Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating bulk transcriptomics and single-cell RNA sequencing reveals protein secretion pathway genes and regulatory mechanisms in osteoporosis

Limin He
The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine
Po Yang
The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine
Dandan Wang
lim_add_oil@163.com
The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine
Tietao Di
The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine
Yan Jiang
The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine

Article

Keywords: Osteoporosis, Protein secretion pathway, Key genes, BM-MSCs, Single-cell RNA sequencing

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11072766/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Osteoporosis (OP) is a systemic skeletal disorder in which altered protein-secretory homeostasis may contribute to impaired bone remodeling. We integrated bulk transcriptomic datasets GSE56815 and GSE2208 with the single-cell RNA-sequencing dataset GSE147287 to identify protein secretion pathway-associated genes and their cellular context in OP. Differential expression, weighted gene co-expression network analysis, three machine-learning algorithms, immune infiltration analysis, gene set enrichment analysis, regulatory network construction, and single-cell analyses were performed, followed by RT-qPCR validation. PPWD1 and TMEM50B were prioritized as key genes and were consistently upregulated in OP. Both genes were associated with proteasome, ribosome, and spliceosome pathways. Single-cell analysis highlighted bone marrow mesenchymal stem cells (BM-MSCs) as a relevant cell population and suggested differences in PPWD1 and TMEM50B expression along pseudotime, together with candidate CXCL12-CXCR4-mediated communication. RT-qPCR confirmed increased expression of both genes in OP samples. These findings identify PPWD1 and TMEM50B as protein secretion pathway-associated genes in OP and provide a transcriptomic framework linking proteostasis-related signatures, BM-MSC states, and osteoimmune alterations while requiring further functional validation.

1. Introduction

Osteoporosis (OP) is a systemic skeletal disorder defined by progressive bone mass loss and deterioration of trabecular microarchitecture, imposing a substantial global public health burden. Over 200 million individuals worldwide live with OP, and population aging continues to increase its prevalence and the health burden associated with fragility fractures [1]. Population-based epidemiological cohorts further demonstrate that one-third of women and one-fifth of men develop OP throughout their lifespans, with more than half of postmenopausal white women experiencing osteoporotic fractures [1–3]. OP pathogenesis arises from multilayered crosstalk between endocrine dysfunction, genetic predisposition, and chronic inflammatory signaling, where the imbalance between bone formation and bone resorption constitutes the central pathological driver of sustained bone loss. Bone formation depends substantially on the osteogenic commitment of bone marrow mesenchymal stem cells (BMSCs). Under osteoporotic pathological conditions, BMSCs exhibit reduced osteogenic capacity alongside a skewed adipogenic differentiation bias, collectively resulting in diminished bone volume and excessive lipid accumulation within the bone marrow niche [4, 5]. Current pharmacological therapies for OP mainly include: anti-resorptive agents (bisphosphonates, denosumab) and bone-anabolic peptides (teriparatide). However, long-term antiresorptive treatment may be associated with rare adverse events, including osteonecrosis of the jaw and atypical femoral fractures, and fracture risk may increase after discontinuation of some agents, particularly denosumab [6, 7]. These clinical limitations underscore the need to investigate the molecular cascades governing OP onset and progression. Systematic identification of disease-associated genes and signaling pathways may improve the understanding of OP biology and provide candidates for subsequent mechanistic and therapeutic studies.

The protein secretion pathway is a central cellular process encompassing endoplasmic reticulum (ER)-based protein folding, post-translational processing, vesicular trafficking, and extracellular secretion; related proteostasis systems, including proteasomal degradation, RNA splicing, and mitochondrial quality control, may interact with secretory homeostasis. Accumulating evidence links disturbances in protein-folding, degradation, and secretion-related processes to bone-cell function and skeletal homeostasis [8]. The ER stress effector XBP1s participates in proteasome inhibition-associated osteogenic responses, supporting crosstalk between ER stress and protein degradation in bone formation [9]. The ubiquitin-proteasome system preserves skeletal homeostasis via targeted proteolysis of master transcription factors that dictate osteoblast and osteoclast differentiation; functional collapse of this machinery can disrupt bone metabolic equilibrium [10, 11]. Spliceosome dysfunction also imposes potent inhibitory effects on bone formation: NIBAN2 remodels RUNX2 alternative splicing through the HNRNPU-centered spliceosome complex to regulate osteoblast maturation, whereas loss-of-function mutations in the spliceosome subunit SNRPB markedly impair osteogenic differentiation and induce severe developmental skeletal malformations [12, 13]. Beyond nuclear and cytoplasmic regulatory modules, proper folding and extracellular secretion of type I collagen constitute an important prerequisite for bone matrix mineralization, and secretory dysfunction may contribute to insufficient extracellular matrix deposition. Mitochondrial proteostasis further governs BMSC fate specification: deletion of TMEM135 disrupts mitochondrial dynamics and impairs osteogenic differentiation, while defective mitophagy accelerates BMSC senescence and further blunts bone-forming potential [14–16]. Despite these observations, which protein-secretion-associated genes are reproducibly associated with OP and in which cell populations they are preferentially expressed remain incompletely characterized. Integrative transcriptomic profiling may help address these gaps and generate testable mechanistic hypotheses for OP.

To address these gaps, the present study integrates bulk transcriptomic datasets and single-cell RNA sequencing (scRNA-seq) profiles to systematically screen candidate key genes associated with the protein secretion pathway in OP. Candidate genes were first prioritized by intersecting differentially expressed genes (DEGs) and co-expressed module genes derived from weighted gene co-expression network analysis (WGCNA) enriched for protein secretion signatures. Three complementary machine learning algorithms—SVM-RFE, XGBoost, and random forest—were subsequently deployed to narrow down reproducible candidate key genes. We then implemented a multi-dimensional analytical pipeline encompassing gene set enrichment analysis (GSEA), quantitative immune cell infiltration profiling, construction of transcription factor (TF)-gene-microRNA (miRNA) co-regulatory networks, protein subcellular localization prediction, and comprehensive scRNA-seq characterization covering cell-type annotation, cell-cell communication mapping, pseudotime trajectory reconstruction, and Reactome pathway enrichment. Finally, reverse transcription quantitative PCR (RT-qPCR) was performed to experimentally validate the differential expression patterns of identified hub genes. Collectively, this integrative transcriptomic framework characterizes associations of protein secretion-related genes with OP and provides a structured basis for prioritizing candidate mechanisms and molecules for subsequent functional validation.

2. Results

2.1 Candidate genes associated with protein secretion in OP

In the training dataset, differential expression testing yielded 412 DEGs between OP and control samples at |log2FC| > 0.15 and p < 0.05. Of these, 163 were increased and 249 were decreased in the OP group (**Supplementary Table S1**, Fig. 1a, b). Evaluation of the protein secretion pathway-related gene set additionally identified 19 genes that differed significantly between groups by the Wilcoxon test (p < 0.05). The corresponding ssGSEA scores also differed significantly between OP and control samples (p < 0.01) (Fig. 1c).

No sample was identified as an outlier during WGCNA of the training cohort (**Supplementary Figure S1**). A soft-thresholding power of 7 was selected because the scale-free fit exceeded the red reference line (R2 = 0.85) while mean connectivity approached 0 (**Supplementary Figure S1**). Merging related co-expression modules produced 6 gene modules after removal of the gray unclassified module, with each module represented by a distinct color (Fig. 1d, e). Applying GS > 0.2 and MM > 0.8 retained 227 module genes, comprising 194 genes in MEblue and 33 in MEbrown (**Supplementary Figure S1**). Intersecting these module genes with the 412 DEGs produced 25 candidate genes (Fig. 1f).

2.2 Functional enrichment and PPI network of candidate genes

GO analysis returned 588 significant annotations (p < 0.05): 512 BP terms, including embryonic placenta development and response to fluid shear stress; 42 CC terms, including Z disc; and 34 MF terms, including non-membrane spanning protein tyrosine kinase activity. KEGG analysis identified 33 significantly enriched pathways (**Supplementary Table S2**, Fig. 2a), with carbohydrate digestion and absorption and the chemokine signaling pathway among the principal findings. A PPI network was subsequently generated from the 25 candidate genes at an interaction-score threshold of ≥ 0.15. The network contained 23 connected nodes and 53 edges, whereas 2 genes had no connections (Fig. 2b). AKT1 and HCK showed particularly prominent interactions with other encoded proteins. Overall, the candidate set mapped to developmental, metabolic, and signaling functions and formed an interconnected protein network relevant to OP-associated regulation.

2.3 Screening and validation of key genes

Feature selection was performed on the 25 candidate genes with all three machine-learning models. SVM-RFE retained 21 genes and achieved an accuracy of 0.738 (Fig. 3a). XGBoost also retained 21 genes with Gain > 0, whereas RF ranked 10 genes as the leading features at ntree = 12 and min error = 0.2875 (Fig. 3c-e). Seven genes were common to the three feature sets: HCK, AKT1, FLNA, PTPN6, PPWD1, SPINT1, and TMEM50B (Fig. 3f).

These seven feature genes were next assessed in both the training and validation datasets using the Wilcoxon test ($p < 0.05$). Only PPWD1 and TMEM50B showed significant differential expression in both cohorts, and both were consistently higher in OP than in controls (Fig. 3g, h). HCK could not be assessed in the validation dataset because expression data were unavailable. On the basis of the cross-dataset consistency, PPWD1 and TMEM50B were retained as the key genes associated with OP.

2.4 Immune infiltration and GSEA of key genes

xCell-based profiling quantified 64 immune and stromal cell types. Megakaryocytes, mesenchymal stem cells (MSCs), and eosinophils had the three highest mean enrichment scores (**Supplementary Table S3**, Fig. 4a). Nine cell types differed significantly between OP and control samples ($p < 0.05$) (Fig. 4b); Th1 cells showed higher inferred abundance in OP, whereas mast cells showed lower inferred abundance. Spearman analysis identified a strong positive association between conventional dendritic cells (cDC) and dendritic cells (DC) ($cor = 0.77$, $p < 0.001$), as well as a positive association between plasma cells and PPWD1 ($cor = 0.41$, $p < 0.001$). PPWD1 and TMEM50B were themselves positively correlated ($cor = 0.73$, $p < 0.001$) (Fig. 4c).

GSEA identified KEGG pathways associated with each key gene using $|NES| > 1$ and $p < 0.05$. PPWD1 was associated with 56 enriched pathways (Fig. 4d), including proteasome, ribosome, and spliceosome, while TMEM50B was associated with 51 pathways (Fig. 4e), again including ribosome, proteasome, and spliceosome (**Supplementary Table S4**). The two genes shared 46 enriched pathways (Fig. 4f), with prominent signals involving proteasome, ribosome, spliceosome, oxidative phosphorylation, RNA degradation, and protein export (Fig. 4g). NES patterns were highly similar across these common pathways. Thus, the OP cohort displayed altered immune-cell enrichment, and both key genes were linked to signatures involving protein degradation, ribosomal activity, and energy metabolism.

2.5 Molecular regulatory network and subcellular localization of key genes

The integrated TF-gene-miRNA network showed that PPWD1 was linked to seven TFs, including CREB1 and GATA2, and to 112 miRNAs, including hsa-let-7a-5p and hsa-let-7b-5p. TMEM50B was connected to five TFs, including ARID3A and HNF4A, and to 116 miRNAs, including hsa-miR-16-5p and hsa-miR-21-5p. In total, the network contained 197 nodes and 240 interaction edges (**Supplementary Table S5**, Fig. 5a).

Predicted subcellular localization differed between the two proteins. PPWD1 showed its highest score for combined cytoplasmic and nuclear localization (cyto_nucl, score = 17.8), followed by nuclear (nucl, score = 15.5) and cytoplasmic (cyto, score = 15.0) localization. By contrast, TMEM50B was predicted primarily at the plasma membrane (plas, score = 31.0) (Fig. 5b). These analyses therefore suggested distinct cellular distributions for the encoded proteins alongside multi-TF and multi-miRNA regulatory associations.

2.6 Single-cell transcriptomic landscape and identification of BM-MSCs as the key cell type

Before filtering, the single-cell dataset contained 9,654 cells and 20,420 genes. Quality-control procedures retained 7,497 cells while preserving 20,420 genes for subsequent analysis (**Supplementary Figure S2**). Normalization followed by variable-feature selection identified 2,000 HVGs, among which GNLY, IGKC, and IBSP were the three most variable transcripts (**Supplementary Figure S2**). Using the first 30 PCs for dimensionality reduction separated the cells into 15 clusters (**Supplementary Figure S2**; Fig. 6a). Marker-based annotation assigned these clusters to nine cell lineages: bone marrow mesenchymal stem cells (BM-MSCs), neutrophils, plasma cells, monocytes, T cells, nucleated red blood cells (NRBCs), cycling cells, natural killer (NK) cells, and B cells (**Supplementary Table S6**; Fig. 6b). Examination of lineage markers showed cell-type-specific expression patterns for the annotated clusters (Fig. 6c).

Mapping PPWD1 and TMEM50B across the nine annotated populations (Fig. 6d) showed that both genes had prominent expression in BM-MSCs (Fig. 6e). Because BM-MSCs displayed the highest average expression pattern for the key genes, this population was selected for the subsequent cell-level analyses.

2.7 Intercellular communication and differentiation dynamics of BM-MSCs

Communication patterns among the nine annotated cell types were compared with emphasis on BM-MSCs. The inferred network indicated that BM-MSCs were prominent signal-sending cells and communicated with several immune and proliferating populations, including T cells and plasma cells (Fig. 7a, b). Ligand-receptor analysis further indicated that CXCL12-CXCR4 mediated predicted signaling from BM-MSCs to T cells, plasma cells, neutrophils, B cells, and cycling cells (Fig. 7c, d).

BM-MSC differentiation dynamics were further examined by pseudotime reconstruction. Re-clustering of BM-MSCs with the top 20 PCs at resolution = 0.3 generated five subpopulations (Fig. 7e). The reconstructed trajectory formed two terminal branches: cells began at node 1, passed through node 2, and then separated, with one branch extending toward the upper pseudospace and dominated by cluster 4, and the other extending leftward and enriched for cluster 2. Along this trajectory, PPWD1 expression was highest from the early to intermediate stages and declined toward the terminal stage. TMEM50B followed the opposite pattern, increasing progressively and reaching its maximum during the intermediate-to-late stages before decreasing slightly (Fig. 7f). The opposing trajectories indicate that the two genes are associated with different temporal states of BM-MSC differentiation.

2.8 Validation of key gene expression by RT-qPCR

RT-qPCR showed significantly higher expression of both PPWD1 and TMEM50B in OP samples than in controls (PPWD1: $p = 0.0056$; TMEM50B: $p = 0.0108$) (Fig. 8). The direction of change agreed with the upregulation observed in the training and validation datasets.

3. Discussion

OP arises from multifaceted pathological cascades dominated by uncoupled bone formation and resorption [1]. Protein secretion-related processes contribute to osteogenic cytokine and extracellular-matrix biosynthesis, yet their core mediators and cell-specific effectors remain poorly defined. This study integrated bulk and single-cell transcriptomic analyses to prioritize two candidate key genes (PPWD1 and TMEM50B) and highlight BM-MSCs as a cell population of interest, providing a multi-layered association framework for subsequent mechanistic investigation in OP.

However, pathway co-enrichment does not by itself demonstrate that PPWD1 and TMEM50B directly control these pathways. Rather, the shared proteasome, ribosome, spliceosome, oxidative phosphorylation, and protein-export signatures suggest that both genes mark a coordinated proteostasis state in circulating monocytes. This interpretation is consistent with recent evidence that maladaptive ER stress can suppress osteoblast differentiation and bone matrix mineralization, whereas pharmacological attenuation of ER stress partially restores bone formation in osteoporotic models [17].

Briefly, PPWD1 functions as a spliceosome C complex subunit controlling pre-mRNA maturation. Prior bone-related WGCNA research linked elevated PPWD1 to low bone mineral density in postmenopausal females [18]. Recent studies have shown that spliceosome dysregulation can impair osteoblast maturation: NIBAN2 modulates RUNX2 alternative splicing via HNRNPU complexes, while SNRPB deficiency impairs osteogenic differentiation [12, 13]. Accordingly, PPWD1 upregulation may be associated with an altered splicing environment in OP; however, its specific RNA substrates and functional effects on BM-MSC osteogenic commitment remain unknown. Recent work demonstrated that NIBAN2 promotes osteoblast differentiation by regulating HNRNPU-centered RUNX2 alternative splicing, providing direct evidence that spliceosomal remodeling can alter bone formation [12]. In this context, PPWD1 may serve as a marker or regulator of an altered splicing environment; however, its specific RNA substrates in BM-MSCs remain to be identified by RNA immunoprecipitation and isoform-level sequencing.

In contrast, TMEM50B is distinct from TMBIM5; the mitochondrial Ca^{2+}/H^{+} antiporter activity characterized for TMBIM5 should not be attributed to TMEM50B [19]. Studies of TMBIM5 and other mitochondrial regulators provide broader context for mitochondrial homeostasis but do not constitute direct functional evidence for TMEM50B [20, 21]. The paralogous mitochondrial regulator TMEM135 similarly balances BM-MSC osteogenic-adipogenic fate, and its ablation triggers severe osteopenia and marrow steatosis [14]. Consistent with these observations, the shared enriched pathways of PPWD1 and TMEM50B center on the ubiquitin-proteasome system and ER stress effector XBP1s, two canonical modules governing skeletal turnover [9–11]. Thus, the current data support an association of TMEM50B with proteostasis-related signatures, but its molecular function in BM-MSCs and OP remains to be established experimentally.

BM-MSCs represent the primary reservoir of osteoblast precursors, and OP is commonly characterized by their attenuated osteogenic capacity and enhanced adipogenic bias [4, 5]. Single-cell atlases of aged bone marrow have documented heterogeneous dysfunction within $Lepr^{+}CXCL12^{+}$ BM-MSC

subpopulations that show increased adipogenic tendency and senescence-associated changes [22, 23]. Accumulated reactive oxygen species and defective mitophagy exacerbate this fate switch [15, 16]. Our pseudotime analysis showed contrasting expression patterns for PPWD1 and TMEM50B: PPWD1 was higher during early-to-intermediate pseudotime, whereas TMEM50B increased toward intermediate-to-late pseudotime states, suggesting associations with distinct BM-MSC states rather than direct stage-specific regulatory effects [14]. CellChat analysis identified BM-MSCs as prominent predicted signal senders via the CXCL12-CXCR4 ligand-receptor axis, a core mediator of stromal-immune communication in the bone niche; CXCL12 overexpression has been reported to enhance bone formation in osteoporotic mice [24]. Reactome pathway enrichment additionally indicated higher inferred activity of DNA damage repair and lipid metabolism in OP BM-MSCs, indicative of persistent stress compensation. These multi-dimensional single-cell data collectively suggest that altered BM-MSC differentiation, niche communication, and metabolic states may contribute to unbalanced bone turnover.

Recent studies further support a mechanistic connection between mitochondrial fitness and BM-MSC fate. Mitochondrial transfer can remodel metabolism and improve osteogenic differentiation in osteoporosis [21], whereas preservation of β-catenin-dependent stemness enhances the osteogenic potential of aging BM-MSCs [22]. These findings are consistent with the present pseudotime results, but the inferred early-versus-late roles of PPWD1 and TMEM50B remain descriptive because pseudotime does not establish lineage direction or causality. Moreover, SDF-1/CXCL12-overexpressing BM-MSCs have been reported to enhance bone formation in osteoporotic mice [24], supporting the biological plausibility of the CXCL12-CXCR4 communication axis. Nevertheless, because the present scRNA-seq dataset contains only one OP sample and lacks a matched healthy bone-marrow control, this axis should be interpreted as a candidate niche interaction rather than an OP-specific mechanism.

Protein secretion is closely connected to ER quality control and may interact with proteasomal degradation, RNA splicing, and mitochondrial proteostasis, processes that can influence bone matrix production and mineralization [8]. The shared pathways enriched by PPWD1 and TMEM50B align with established bone regulatory axes: unresolved ER stress induces osteoblast apoptosis, while proteasomal over-degradation and abnormal splicing both disrupt RUNX2-dependent osteogenesis [12, 17, 25]. These results link pathway-level proteostasis signatures with osteoporotic phenotypes, but functional experiments are needed to establish directionality.

Immune infiltration analysis revealed higher inferred Th1-cell enrichment and lower inferred mast-cell enrichment in the training cohort, alongside a positive correlation between PPWD1 expression and plasma-cell enrichment. Th1-secreted IFN-γ exerts dual bone-regulatory effects: it inhibits osteoclastogenesis under homeostatic conditions yet sustains pro-resorptive inflammation in high-RANKL microenvironments [26]. Mast cells directly drive osteoclast formation via surface RANKL, so the lower mast-cell score may reflect changes in immune composition or transcriptional state [27]. Broad osteoimmunology research confirms sustained immune niche remodeling exacerbates bone turnover imbalance, supporting an association between PPWD1 upregulation and altered immune-cell enrichment [28]. Recent single-cell studies have identified pro-inflammatory T-cell, macrophage, neutrophil, and B-cell

subpopulations that activate cytokine and osteoclast-differentiation programs in OP [26, 28]. This evidence supports the observed Th1 shift, but the lower mast-cell score and PPWD1-plasma-cell correlation should not be interpreted as direct cellular depletion or regulation because xCell estimates relative enrichment from bulk monocyte expression profiles. Flow cytometry and bone-marrow tissue validation are required.

Finally, the TF-gene-miRNA network predicted extensive potential upstream transcriptional and post-transcriptional regulation of both key genes: PPWD1 was linked to seven TFs and 112 miRNAs, whereas TMEM50B was linked to five TFs and 116 miRNAs. Transcription factors are well-established major regulators of osteoblast/osteoclast transcriptional programs [10, 11], suggesting candidate mechanisms that may contribute to PPWD1/TMEM50B dysregulation. This tiered framework, spanning proteostasis breakdown, osteoimmune reshaping and complex transcriptional control, provides a hypothesis-generating framework for prioritizing upstream regulators.

Although this integrative transcriptomic framework suggests coordinated associations of PPWD1 and TMEM50B with BM-MSC-related processes, several key limitations must be acknowledged. First, all transcriptomic and single-cell datasets were retrieved from public repositories with limited sample sizes, lacking validation via multi-center prospective cohorts across diverse ethnic groups and OP disease stages. Second, while BM-MSCs were identified as the cell population with the highest average expression, the divergent biological roles of distinct BM-MSC subclones remain underexplored and may be addressed using spatial transcriptomics to map in situ niche activity. Most importantly, our expression validation is limited to RT-qPCR at the mRNA level; loss-of-function and overexpression cellular assays, combined with gene-edited osteoporotic animal models, are needed to test the functional effects of PPWD1/TMEM50B on BM-MSC osteogenic-adipogenic fate switching. Future work integrating spatial multi-omics and in vivo functional manipulation will further refine our mechanistic model and clarify the translational relevance of these findings.

Collectively, this study integrates bulk transcriptomic and scRNA-seq data to identify PPWD1 and TMEM50B as protein secretion pathway-associated genes in OP and to highlight BM-MSCs as a relevant cell population. The findings connect these genes with proteostasis-related pathways, osteoimmune alterations, and stromal-cell communication, while providing testable hypotheses for subsequent mechanistic studies. Further validation in independent bone-marrow cohorts and functional models is required before these genes can be considered therapeutic targets.

4. Methods

4.1 Data source

Publicly available transcriptomic and scRNA-seq profiles for OP were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). GSE56815 (GPL96 platform) served as the training set and included blood-monocyte samples from 40 individuals with OP and 40 controls. The

validation dataset, GSE2208 (GPL96 platform), consisted of blood-monocyte microarray data from 9 OP samples and 10 control samples. For single-cell analysis, GSE147287 (GPL24676 platform) provided high-throughput transcriptomic data from 1 OP bone marrow mononuclear-cell sample; 1 osteoarthritis sample was excluded. In addition, 96 protein secretion-related genes were obtained from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb) by downloading the "HALLMARK_PROTEIN_SECRETION" gene set and removing redundant entries; the complete list is provided in **Supplementary Table S7**.

4.2 Acquisition of candidate genes

Differential expression in the training cohort was assessed with the "limma" package (v3.58.1) [29]. Genes meeting $|\log_2 \text{Fold Change (FC)}| > 0.15$ and p-value (p) < 0.05 were designated as DEGs [30]. A Wilcoxon test was also used to evaluate protein secretion-related genes between the OP and control groups ($p < 0.05$). For genes showing significant differential expression, single-sample gene set enrichment analysis (ssGSEA) scores were calculated across all training samples with the "GSVA" package (v1.50.0) [31].

Weighted gene co-expression network analysis (WGCNA) was then carried out for the full training set with the "WGCNA" package (v1.72.5) [32] to identify modules most strongly associated, in either direction, with the ssGSEA scores. Samples were first hierarchically clustered to screen for outliers. The soft-thresholding power was chosen to achieve a scale-free topology with $R^2 > 0.85$ while maintaining sufficient connectivity. Modules were defined using minModuleSize = 100 and mergeCutHeight = 0.25. Correlations between module eigengenes and ssGSEA scores were computed with the "cor" function in "WGCNA" (v1.72.5). Genes satisfying gene significance (GS) > 0.2 and module membership (MM) > 0.8 within the modules showing the strongest positive or negative associations with ssGSEA scores constituted the WGCNA module gene set. Finally, the "ggvenn" package (v0.1.10) [33] was used to intersect this set with the DEGs; genes shared by both sets were retained as candidate genes.

4.3 Functional enrichment and protein-protein interaction (PPI) network analyses of candidate genes

Functional interpretation of the candidate genes was performed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment [34, 35]using the "clusterProfiler" package (v4.8.3) [36], with $p < 0.05$ as the significance criterion. GO annotations were summarized within Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Protein-protein interaction information was downloaded from the search tool for the retrieval of interacting genes/proteins (STRING) database (https://www.string-db.org) through the "curl" package (v5.1.0). The resulting PPI network was then assembled and evaluated with the "igraph" package (v2.0.3), applying an interaction-score cutoff of ≥ 0.15.

4.4 Screening of key genes

Three machine-learning approaches were applied to the training set to prioritize key feature genes with potential diagnostic relevance to OP. Support Vector Machine-Recursive Feature Elimination (SVM-RFE) was implemented with 5-fold cross-validation using the "caret" package (v6.0-94). Extreme Gradient Boosting (XGBoost) was run with the "xgboost" package (v2.0.3.1), setting the maximum tree depth to 3, the learning rate to 0.3, and the number of training iterations to 25; genes with Gain > 0 were retained as XGBoost features. Random Forest (RF) modeling used the "randomForest" package (v4.7.1.1) with 5-fold cross-validation, and the 10 genes with the highest importance scores were selected. Feature sets generated by SVM-RFE, XGBoost, and RF were then intersected with the "ggvenn" package (v0.1.10) to obtain the key feature genes.

The reproducibility of key feature-gene expression was examined in both the training and validation cohorts. OP and control expression levels were compared with the Wilcoxon test, using $p < 0.05$ to define a significant difference. Genes that remained significant and showed the same direction of expression change in both datasets were classified as key genes.

4.5 Immune infiltration analysis

Immune and stromal composition in the training dataset was estimated for 64 cell types [37] using the xCell algorithm implemented in the "IOBR" package (v0.99.8) [38]. OP-versus-control differences in estimated cell abundance were tested with the "wilcox_test" function from "rstatix" (v0.7.2), with $p < 0.05$ considered significant. Spearman correlations were subsequently calculated with the "psych" package (v2.4.3) for pairs of differentially abundant cell types and for relationships between these cell types and the key genes. Correlations were considered significant when |correlation coefficient (cor)| > 0.30 and $p < 0.05$.

4.6 GSEA

Gene set enrichment analysis was used to characterize pathways associated with each key gene. KEGG pathway gene sets were obtained from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb) through the "msigdbr" package (v7.5.1) [39]. For each key gene, every gene in the training dataset was ranked according to its Spearman correlation with that key gene, calculated with the "stats" package (v4.3.1) [40]. GSEA was then performed using "clusterProfiler" (v4.8.3), and pathways meeting |Normalized Enrichment Score (NES)| > 1 and $p < 0.05$ were retained.

4.7 Molecular regulatory network construction and subcellular localization

Potential upstream regulation of the key genes was explored by building a TF-gene-miRNA network. Candidate transcription factors and miRNAs were predicted from the JASPAR database (https://jaspar.genereg.net) and TarBase database (http://microrna.gr/tarbase), respectively, through the NetworkAnalyst platform (https://www.networkanalyst.ca). TF-miRNA regulatory links were subsequently obtained from the RegNetwork database (http://regnetworkweb.org) and combined with the predicted TF-target and miRNA-target relationships. The integrated network was displayed in

Cytoscape (v3.10.3) [41]. In parallel, protein FASTA sequences for the key genes were downloaded from UniProt (https://www.uniprot.org) and analyzed with WoLF PSORT (https://wolfpsort.hgc.jp) to predict subcellular localization.

4.8 Dimensionality reduction, clustering and cell type annotation of scRNA-seq data

Quality control of the scRNA-seq dataset was conducted in Seurat (v5.0.1) [42], with mitochondrial content calculated using the "PercentageFeatureSet" function. Genes detected in at least three cells were kept. Cells were retained when nFeature_RNA ranged from 200 to 5,000, percent.mt was < 20%, and nCount_RNA was < 20,000. The raw matrix was normalized with "NormalizeData", after which data layers were combined with "JoinLayers". The 2,000 most highly variable genes (HVGs) were selected by "FindVariableFeatures" using the "vst" method. Data were then scaled with "ScaleData", and principal component analysis (PCA) was performed using "RunPCA". Principal-component significance was assessed with "JackStraw" and "ScoreJackStraw", while the variance profile was inspected using "ElbowPlot". On the basis of these evaluations, the first 30 PCs were carried forward for downstream analyses.

A shared nearest neighbor (SNN) graph was generated from the first 30 PCs with "FindNeighbors", and clusters were called by "FindClusters" at resolution = 0.3. Two-dimensional visualization was obtained with "RunUMAP" for Uniform Manifold Approximation and Projection (UMAP). Cell identities were assigned by integrating the clustering results with cell-type-specific marker genes reported previously [43] and those listed in the CellMarker 2.0 database (https://bio-bigdata.hrbmu.edu.cn/CellMarker/) (accessed on 2 June, 2026). The distribution of each key gene was then compared across annotated cell types, and the cell type showing the highest average expression of the key genes was taken forward as the key cell type.

4.9 Cell communication, pseudotime trajectory, and ReactomeGSA enrichment analysis

Cell-cell signaling within the single-cell dataset was inferred with the "CellChat" package (v1.6.1) [44]. Ligand-receptor communication probabilities were evaluated across all annotated cell types, with focused examination of the key cell type identified above to characterize the interaction signals associated with each cell population.

Differentiation-related heterogeneity within the key cell type was investigated by pseudotime analysis. These cells underwent a second round of dimensionality reduction and clustering using the first 20 PCs at resolution = 0.3 to define subclusters. The "Monocle2" package (v2.30.1) [45] was then run with default parameters to embed the key-cell subpopulations in a low-dimensional UMAP space. A predefined "pseudoroot" state was used as the starting point for construction of the pseudotime trajectory representing possible developmental paths during OP progression. Expression of the key genes was subsequently mapped along this trajectory to depict temporal expression changes.

Cell-type-level pathway activity was assessed with the "analyse_sc_clusters" function in the "ReactomeGSA" package (v3.20) [46]. For each pathway, the difference between the highest and lowest activity scores across cell types (max_difference) was calculated. The 10 pathways with the largest max_difference values were selected and presented as a heatmap.

4.10 RT-qPCR

RT-qPCR was used for experimental verification of key-gene expression. Frozen whole-blood samples were obtained from six patients with OP and six controls. The study received approval from the Ethics Committee of our hospital (No. Lw20260329), and all participants provided written informed consent. All procedures in this study were performed in accordance with the Declaration of Helsinki. RNA was isolated from each specimen with TRIzol (Novazon, Nanjing, China) following the manufacturer's instructions, and RNA concentration was determined on a NanoPhotometer N50 (Implen, Germany). Complementary DNA (cDNA) was generated from total RNA using the HP All-in-one qRT Master Mix II (Kunming Yungen Biotechnology Co., Ltd., Kunming, China) according to the manufacturer's instructions. Amplification was performed for 39 cycles on a CFX Connect Real-Time PCR Instrument (BIO-RAD, USA) with 2×Universal Blue SYBR Green qPCR Master Mix (Kunming Yungen Biotechnology Co., Ltd., Kunming, China). Primer sequences are listed in **Supplementary Table S8**, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) served as the internal reference. Relative key-gene expression was determined by the $2 - \Delta\Delta Ct$ method. Each experiment was independently repeated at least three times. GraphPad Prism (v10.1.2) [47] was used for statistical analysis and visualization, and $p < 0.05$ denoted statistical significance.

4.11 Statistical analysis

All statistical analyses were carried out in R software (v4.2.2). Between-group comparisons used the Wilcoxon test, with $p < 0.05$ defining statistical significance. Significance symbols were assigned as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$; ns denotes a non-significant result.

Abbreviations

Abbreviation	Definition
OP	Osteoporosis
GEO	Gene Expression Omnibus
DEG	Differentially expressed gene
WGCNA	Weighted gene co-expression network analysis
GSVA	Gene set variation analysis
ssGSEA	Single-sample gene set enrichment analysis
PPI	Protein-protein interaction
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
SVM-RFE	Support vector machine-recursive feature elimination
XGBoost	eXtreme Gradient Boosting
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
AUC	Area under the receiver operating characteristic curve
GSEA	Gene set enrichment analysis
MSigDB	Molecular Signatures Database
TF	Transcription factor
miRNA	MicroRNA
scRNA-seq	Single-cell RNA sequencing
PCA	Principal component analysis
SNN	Shared nearest neighbor
UMAP	Uniform manifold approximation and projection
BM-MSC	Bone marrow mesenchymal stem cell
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
NES	Normalized enrichment score

Declarations

Acknowledgements

The authors would like to express their sincere gratitude to the Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine for generously offering the platform. The contributions of all participants in this study are greatly appreciated.

Author contributions

L.H.: Conceptualization, Data curation, Validation, Visualization, Writing - original draft, Writing - review & editing. D.W.: Data curation, Validation, Visualization, Writing - review & editing. P.Y.: Validation, Visualization, Writing - review & editing. T.D.: Conceptualization, Supervision, Writing - review & editing. Y.J.: Conceptualization, Project administration, Supervision, Writing - review & editing. All authors reviewed and approved the final manuscript and agree to be accountable for the work.

Data availability statement

The transcriptomic and single-cell RNA sequencing datasets analyzed in this study are publicly available from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE56815, GSE2208, and GSE147287. Protein secretion pathway-related genes were obtained from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/). All other data generated or analyzed during this study are included in this article and its Supplementary Information files. Additional information supporting the findings is available from the corresponding author upon reasonable request.

Additional Information

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine (No. Lw20260329). Written informed consent was obtained from all participants. All methods in this study were performed in accordance with the principles outlined in the Declaration of Helsinki.

Consent for publication

Not applicable.

References

1. Girdler, S. J., Lindsey, M. H., Sebastian, A. S. & Nassr, A. Osteoporosis Evaluation and Management in Spine Surgery. *J. Am. Acad. Orthop. Surg.* **32**, e909–e918 (2024).
2. Stevenson, J. Prevention and treatment of osteoporosis in women. *Post. Reprod. Health*. **29**, 11–14 (2023).
3. Porter, J. L. & Varacallo, M. *A. in StatPearls* (StatPearls Publishing, 2026).
4. Qi, M. et al. Erratum: Autophagy Maintains the Function of Bone Marrow Mesenchymal Stem Cells to Prevent Estrogen Deficiency-Induced Osteoporosis: Erratum. *Theranostics* **13**, 4287 (2023).
5. An, F. et al. Research progress on the role of lncRNA-miRNA networks in regulating adipogenic and osteogenic differentiation of bone marrow mesenchymal stem cells in osteoporosis. *Front. Endocrinol. (Lausanne)*. **14**, 1210627 (2023).
6. Kobayashi, T., Morimoto, T., Ito, K., Mawatari, M. & Shimazaki, T. Denosumab vs. bisphosphonates in primary osteoporosis: a meta-analysis of comparative safety in randomized controlled trials. *Osteoporos. Int.* **35**, 1377–1393 (2024).
7. Bal, R., Ambade, R., Singh, N. & Upadhyay, P. Bisphosphonates and Atypical Femur Fractures: Is the Relationship Causal or Casual. *Cureus* **15**, e48141 (2023).
8. Zhong, M., Wu, Z., Chen, Z., Ren, Q. & Zhou J. Advances in the interaction between endoplasmic reticulum stress and osteoporosis. *Biomed. Pharmacother*. **165**, 115134 (2023).
9. Zhang, D. et al. ER stress arm XBP1s plays a pivotal role in proteasome inhibition-induced bone formation. *Stem Cell. Res. Ther.* **11**, 516 (2020).
10. Fan, X. et al. Role of ubiquitination in the occurrence and development of osteoporosis (Review). *Int J. Mol. Med* **54** (2024).
11. Dong, Y., Chen, Y., Ma, G. & Cao, H. The role of E3 ubiquitin ligases in bone homeostasis and related diseases. *Acta Pharm. Sin B*. **13**, 3963–3987 (2023).
12. Zhang, S. et al. Identification of NIBAN2-Regulated RUNX2 Alternative Splicing Presents Novel Strategies for Antagonizing Osteoporosis. *Adv. Sci. (Weinh)*. **12**, e2416536 (2025).
13. Knill, C. et al. Defects of the spliceosomal gene SNRPB affect osteo- and chondro-differentiation. *FEBS J.* **291**, 272–291 (2024).
14. Liu, J. et al. TMEM135 maintains the equilibrium of osteogenesis and adipogenesis by regulating mitochondrial dynamics. *Metabolism* **152**, 155767 (2024).
15. Xiang, K. et al. Tobacco toxins trigger bone marrow mesenchymal stem cells aging by inhibiting mitophagy. *Ecotoxicol. Environ. Saf.* **277**, 116392 (2024).
16. Zhao, B. et al. Leonurine Protects Bone Mesenchymal Stem Cells from Oxidative Stress by Activating Mitophagy through PI3K/Akt/mTOR Pathway. *Cells* 11 (2022).
17. Al-Daghestani, H. et al. Pharmacological inhibition of endoplasmic reticulum stress mitigates osteoporosis in a mouse model of hindlimb suspension. *Sci. Rep.* **14**, 4719 (2024).

18. Qian, G. F. et al. PPWD1 is associated with the occurrence of postmenopausal osteoporosis as determined by weighted gene co–expression network analysis. *Mol. Med. Rep.* **20**, 3202–3214 (2019).
19. Austin, S. et al. TMBIM5 is the Ca(2+) /H(+) antiporter of mammalian mitochondria. *EMBO Rep.* **23**, e54978 (2022).
20. Zhang, L. et al. TMBIM5 loss of function alters mitochondrial matrix ion homeostasis and causes a skeletal myopathy. *Life Sci. Alliance* **5** (2022).
21. Cai, W. et al. Mitochondrial Transfer Regulates Cell Fate Through Metabolic Remodeling in Osteoporosis. *Adv. Sci. (Weinh)*. **10**, e2204871 (2023).
22. Jiang, N. et al. Strategic targeting of miR-183 and beta-catenin to enhance BMSC stemness in age-related osteoporosis therapy. *Sci. Rep.* **14**, 21489 (2024).
23. Zhong, L. et al. Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9** (2020).
24. Wang, Y. et al. Bone marrow mesenchymal stem cells overexpressing stromal cell- derived factor 1 aid in bone formation in osteoporotic mice. *BMC Musculoskelet. Disord*. **25**, 878 (2024).
25. Shen, J. et al. E3 Ubiquitin Ligase-Mediated Regulation of Osteoblast Differentiation and Bone Formation. *Front. Cell. Dev. Biol.* **9**, 706395 (2021).
26. Yang, W., Wang, Y., Mo, K., Chen, W. & Xie, X. Single-cell RNA sequencing reveals multiple immune cell subpopulations promote the formation of abnormal bone microenvironment in osteoporosis. *Sci. Rep.* **14**, 29493 (2024).
27. Ng, C. W. et al. Human mast cells induce osteoclastogenesis through cell surface RANKL. *Inflamm. Res.* **71**, 1261–1270 (2022).
28. Yang, W., Wang, M., Hu, J., Mo, K. & Xie, X. The complex association between the immune system and the skeletal system in osteoporosis: A study of single-cell RNA sequencing. *Int. Immunopharmacol.* **138**, 112611 (2024).
29. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. https://doi.org:10.1093/nar/gkv007 (2015).
30. Xia, Y. et al. Identification of biomarkers associated with endoplasmic reticulum stress-related cell death in osteoporosis based on bulk and single-cell transcriptomic analyses and experimental validation. *Sci Rep* **16** (2026).
31. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. https://doi.org:10.1186/1471-2105-14-7 (2013).
32. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. https://doi.org:10.1186/1471-2105-9-559 (2008).
33. Gao, C. H., Yu, G., Cai, P. & ggVennDiagram An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Front. Genet.* **12**, 706907 (2021).

34. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–d361. https://doi.org:10.1093/nar/gkw1092 (2017).
35. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–d592. https://doi.org:10.1093/nar/gkac963 (2023).
36. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287. https://doi.org:10.1089/omi.2011.0118 (2012).
37. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
38. Zeng, D. et al. Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* **12**, 687975 (2021).
39. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell. Syst.* **1**, 417–425 (2015).
40. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. https://doi.org:10.1101/gr.1239303 (2003).
41. Liu, P., Xu, H., Shi, Y., Deng, L. & Chen, X. Potential Molecular Mechanisms of Plantain in the Treatment of Gout and Hyperuricemia Based on Network Pharmacology. *Evid Based Complement Alternat Med* 3023127 (2020). (2020).
42. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502. https://doi.org:10.1038/nbt.3192 (2015).
43. Peng, J. et al. Identification of lactylation-related biomarkers in osteoporosis from transcriptome and single-cell data. *Front. Endocrinol. (Lausanne)*. **16**, 1621878 (2025).
44. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
45. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
46. Griss, J. et al. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol. Cell. Proteom.* **19**, 2115–2125 (2020).
47. Mitteer, D. R., Greer, B. D., Fisher, W. W. & Cohrs, V. L. Teaching behavior technicians to create publication-quality, single-case design graphs in graphpad prism 7. *J. Appl. Behav. Anal.* **51**, 998–1010 (2018).

Figures

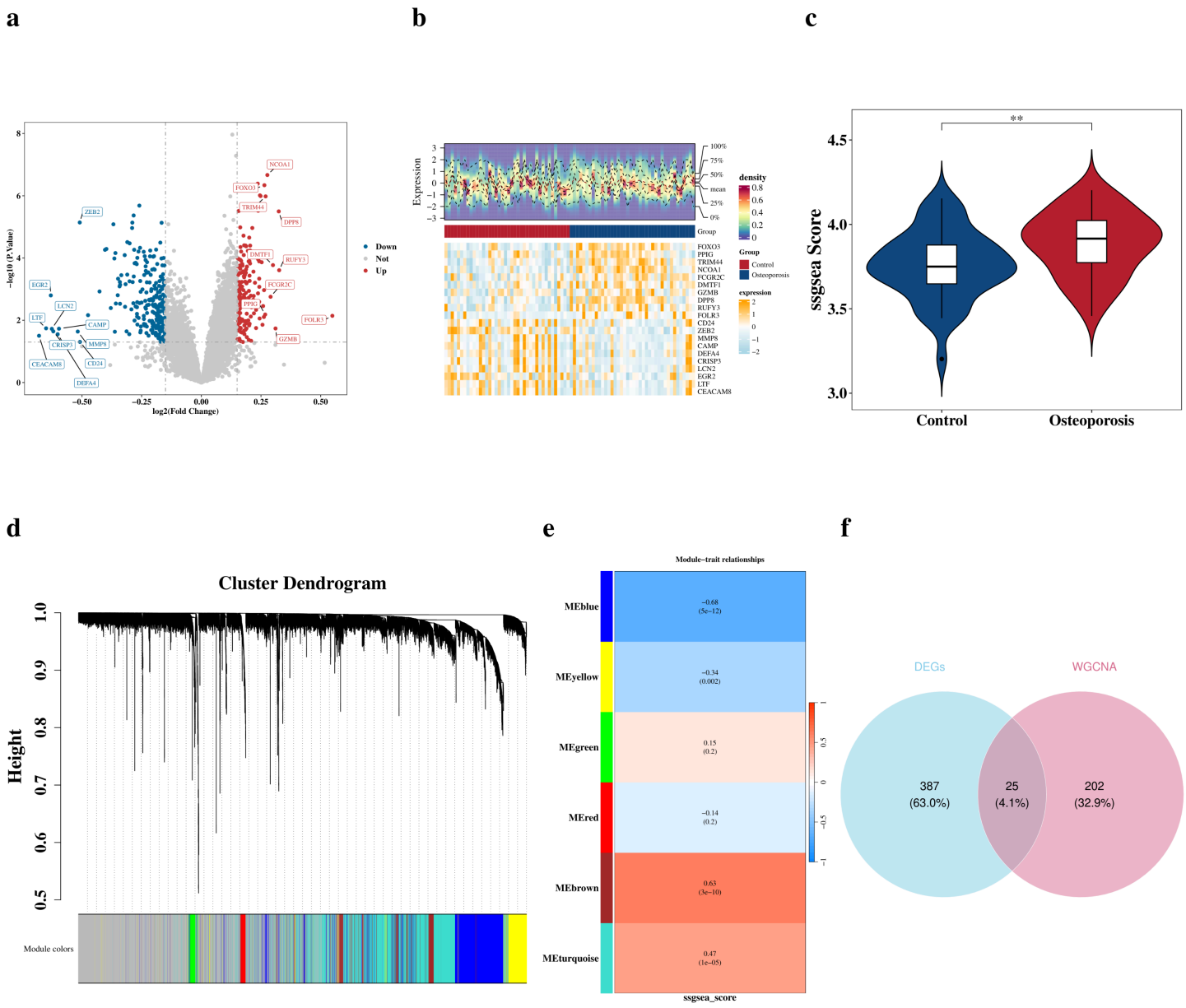


Figure 1

Identification of candidate genes associated with protein secretion in osteoporosis (OP). (a) Volcano plot of differentially expressed genes (DEGs). (b) Heatmap of top DEGs. (c) The single-sample gene set enrichment analysis (ssGSEA) scores of protein secretion-related DEGs. (d) Cluster dendrogram of merged gene modules. (e) Heatmap of module-trait relationships. (f) Venn diagram of DEGs and module genes

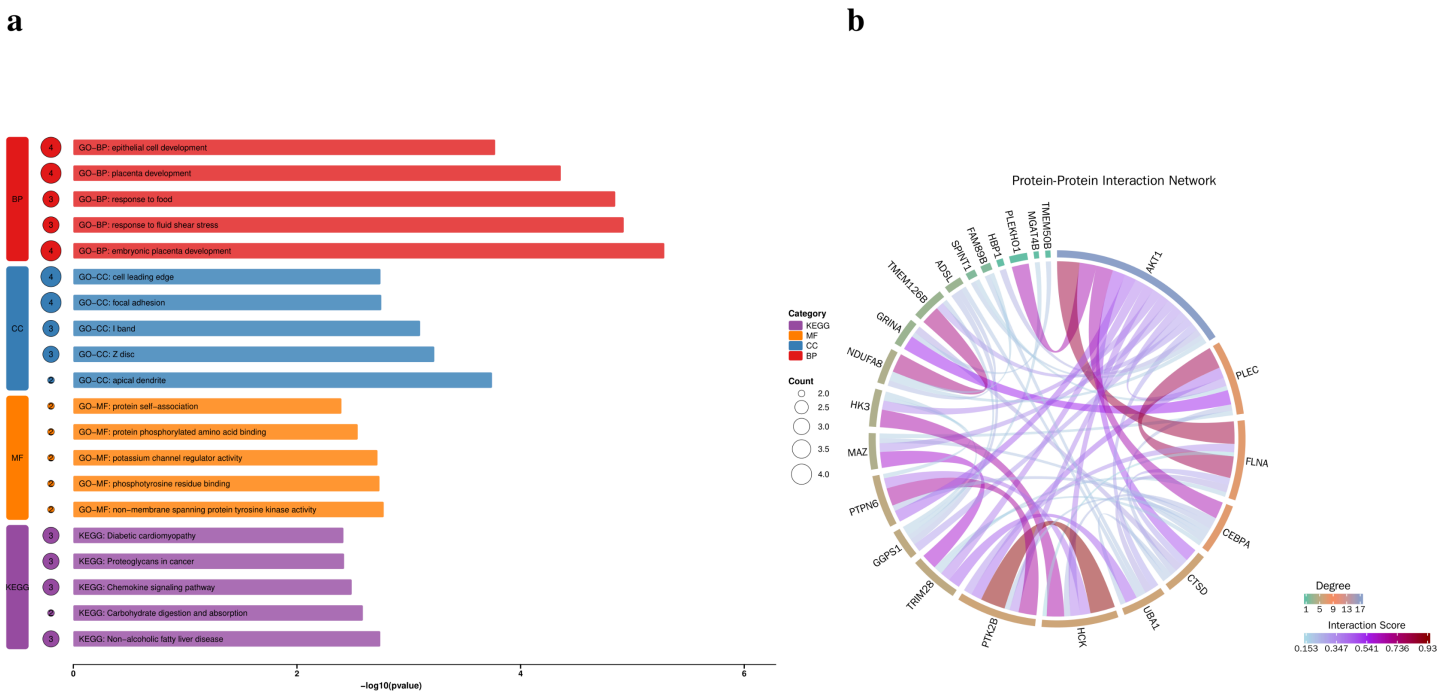


Figure 2

Functional enrichment and protein-protein interaction (PPI) network of candidate genes. (a) Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of candidate genes. (b) PPI network of candidate genes



Figure 3

Screening and validation of key genes. (a) Support vector machine recursive feature elimination (SVM-RFE) accuracy rate curve. (b) SVM-RFE error rate curve. (c) The eXtreme gradient boosting (XGBoost) gene importance. (d) Random forest (RF) feature importance. (e) RF cross-validation error rate. (f) Venn diagram of feature genes from three algorithms. (g) Expression of key genes in the training set. (h) Expression of key genes in the validation set

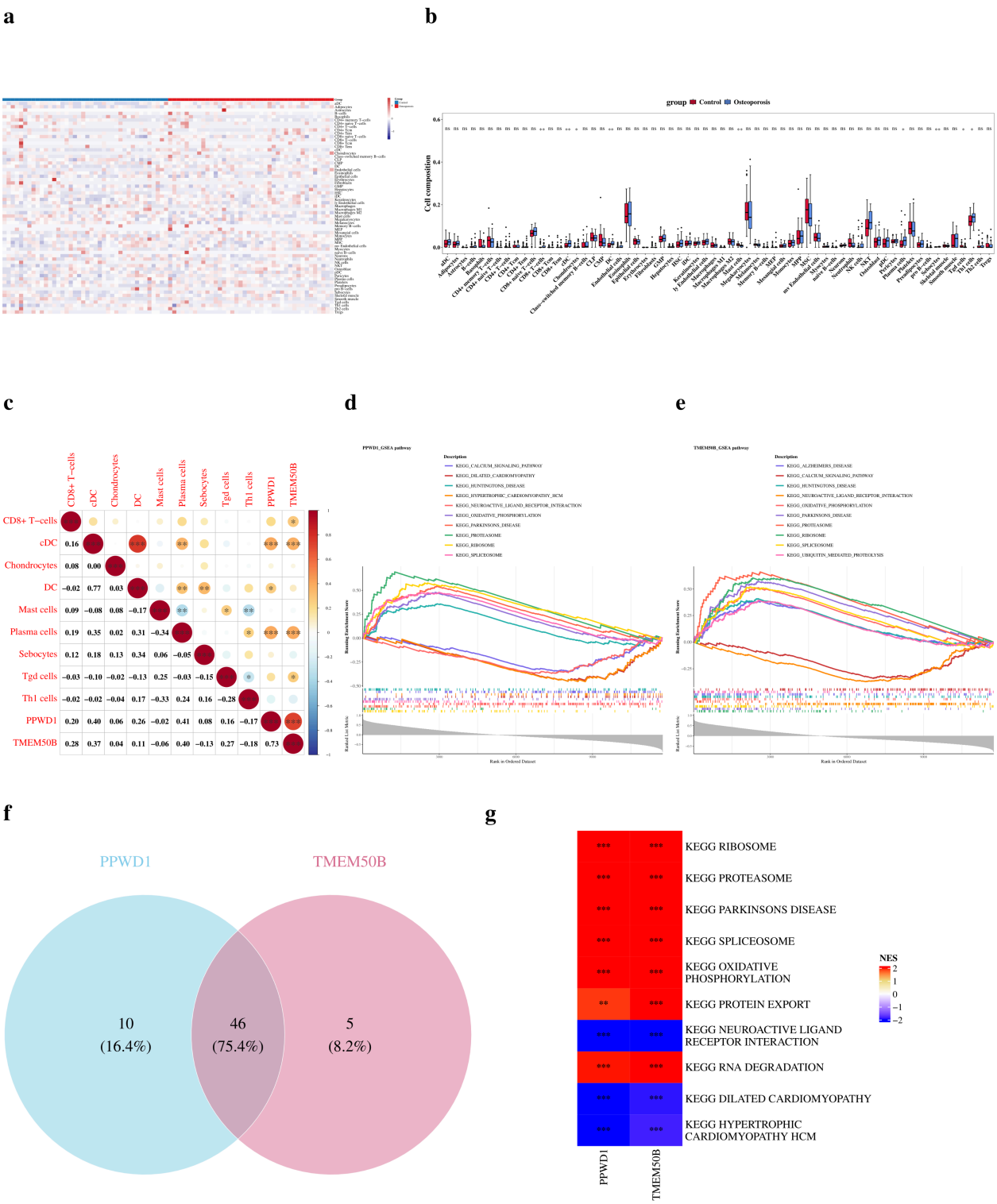


Figure 4

Immune infiltration and gene set enrichment analysis (GSEA) of key genes. (a) Heatmap of cell infiltration. (b) Immune cell types with significant abundance differences between the OP and control groups. (c) Correlation heatmap of differentially abundant immune cells and key genes (* P < 0.05, ** P < 0.01, *** P < 0.001; NS, not significant). (d) GSEA enrichment plot for PPWD1. (e) GSEA enrichment plot

for TMEM50B. (f) Venn diagram of commonly enriched KEGG pathways. (g) Normalized enrichment score (NES) heatmap of shared pathways

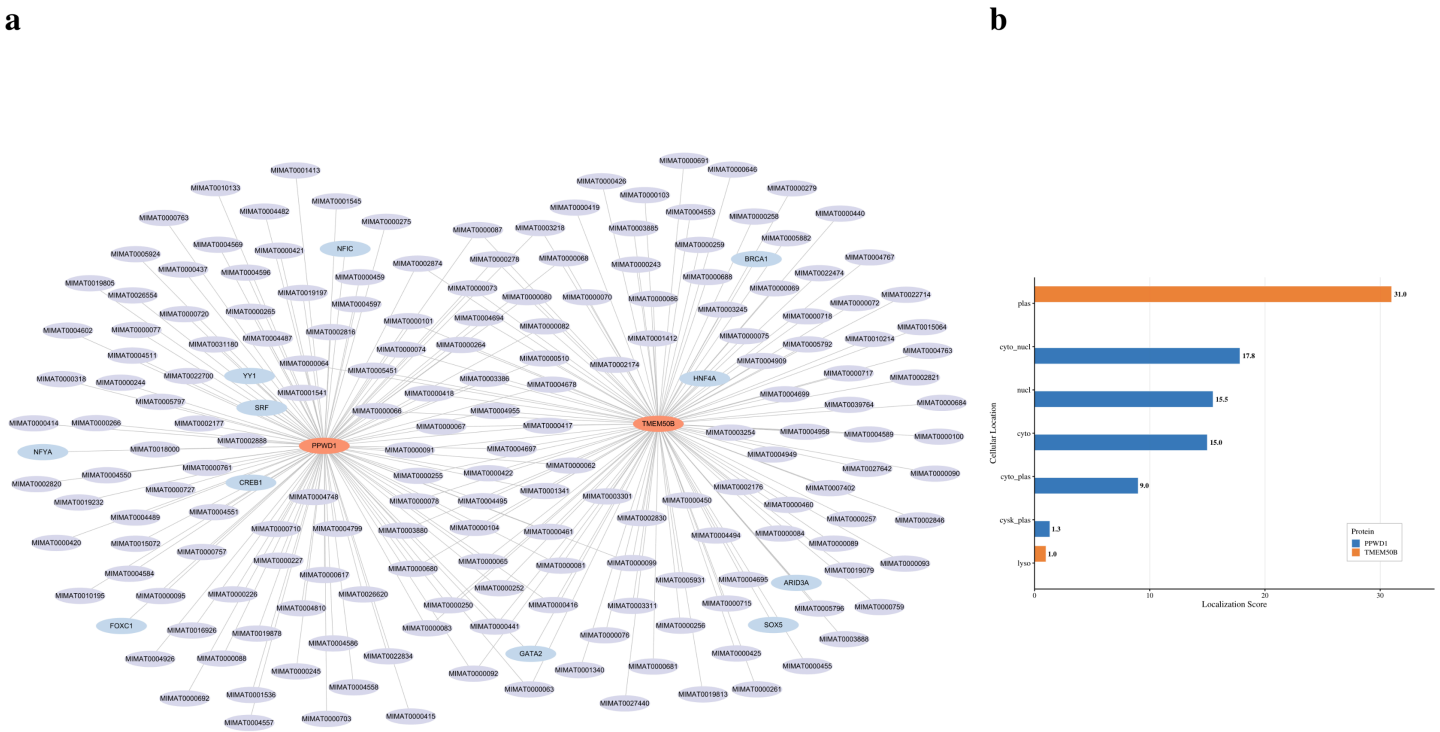


Figure 5

Molecular regulatory network and subcellular localization of key genes. (a) The transcription factor (TF)-gene-microRNA (miRNA) coregulatory network of the key genes. (b) Subcellular localization prediction of PPWD1 and TMEM50B proteins

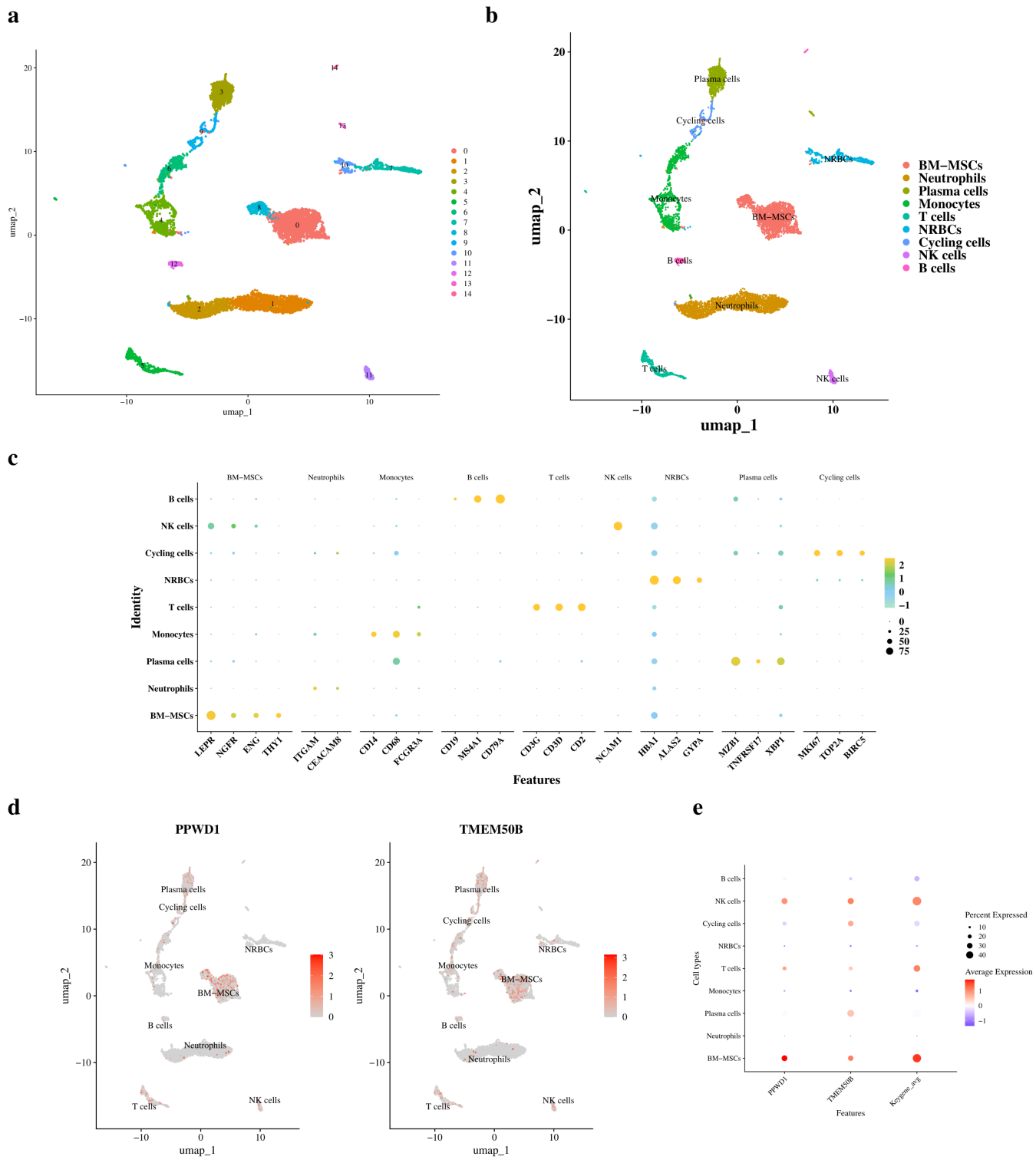


Figure 6

Single-cell transcriptomic landscape and identification of bone marrow mesenchymal stem cells (BM-MSCs) as the key cell type. (a) Uniform manifold approximation and projection (UMAP) plot of cell clusters. (b) UMAP visualization of major cell types. (c) Bubble plot of marker gene expression across the annotated cell types. (d) Expression distribution of PPWD1 and TMEM50B across the nine cell types. (e) Dot plot of PPWD1 and TMEM50B expression across cell types Pathway activity scores were

compared across cell types to assess functional heterogeneity. The 10 pathways with the largest between-cell-type differences included MGMT-mediated DNA damage repair and intestinal lipid absorption, both of which showed particularly high inferred activity in BM-MSCs. In contrast, TASK-type acid-sensitive potassium channel, GABA reuptake, and PMN phagocytosis pathways showed lower activity in this population (**Figure 7g**). Within the analyzed dataset, these patterns indicated comparatively greater inferred DNA damage-repair and lipid-metabolism activity in BM-MSCs.

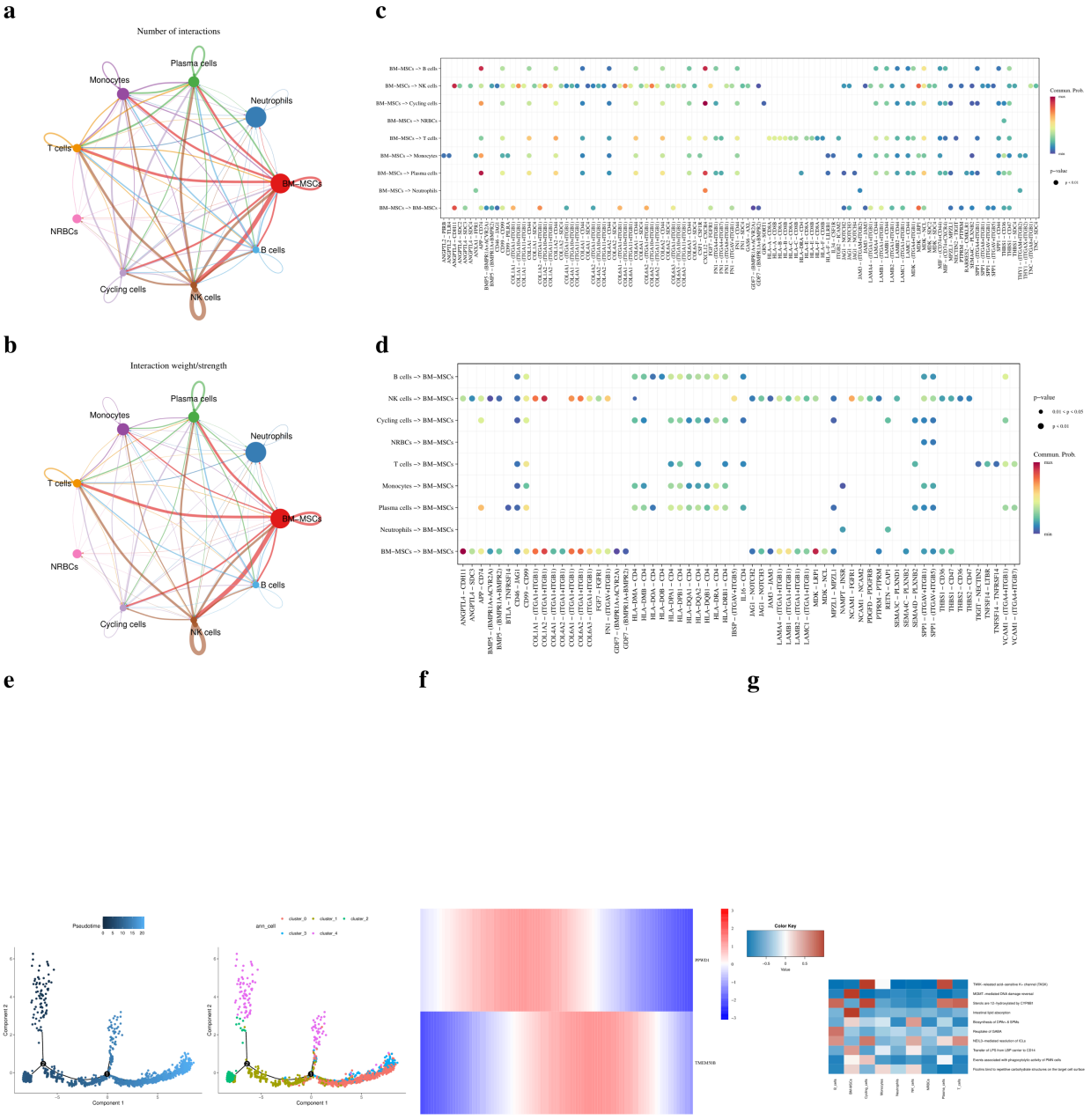


Figure 7

Intercellular communication and differentiation dynamics of BM-MSCs. (a) Cell-cell communication network showing interaction counts. (b) Cell-cell communication network showing interaction strengths. (c) Ligand-receptor interactions with BM-MSCs as signal senders. (d) Ligand-receptor interactions with BM-MSCs as signal recipients. (e) Pseudotime trajectory of BM-MSCs. (f) Expression profiles of PPWD1 and TMEM50B along pseudotime. (g) Pathway activity heatmap of the top 10 differentially active pathways across cell types

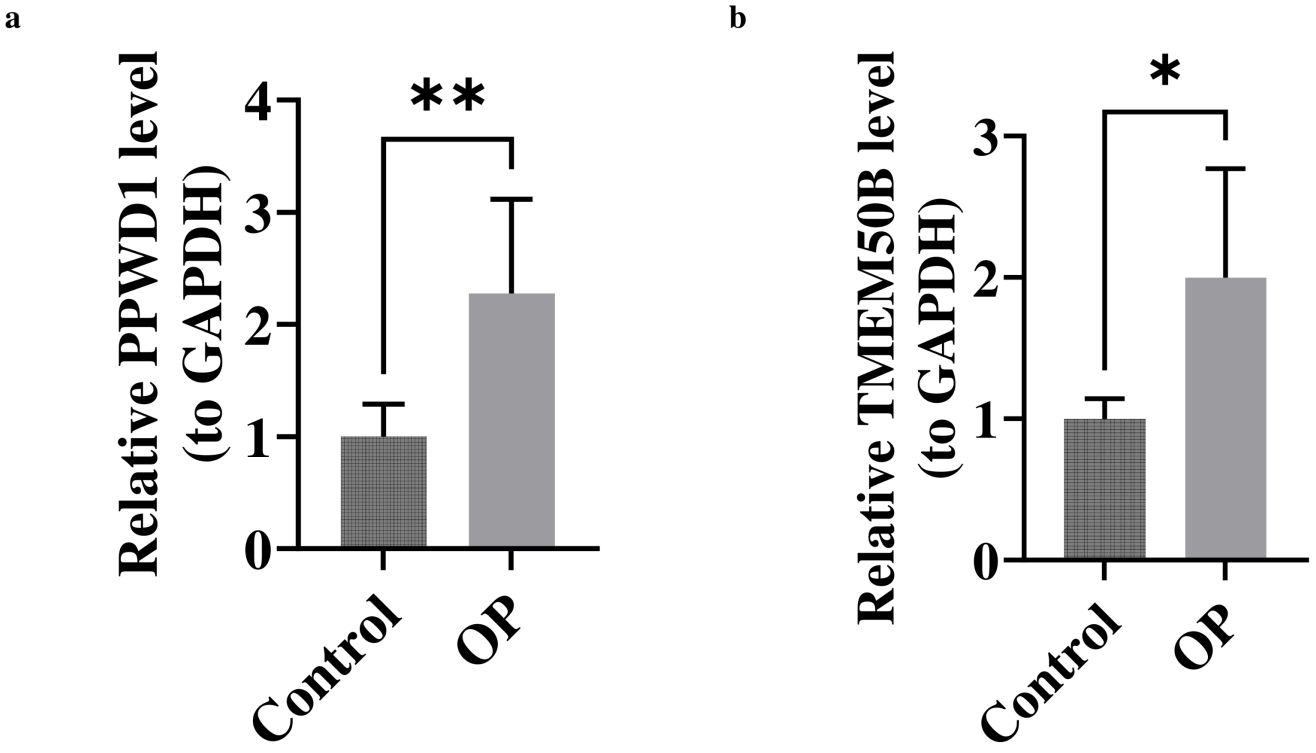


Figure 8

Relative expression levels of (a) PPWD1 and (b) TMEM50B in the OP and control groups