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.

# Integrated multi-omics and machine learning analysis identifies CDK1 as a key regulator of Dieda Qili Tablet-mediated osteogenic differentiation and immune microenvironment remodeling

**Shulong Sun**

dr.sun2012@hotmail.com

Henan University of Chinese Medicine Henan Province Hospital of Traditional Chinese Medicine

**Hua Li**

Henan University of Chinese Medicine Henan Province Hospital of Traditional Chinese Medicine

**Dongfang Li**

Henan University of Chinese Medicine Henan Province Hospital of Traditional Chinese Medicine

**Yujing Cao**

Henan University of Chinese Medicine Henan Province Hospital of Traditional Chinese Medicine

**Ji Li**

Hebei University of Chinese Medicine

**Research Article**

**Keywords:** fracture healing, hub targets, osteoblast, cell cycle, machine learning

**Posted Date:** September 22nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10907212/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

Dieda Qili Tablet (DDQLT), a traditional Chinese medicine formula, has been clinically applied for fracture treatment; however, its molecular mechanisms underlying osteogenic effects remain unclear. This study aimed to elucidate the potential mechanisms of DDQLT in regulating osteoblast differentiation using an integrated bioinformatics and computational framework. Active compounds and potential targets of DDQLT were identified through network pharmacology. Transcriptomic datasets of MC3T3-E1 osteoblast differentiation were integrated for differential expression analysis and weighted gene co-expression network analysis (WGCNA). Candidate genes were screened by intersecting DDQLT-related targets with differentially expressed genes and WGCNA-derived hub genes. Protein–protein interaction (PPI) analysis combined with LASSO, random forest (RF), and support vector machine (SVM) algorithms was performed to identify key regulatory genes. Candidate targets were further validated using the independent human osteoblast dataset GSE157322. Molecular docking and immune infiltration analyses were conducted to explore potential mechanisms. A total of 127 active compounds and 780 potential targets of DDQLT were identified. Transcriptomic analysis revealed 1,165 differentially expressed genes, while WGCNA identified four osteogenesis-associated modules containing 5,628 genes. Functional enrichment analysis demonstrated that DDQLT-associated targets were primarily involved in cell-cycle regulation and osteogenic processes. CDK1 was consistently identified as a key hub gene across multiple approaches and was significantly upregulated during human osteoblast differentiation in the GSE157322 dataset. Molecular docking suggested potential interactions between CDK1 and DDQLT-derived flavonoids, including quercetin and kaempferol. Immune infiltration analysis revealed increased abundance of resting memory $CD4^+$ T cells during osteogenic differentiation. This study suggests that DDQLT may promote osteoblast differentiation through a multi-component regulatory network involving CDK1-associated cell-cycle regulation and osteoimmune microenvironment modulation. These findings provide new insights into the molecular basis of DDQLT-mediated bone regeneration and highlight CDK1 as a potential target for osteogenic therapy.


# Introduction

Fracture is defined as the disruption of the structural continuity of cortical bone, usually caused by traumatic injury, and is commonly accompanied by local swelling, pain, and functional impairment[1]. With the rapid increase in the aging population and the growing prevalence of osteoporosis, the global incidence of fractures has continued to rise. Epidemiological studies indicate that fracture occurrence in Asia and Southeast Asia is increasing rapidly, and approximately half of all fractures worldwide are expected to occur in Asia by 2050[2]. The increasing incidence of fractures, together with prolonged hospitalization and fracture-related complications, has contributed to substantial medical and socioeconomic burdens[3]. Therefore, developing effective strategies to accelerate fracture healing and improve patient outcomes remains an important global healthcare challenge.

Fracture healing is a complex regenerative process involving three overlapping phases: inflammatory hematoma formation, callus formation, and callus remodeling[4,5]. Although uncomplicated fractures

generally heal within approximately 20 weeks[6], this process is influenced by multiple biological factors, including inflammatory regulation, angiogenesis, extracellular matrix formation, and bone remodeling[7,8]. The successful repair of bone tissue requires coordinated interactions among immune cells, vascular cells, fibroblasts, chondrocytes, and osteoblasts[9]. Among these cell populations, osteoblasts play a central role in bone formation by synthesizing extracellular matrix and promoting mineral deposition[10]. Previous studies have demonstrated that osteoblast proliferation, migration, differentiation, and mineralization are essential processes during intramembranous ossification and fracture repair[11–14]. Therefore, enhancing osteoblast activity represents a promising therapeutic strategy for improving bone regeneration.

Increasing evidence indicates that osteoblast differentiation is regulated by complex molecular networks involving transcription factors, signaling pathways, cell-cycle regulators, and interactions with the surrounding microenvironment[15]. Appropriate control of osteoblast proliferation and cell-cycle progression is particularly important because sufficient numbers of functional osteoblasts are required for matrix production and subsequent mineralization. In addition, recent advances in osteoimmunology have revealed that immune cells and inflammatory mediators actively participate in bone regeneration by regulating osteoblast and osteoclast activity. Thus, a comprehensive understanding of the molecular mechanisms controlling osteoblast differentiation and immune regulation may provide new insights into fracture repair.

Traditional Chinese Medicine (TCM) has been widely used for fracture treatment due to its advantages in reducing inflammation, alleviating pain, and promoting tissue repair. Historical records of fracture treatment in TCM can be traced back to ancient medical texts, and Dieda Qili Tablets (DDQLT) represents one of the classic prescriptions clinically applied for fracture and traumatic injury management. DDQLT is derived from the traditional formula “Qili Powder” and consists of multiple herbal components, including Olibanum, Myrrha, Carthami Flos, Catechu, Borneolum Syntheticum, Sanguis Draconis, Moschus Moschiferus, Panax Notoginseng, Angelicae Sinensis, and Cinnabaris. Through the synergistic actions of multiple components, DDQLT has been reported to promote blood circulation, relieve swelling and pain, and facilitate tissue repair[16]. However, due to its complex composition and multi-component characteristics, the molecular mechanisms underlying the therapeutic effects of DDQLT remain largely unclear[17]. Conventional experimental approaches focusing on individual compounds or targets are insufficient to fully elucidate the integrated regulatory network of complex TCM formulations.

Network pharmacology and bioinformatics approaches provide powerful tools for systematically investigating the mechanisms of multi-component drugs by integrating chemical composition, target prediction, gene expression profiles, and biological pathways[18]. In combination with transcriptomic analysis, weighted gene co-expression network analysis (WGCNA), and machine learning algorithms, these approaches enable more comprehensive identification of key regulatory genes associated with complex biological processes. However, candidate genes obtained from high-throughput analyses often require further prioritization and independent validation to improve reliability. Therefore, integrating

multiple computational strategies may provide a more robust framework for identifying potential therapeutic targets of DDQLT.

To investigate the potential mechanisms of DDQLT in promoting osteogenesis, we selected osteoblast differentiation as the biological focus and analyzed publicly available GEO datasets derived from the MC3T3-E1 osteoblast model. MC3T3-E1 cells are a well-established osteoblast precursor cell line derived from mouse calvaria and have been extensively used to study osteoblast proliferation, differentiation, extracellular matrix maturation, and mineralization[19]. Previous studies have demonstrated that MC3T3-E1 cells provide a reliable in vitro model for evaluating the effects of drugs, biomaterials, and signaling molecules on osteogenic activity and bone formation[20–25]. Therefore, integrating MC3T3-E1 transcriptomic data with advanced computational approaches may provide valuable insights into the molecular mechanisms through which DDQLT regulates osteoblast differentiation.

In this study, we combined transcriptomic analysis, WGCNA, network pharmacology, machine learning-based hub gene identification, molecular docking, and immune infiltration analysis to systematically explore the potential mechanisms of DDQLT in osteogenic differentiation. Through this integrated strategy, we aimed to identify key regulatory genes, clarify the potential roles of active compounds, and investigate the relationship between DDQLT, osteogenic regulation, and immune microenvironment remodeling. This study provides new molecular insights into the traditional application of DDQLT and offers potential targets for improving bone regeneration and fracture healing.

## Materials and methods

# Data sources and selection criteria

The datasets analyzed in this study were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds/)[26]. Dataset selection was performed according to the following inclusion criteria: (1) the dataset was generated from studies investigating differentially expressed genes associated with osteogenic differentiation in the mouse osteoblast cell line MC3T3-E1, with array expression profiles available as the primary data type; and (2) the datasets were publicly accessible and available for independent analysis[27].

## Data preprocessing and dataset integration

R software (version 4.2.2; R Foundation for Statistical Computing, 2025; https://www.r-project.org/) was used for data preprocessing and integration of the GEO datasets. Briefly, the series matrix files and corresponding platform annotation files were downloaded from the GEO database. Gene annotation was performed using the 'GEOquery' package[28], and gene IDs were standardized and merged using the 'dplyr' and 'tidyr' packages[29]. Duplicate entries and missing values were removed using the 'limma' package to

generate the processed expression matrix[30]. The processed datasets were integrated using the 'inner_join' function in 'dplyr'[31]. Batch effects were subsequently corrected using the 'removeBatchEffect' function from the 'limma' package, and data distribution before and after correction was visualized using box plots[32]. Principal component analysis (PCA) was performed using the 'FactoMineR', 'stats', and 'factoextra' packages to assess sample heterogeneity, and hierarchical clustering analysis was conducted using the 'hclust' function to evaluate sample relationships[33].

## Screening of DEGs

DEGs were identified using R software (version 4.2.2) based on the processed expression matrix. Differential expression analysis was performed using the 'limma' package, with $|log2FC| \geq 1$ and adjusted $P < 0.05$ as the screening criteria[34]. Volcano plots, waterfall plots, and heatmaps were generated using the 'ggplot2', 'patchwork', and 'heatmap' packages[35,36]. The top 25 DEGs were selected for visualization and clustering analysis[37].

## WGCNA analysis and gene modules construction

Weighted Gene Co-expression Network Analysis (WGCNA) is a widely used bioinformatics approach for identifying gene co-expression patterns and exploring gene–phenotype associations[38]. In this study, the 'WGCNA' R package was applied to identify candidate hub gene modules. Briefly, data quality control was performed to detect and remove potential outlier samples[39]. The optimal soft-thresholding power was determined using the 'pickSoftThreshold' function based on the scale-free topology criterion, followed by the construction of a Topological Overlap Matrix (TOM)[40,41]. Hierarchical clustering and dynamic tree cutting were subsequently performed based on TOM dissimilarity to identify co-expressed gene modules[42,43].

## Screening of active ingredients and potential targets of Dieda Qili Tablet

The active ingredients of Olibanum (Ob), Myrrh (Mr), Carthami Flos (CF), Catechu (Cc), Borneolum Syntheticum (BS), Angelicae Sinensis (AS), and Panax Notoginseng (PN) were initially retrieved from the Traditional Chinese Medicine Systems Pharmacology (TCMSP) database (https://www.tcmsp-e.com/)[44] using the screening criteria of oral bioavailability (OB) ≥ 30% and drug-likeness (DL) ≥ 0.18[45,46]. For herbs unavailable in the TCMSP database, including Sanguis Draconis (SD), Moschus Moschiferus (MM), and Cinnabaris (Cb), candidate compounds were supplemented using the HERB database (http://herb.ac.cn/)[47]. The SwissADME database (http://www.swissadme.ch/) was further applied to evaluate compound pharmacokinetic properties. Compounds with high gastrointestinal absorption and positive drug-likeness predictions according to the Lipinski, Ghose, Veber, Egan, and Muegge rules were

retained[48]. The Canonical SMILES of the selected compounds were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and submitted to SwissTargetPrediction (http://www.swisstargetprediction.ch/) for target prediction[49,50]. Target prediction was performed with the species set to *Homo sapiens*, and targets with a prediction probability of zero were excluded[51,52]. Finally, a compound–target network was constructed using Cytoscape 3.9.1 (https://cytoscape.org/) to visualize the interactions among Dieda Qili Tablet components, active compounds, and their potential targets[53,54].

# Acquisition of common targets and GO/KEGG enrichment analysis

The intersection between DEG- and WGCNA-derived genes identified in our preliminary study and the targets of active DDQLT ingredients was performed using the online tool jvenn 2.0.2 (https://jvenn.toulouse.inrae.fr/app/index.html), and the shared targets were visualized using Venn diagrams[55]. A regulatory network of common targets was subsequently constructed using Cytoscape 3.9.1. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to explore the biological functions and pathways associated with the common targets[56,57]. Briefly, genes were submitted to the Metascape database (https://metascape.org/gp/index.html#/main/step1) for custom analysis, with the species set to *Homo sapiens*. The enrichment parameters were set as follows: Min Overlap = 3, *P* value cutoff = 0.01, and Min Enrichment = 1.5[58]. The enriched terms were visualized using bar plots, bubble plots, and Sankey diagrams generated by the online visualization platform Bioinformatics (https://www.bioinformatics.com.cn/)[59]. Finally, a drug–ingredient–target–pathway interaction network was constructed using Cytoscape 3.9.1.

## Hub target identification based on protein–protein interaction (PPI) network analysis

The common targets were imported into the STRING database (https://cn.string-db.org/) to construct a PPI network using the following parameters: multiple proteins as the search type and a high-confidence interaction score (> 0.700)[60]. The network was further visualized and optimized using Cytoscape 3.9.1. The CytoHubba plugin was then applied to calculate four topological parameters, including Degree, Closeness, Betweenness, and MCC[61,62]. The top 20 targets ranked by each algorithm were selected, and their intersection was defined as the final set of hub targets.

## Hub target identification using machine learning models

To further identify hub targets from the preliminary PPI network analysis, three robust machine learning models were developed using the R packages 'glmnet', 'DALEX', 'randomForest', and 'e1071'[63]. The

workflow included data preprocessing, model construction, optimization, and performance evaluation. Briefly, the common targets identified in the initial analysis were cleaned and randomly divided into training (70%) and testing (30%) sets using the 'caret' package[64]. Model training and optimization were performed on the training set, with 10-fold cross-validation applied to improve model reliability. Model performance was assessed using residual distributions, defined as the difference between observed and predicted probabilities, with the 'DALEX' package used for model interpretation and evaluation[65–67]. For the Least Absolute Shrinkage and Selection Operator (LASSO) model, the penalty parameter (λ) was optimized using 10-fold cross-validation implemented in the 'glmnet' package. The optimal λ value was selected, and genes with non-zero coefficients were considered candidate core targets[68]. For the Support Vector Machine (SVM) model, feature importance was evaluated based on prediction performance, and 10-fold resampling was applied to improve model stability. The most informative feature genes were subsequently selected[69]. A random forest (RF) model was constructed using the 'randomForest' package with 500 decision trees[70]. Gene importance was assessed using the Gini impurity index, and the top-ranked genes were selected as candidate targets[71]. Finally, the candidate targets identified by the three machine learning approaches were intersected to determine the final hub targets, and a Venn diagram was generated to visualize the overlap[72].

# Independent dataset validation of the identified hub gene

To further assess the robustness of the identified hub gene, an independent human osteoblast dataset was retrieved from the GEO database for cross-dataset validation[73]. This dataset included gene expression profiles from osteogenic differentiation and normal control samples[74]. The gene expression profiles were normalized to reduce potential batch effects and technical variations among samples[75]. The distribution of gene expression levels before and after normalization was visualized using box plots to evaluate the effectiveness of data preprocessing[76]. Subsequently, differential expression analysis was performed between the osteogenic differentiation and control groups, and volcano plots were generated to illustrate the global distribution of DEGs[77]. The expression pattern of the hub gene was further examined by comparing its expression levels between the two groups[78]. Statistical significance was evaluated using the Wilcoxon rank-sum test, with $P < 0.05$ considered statistically significant[79,80].

# Molecular docking validation of hub targets

A Sankey diagram was first constructed to visualize the relationships between the preliminarily identified hub targets and their corresponding active compounds. Subsequently, a compound-target network was established using Cytoscape, and the Degree values of compounds were calculated using the CytoHubba plugin to identify key ingredients. The InChIKey identifiers of selected compounds were obtained from the TCMSP database, and their 3D structures were downloaded from the PubChem database and saved in SDF format[81]. Target UniProt Entry IDs were retrieved from the UniProt database (https://www.uniprot.org/) and used to obtain corresponding protein structures from the PDB database (https://www.rcsb.org/), which were saved in PDB format[82]. Finally, ligand and receptor structures were

uploaded to the CB-Dock2 platform (https://cadd.labshare.cn/cb-dock2/) for blind molecular docking analysis, and the binding mode with the lowest binding energy was selected as the optimal docking result[83].

# Immune infiltration analysis of common targets

To investigate the immune microenvironment associated with DDQLT-induced osteogenic differentiation, immune cell infiltration profiles were characterized using the CIBERSORT algorithm (v1.0) based on the common targets shared by DDQLT and osteoblast differentiation[84]. The relative proportions of 22 immune cell subsets were estimated, and samples with valid deconvolution results were retained[85]. Differences in immune cell infiltration between groups were evaluated using the Wilcoxon rank-sum test ($P < 0.05$)[86,87]. Immune cell infiltration profiles were visualized using stacked bar plots and boxplots. Statistical analyses were performed in R using the 'CIBERSORT', 'dplyr', 'ggplot2', and 'ggpubr' packages[88–90].

## Results

# Dataset collection and preprocessing

Based on the predefined inclusion criteria, three GEO datasets (GSE66416, GSE46400, and GSE273620) were selected for analysis. These datasets, containing 2, 6, and 4 samples, respectively, consist of matrix files of differentially expressed genes associated with osteogenic differentiation in the mouse osteoblast cell line MC3T3-E1[91]. The corresponding expression matrix files and platform annotation files (GPL16570, GPL6246, and GPL1261) were downloaded from the GEO database. Data preprocessing was performed using R software (version 4.2.2), generating processed expression matrices (Tables 1–3). The datasets were subsequently integrated into a combined matrix dataset, followed by normalization, heterogeneity assessment, and batch effect correction (Table 4). The effectiveness of data processing was evaluated using boxplots, hierarchical clustering dendrograms, and principal component analysis (PCA) plots (Figs. 1a–c). After correction, samples from different datasets showed improved clustering patterns without evident batch effects, indicating the reliability and comparability of the integrated dataset.

# Identification of DEGs in the dataset

Compared with the normal group, the differentiated group exhibited 1,165 DEGs, with 713 genes upregulated and 452 genes downregulated (Fig. 2a and Table 5). The MA plot further illustrated the distribution of DEGs according to their average expression levels and fold changes (Fig. 2b). A waterfall plot highlighted the top 10 ranked DEGs, including five upregulated genes (IBSP, CDK1, F13A1, APOD, and MMP13) and five downregulated genes (CENPE, MKI67, HS6ST2, CCNB1, and KIF15) (Fig. 2c). A circular heatmap was generated to visualize the expression patterns of representative DEGs, revealing distinct expression profiles between the differentiated and normal groups (Fig. 2d). Overall, these findings reveal

extensive transcriptional remodeling during differentiation and identify candidate genes for subsequent functional enrichment and machine learning-based hub gene analysis.

# Construction of the WGCNA network and identification of core gene modules

The merged dataset was analyzed using the R package 'WGCNA'. Sample clustering was performed with a threshold height of 95, and no outlier samples were identified (Fig. 3a). The optimal soft-thresholding power was determined using the 'pickSoftThreshold' function. A soft-threshold power of 8 was selected, as it achieved a scale-free topology fit index close to 0.90 and resulted in stable average connectivity, indicating that the constructed network approximated a scale-free topology (Fig. 3b). Based on the topological overlap matrix (TOM) dissimilarity, candidate genes were hierarchically clustered according to their expression similarity. Using the dynamic tree cut algorithm with minModuleSize = 30 and mergeCutHeight = 0.25, 16,234 candidate genes were assigned to 56 co-expression modules (Fig. 3c-1,2). Highly similar modules were subsequently merged based on eigengene similarity. A TOM-based network heatmap was generated using 400 representative genes to visualize gene connectivity (Fig. 3d). Finally, module–trait correlation analysis identified four modules significantly associated with the studied phenotype, including the darkgrey module (r = 0.78, *P* = 0.003), darkred module (r = 0.74, *P* = 0.005), grey60 module (r = 0.66, *P* = 0.020), and turquoise module (r = 0.81, *P* = 0.002) (Fig. 3e and Tables 6–9). These four modules were combined, resulting in a total of 5,628 genes identified as core module genes for subsequent analyses.

# Identification of active ingredients and targets of Dieda Qili Tablet

A total of 99 active ingredients of Dieda Qili Tablet were obtained from the TCMSP database, including 8 from Olibanum, 45 from Myrrha, 22 from Carthami Flos, 11 from Catechu, 3 from Borneolum Syntheticum, 2 from Angelicae Sinensis, and 8 from Panax Notoginseng (Table 10). In addition, HERB and SwissADME databases were used to identify 28 additional active ingredients, including 7 from Sanguis Draconis and 21 from Moschus Moschiferus (Table 11). After data integration and deduplication, 127 active ingredients associated with Dieda Qili Tablet were identified. Subsequently, 780 potential targets were obtained after target prediction and redundancy removal. The targets included 160 for Olibanum (Ob), 508 for Myrrha (Mr), 417 for Carthami Flos (CF), 229 for Catechu (Cc), 202 for Borneolum Syntheticum (BS), 303 for Sanguis Draconis (SD), 308 for Moschus Moschiferus (MM), 46 for Angelicae Sinensis (AS), and 254 for Panax Notoginseng (PN), whereas no targets were identified for Cinnabaris. The drug–ingredient–target interaction network was constructed using Cytoscape (Fig. 4a), comprising 913 nodes and 6801 edges, including 9 drugs, 106 active ingredients, and 780 targets. Network analysis revealed four major overlapping active ingredients (MOL000422, MOL000449, MOL000098, and MOL000358). Among them, kaempferol, quercetin, β-sitosterol, and stigmasterol were

identified as potential core active ingredients of Dieda Qili Tablet, interacting with 100, 100, 44, and 41 targets, respectively (Fig. 4b).

# GO and KEGG enrichment analysis of common targets

By integrating the DEGs and WGCNA-derived genes with DDQLT-related targets, 68 common targets were identified. Their distribution and interactions are presented in the Venn diagram and network plot (Figs. 5a and b). GO enrichment analysis showed that these targets were mainly enriched in 62 cellular components (CC), including the chromosomal region, extracellular matrix, and nuclear chromosome. For molecular functions (MF), 54 terms were identified, with predominant enrichment in protein kinase activity, serine-type peptidase activity, and kinase binding. In addition, 370 biological processes (BP) were enriched, mainly involving positive regulation of cell migration, protein phosphorylation, cell cycle phase transition, and positive regulation of the cell cycle (Table 12). The top 10 enriched terms in each GO category are visualized using circular and bar plots (Figs. 5c and d). KEGG enrichment analysis identified 42 significantly enriched pathways, mainly including the cell cycle, PI3K-Akt signaling pathway, and complement and coagulation cascades (Table 13). The relationships between pathways and target genes were illustrated using a Sankey diagram (Fig. 5e). Furthermore, a drug–component–target–pathway network was constructed using Cytoscape 3.9.1 (Fig. 5f).

# Identification of hub targets based on the PPI network

A total of 68 common targets were submitted to the STRING database to construct a PPI network. The network consisted of 68 nodes and 261 edges, with an average node degree of 7.68, a clustering coefficient of 0.589, and significant PPI enrichment ($p < 1.0 \times 10^{-16}$) (Fig. 6a). The PPI network was visualized using Cytoscape, and the CytoHubba plugin was applied to evaluate the topological characteristics of each target based on Degree, MCC, Closeness, and Betweenness centrality (Fig. 6b and Table 14). The top 20 targets were subsequently selected for integrative analysis, including CDK1, CCNA2, AURKB, EZH2, PARP1, MCL1, and DNMT1 (Fig. 6c).

# The hub targets identified by machine learning

To further identify hub genes from the PPI network, three machine learning models (LASSO, RF, and SVM) were constructed. The performance of these models was initially evaluated by comparing residual distributions. Overall, the three models showed comparable predictive performance. However, the reverse cumulative distribution analysis of residuals indicated that the SVM model exhibited a greater proportion of low residual values in the left tail, whereas the LASSO model showed more high residual values in the right tail of the distribution (Fig. 7a). Among the three models, the RF model achieved the lowest median absolute residual value, suggesting superior predictive performance based on this metric (Fig. 7b). Using the LASSO model, five hub genes were selected based on the optimal λ value (Fig. 7c and Table 15). The SVM model identified the five most significant genes according to the lowest classification error rate and highest average ranking (Fig. 7d and Table 16). In addition, the RF model was

used to calculate gene importance based on 500 decision trees, and the five most important genes were selected (Fig. 7e, f and Table 17). Finally, CDK1 was identified as the hub gene through the intersection of candidate genes obtained from the three machine learning approaches (Fig. 7g).

# Independent dataset validation of CDK1 expression in human osteoblasts

To independently validate CDK1 as a potential hub gene, the GSE157322 human osteoblast dataset was analyzed (Table 18). After normalization, sample expression profiles showed improved consistency and comparability, supporting the reliability of subsequent analyses (Fig. 8a). Differential expression analysis revealed distinct transcriptional profiles between the osteogenic differentiation and control groups (Table 19). The volcano plot illustrated widespread transcriptional changes, with 130 genes upregulated and 84 genes downregulated (Fig. 8b). Further ranking analysis of DEGs identified CDK1 as one of the top-ranked differentially expressed genes, highlighting its potential involvement in the regulation of osteogenic differentiation (Fig. 8c). Consistent with the findings from the discovery dataset, CDK1 expression was significantly increased in osteogenic differentiation samples compared with controls ($P = 0.0033$; Fig. 8d). These results validated the robust expression pattern of CDK1 in an independent dataset, highlighting its potential significance as a key regulatory gene in osteogenic differentiation.

# Molecular docking

Based on the hub target CDK1 identified in the previous analysis, 17 corresponding compounds were screened (Fig. 9a). Network analysis combined with the CytoHubba plugin further identified key compounds with Degree > 1, including kaempferol (MOL000422) and quercetin (MOL000098) (Fig. 9b and Table 20). Molecular docking was subsequently performed between CDK1 and these two compounds using the CB-Dock database. Five potential binding pockets (C1–C5) were identified for each docking group. The results indicated that the most favorable binding pocket for both CDK1-kaempferol and CDK1-quercetin interactions was located in the C2 region. The lowest binding energy was − 8.0 kcal/mol for CDK1-kaempferol (Fig. 9c and Table 21) and − 8.2 kcal/mol for CDK1-quercetin (Fig. 9d and Table 22), suggesting strong binding affinity between CDK1 and these compounds.

# Immune infiltration landscape of common targets

To further investigate the potential involvement of immune regulation in osteogenic differentiation, we characterized the immune landscape associated with the 68 common targets between DDQLT and osteoblast differentiation. These targets exhibited distinct expression patterns between the osteogenic differentiation and control groups (Fig. 10a and Table 23), suggesting their potential regulatory roles in osteogenesis. Immune deconvolution analysis using CIBERSORT revealed marked heterogeneity in the infiltration profiles of 22 immune cell subsets among samples (Fig. 10b and Table 24), suggesting dynamic remodeling of the immune microenvironment during osteoblast differentiation. Comparative analysis further demonstrated distinct immune cell composition patterns between the differentiation

and control groups (Fig. 10c). Notably, resting memory CD4⁺ T cells exhibited a significant increase in the differentiation group compared with the control group ($P = 0.017$; Fig. 10d), indicating their potential involvement in the immune regulation associated with osteoblast differentiation.

# Discussion

Bone fracture repair is a highly coordinated biological process involving inflammation, angiogenesis, osteoblast proliferation, extracellular matrix deposition, and mineralization[92]. With the rapid growth of the aging population, osteoporosis-related fractures have become a major public health challenge, resulting in substantial medical and socioeconomic burdens[93]. Therefore, identifying safe and effective strategies to enhance bone regeneration remains an important clinical goal. Bone tissue is continuously remodeled through the coordinated activities of osteoblasts, osteocytes, and osteoclasts, maintaining skeletal integrity and mineral homeostasis[94,95]. Among these cell populations, osteoblasts are the primary bone-forming cells responsible for extracellular matrix synthesis and mineralization[96], and their dysfunction contributes directly to impaired fracture healing and osteoporosis progression[97–100]. Although DDQLT has been clinically used for many years to promote fracture healing and alleviate soft tissue injury, its underlying molecular mechanisms remain incompletely understood. In this study, we integrated network pharmacology, transcriptomic analysis, WGCNA, machine learning, molecular docking, and immune infiltration analysis to systematically investigate the potential mechanisms of DDQLT during osteoblast differentiation. Our findings identified CDK1 as a key regulatory gene and suggested that quercetin and kaempferol may represent important active compounds involved in DDQLT-mediated osteogenic regulation.

Human and mouse osteoblasts share highly conserved molecular characteristics during skeletal development and bone repair[101]. Therefore, three GEO datasets derived from the mouse osteoblast precursor cell line MC3T3-E1 were selected for transcriptomic analysis. MC3T3-E1 cells are widely used as an established in vitro model for studying osteoblast proliferation, matrix maturation, and mineralization, and previous studies have demonstrated coordinated activation of proliferation-associated and osteogenesis-associated genes during differentiation[102]. After integration, normalization, and batch effect correction, the combined dataset exhibited improved sample clustering and reduced technical heterogeneity, supporting the reliability of subsequent analyses. Differential expression analysis identified 1,165 DEGs between osteogenic differentiation and control groups, indicating extensive transcriptional remodeling during osteoblast maturation. Furthermore, WGCNA identified four phenotype-associated modules containing 5,628 core genes, providing a systematic representation of gene networks associated with osteogenic differentiation.

Network pharmacology analysis revealed 127 active compounds and 780 potential targets associated with DDQLT, supporting the characteristic multi-component and multi-target properties of traditional Chinese medicine formulations[103]. Among these compounds, kaempferol, quercetin, β-sitosterol, and stigmasterol exhibited high connectivity within the drug–component–target network, suggesting that

these compounds may contribute significantly to the biological effects of DDQLT. Previous studies have demonstrated that quercetin possesses antioxidant, anti-inflammatory, and osteoprotective activities, including enhancement of osteogenic differentiation and regulation of osteoblast-related markers[104]. Similarly, kaempferol has been reported to regulate inflammatory responses and oxidative stress pathways, which are closely associated with bone metabolism and regeneration[105]. Among the constituent herbs of DDQLT, Myrrha showed the highest number of predicted compounds and targets, suggesting a potentially important pharmacological contribution. Previous studies have demonstrated that Myrrha exhibits anti-inflammatory, analgesic, and immunomodulatory activities through pathways including TNF and Toll-like receptor signaling[106–108]. By integrating DDQLT-associated targets with osteogenesis-related DEGs and WGCNA-derived genes, 68 common targets were identified, providing candidate molecular mechanisms underlying the osteogenic effects of DDQLT.

Functional enrichment analysis demonstrated that these common targets were mainly associated with chromosomal regions, extracellular matrix components, protein kinase activity, cell migration, and cell-cycle regulation. KEGG analysis further revealed that the cell cycle pathway was among the most significantly enriched pathways, suggesting that DDQLT may regulate osteoblast differentiation through modulation of cell proliferation and cell-cycle progression. The cell cycle is a fundamental biological process involving sequential transitions through G1, S, G2, and M phases, which are tightly regulated to maintain cellular proliferation and differentiation[109,110]. During fracture healing, sufficient osteoblast proliferation is required to increase the number of bone-forming cells and support extracellular matrix production and mineralization[111]. Therefore, regulation of cell-cycle-related genes may represent an essential mechanism through which DDQLT facilitates osteogenic differentiation.

Cell-cycle progression is primarily controlled by cyclin-dependent kinases (CDKs), which regulate key transition points through phosphorylation-dependent signaling[112–114]. Among these regulators, CDK1 is considered the central kinase controlling mitotic entry and G2/M transition. CDK1 forms complexes with cyclin B to generate maturation-promoting factor (MPF), which regulates chromosome segregation, centrosome duplication, and mitotic progression[115,116]. Activation of cyclin B-CDK1 promotes G2-to-M transition, whereas its precise regulation by WEE1, MYT1, and CDC25 ensures appropriate timing of mitosis[117–119]. Beyond its canonical role in cell-cycle regulation, CDK1 has also been implicated in transcriptional regulation, mitochondrial metabolism, cellular adhesion, and stem cell differentiation[120–125]. Therefore, CDK1 may serve as a molecular bridge linking osteoblast proliferation, differentiation, and metabolic adaptation.

To improve the reliability of hub gene identification, this study combined PPI network analysis with three machine learning algorithms, including LASSO, random forest (RF), and support vector machine (SVM). Machine learning-based feature selection has been increasingly applied in biomedical research because it can identify stable predictive features from high-dimensional biological datasets while reducing potential bias associated with single statistical approaches[126]. Although the three models demonstrated different predictive characteristics, including lower median residual values in RF and a higher proportion

of low-residual predictions in SVM, all algorithms consistently identified CDK1 as the overlapping candidate gene. This multi-model consensus strategy improves confidence in candidate selection and reduces the possibility of false-positive discoveries.

Furthermore, the reliability of CDK1 identification was strengthened by independent validation using the human osteoblast dataset GSE157322. Transcriptomic biomarkers identified from a single cohort may be influenced by platform-specific variation, batch effects, and biological heterogeneity, which may limit their reproducibility and translational value[127]. Therefore, external validation using independent datasets represents an important strategy for evaluating the robustness of candidate biomarkers[128]. In the present study, CDK1 expression was significantly elevated during human osteoblast differentiation compared with control samples ($P = 0.0033$), and CDK1 ranked among the top differentially expressed genes. This expression pattern was consistent with the findings obtained from the integrated mouse MC3T3-E1 datasets, suggesting that CDK1 activation during osteogenic differentiation may represent a conserved regulatory event across species. These results indicate that the combination of machine learning-based screening and independent human validation provides stronger evidence supporting CDK1 as a potential biomarker associated with osteoblast differentiation.

To further explore the molecular basis by which DDQLT may regulate CDK1, molecular docking analysis was performed between CDK1 and key active compounds identified through network pharmacology. Kaempferol and quercetin exhibited favorable binding affinities toward CDK1, with docking energies of −8.0 kcal/mol and − 8.2 kcal/mol, respectively. Previous studies have demonstrated that quercetin can promote osteogenic differentiation by enhancing the expression of osteogenesis-related factors, including BMP2 and osteopontin, and improve bone formation in animal models[104]. In addition, flavonoids such as kaempferol have demonstrated protective effects on bone metabolism through regulation of oxidative stress, inflammatory responses, and osteoblast activity[105]. The favorable binding energies observed in the present study suggest that these compounds may interact stably with CDK1 and potentially contribute to the pharmacological effects of DDQLT. However, molecular docking mainly predicts possible ligand–protein interactions based on computational algorithms and cannot fully represent the dynamic intracellular environment. Therefore, further experimental validation, including surface plasmon resonance, enzyme activity assays, and molecular dynamics simulations, is required to confirm the direct interaction and functional consequences of CDK1 modulation by these compounds[129].

In addition to regulating intrinsic osteoblast functions, increasing evidence indicates that immune regulation plays a critical role in bone regeneration and fracture repair. Bone healing is not only dependent on osteogenic cells but also involves complex interactions among immune cells, inflammatory mediators, vascular components, and extracellular matrix networks[130]. Immune cells can regulate osteoblast and osteoclast activity through cytokine secretion and cell–cell communication, thereby influencing bone remodeling processes[131]. In this study, immune infiltration analysis based on 68 common targets revealed dynamic changes in the immune microenvironment during osteoblast

differentiation. Among the 22 immune cell populations analyzed using CIBERSORT, resting memory $CD4^+$ T cells showed significantly increased infiltration in the osteogenic differentiation group compared with the control group ($P = 0.017$).

$CD4^+$ T cells have been increasingly recognized as important regulators of skeletal homeostasis. Depending on their activation state and cytokine profile, different $CD4^+$ T-cell subsets can exert either anabolic or catabolic effects on bone metabolism[132]. Regulatory immune responses may promote tissue repair by maintaining an appropriate inflammatory environment, whereas excessive immune activation can impair osteogenesis and enhance bone resorption[133]. The increased proportion of resting memory $CD4^+$ T cells observed in osteogenic samples suggests that DDQLT-associated targets may participate in shaping an immune environment favorable for osteoblast differentiation. Although the precise mechanisms remain unclear, this finding provides a potential link between DDQLT-mediated osteogenesis and osteoimmune regulation. Future studies combining single-cell sequencing, immune cell depletion models, and cytokine profiling will be necessary to clarify the functional role of specific immune populations during DDQLT-mediated bone regeneration.

Collectively, this study provides a comprehensive computational framework integrating network pharmacology, transcriptomics, WGCNA, machine learning, molecular docking, and immune infiltration analysis to elucidate the potential mechanism of DDQLT in promoting osteogenic differentiation[134]. The results suggest that DDQLT may exert osteogenic effects through a multi-target regulatory network involving flavonoid compounds such as quercetin and kaempferol, CDK1-associated cell-cycle regulation, and remodeling of the osteoimmune microenvironment. The identification of CDK1 as a conserved hub gene provides new molecular insights into the traditional application of DDQLT and highlights potential targets for improving bone regeneration.

Several limitations should be considered. First, although multiple bioinformatics approaches and independent human dataset validation were performed, the conclusions remain primarily based on computational analyses and require experimental verification. Functional studies using CDK1 knockdown, overexpression, or pharmacological inhibition models are needed to determine whether CDK1 is essential for the osteogenic effects of DDQLT. Second, molecular docking provides theoretical evidence for compound–target interactions but does not directly confirm biological activity. Additional biochemical and structural studies are required to validate the interactions between CDK1 and DDQLT-derived compounds. Third, DDQLT contains numerous bioactive components, and the synergistic interactions among different compounds remain largely unknown. Fourth, although immune infiltration analysis suggested potential involvement of resting memory $CD4^+$ T cells, the causal relationship between immune modulation and osteoblast differentiation requires further investigation.

Future studies should combine in vitro osteoblast differentiation models, CDK1 functional manipulation, fracture animal models, molecular dynamics simulation, phosphoproteomic analysis, and single-cell transcriptomic approaches to further clarify the regulatory network involving DDQLT, CDK1, immune cells, and osteogenic differentiation. Such investigations may facilitate the translation of DDQLT from

empirical clinical application toward evidence-based therapeutic development for fracture healing and bone regeneration.

## Conclusion

In conclusion, this study established an integrated bioinformatics and computational framework combining network pharmacology, transcriptomics, WGCNA, machine learning, molecular docking, and immune infiltration analysis to elucidate the potential mechanisms of DDQLT in osteogenic differentiation. Our results demonstrated that DDQLT acts through a multi-component and multi-target regulatory network, with CDK1 identified as a potential core target by multiple machine learning algorithms and validated in an independent human osteoblast dataset. Functional analyses suggested that CDK1-associated cell-cycle regulation may represent an important mechanism underlying the osteogenic effects of DDQLT. Furthermore, quercetin and kaempferol were identified as potential active compounds interacting with CDK1, while immune infiltration analysis indicated a possible role of immune microenvironment modulation, particularly involving resting memory $CD4^+$ T cells, in DDQLT-mediated osteogenesis. These findings provide novel molecular insights into the traditional use of DDQLT for fracture healing and highlight potential targets for future experimental validation and bone regeneration therapies.

## Declarations

# Ethics and Consent to Participate declarations

Not applicable

# 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.

# Funding

The author(s) received no specific funding for this work.

# Author Contribution

S.S. wrote the main manuscript text and was responsible for the overall progress of the research project. Y.C. and J.L.carried out comprehensive quality control and supervision for this research project. H.L. conducted the initial data screening and processing, while D.L. optimized the images. All authors reviewed the manuscript.

# Acknowledgement

We would like to thank Yujing Cao and Xiuhuan Wang for the help of the collection and analysis of experimental data related experiments and constructive discussion.


# Data Availability

The datasets analyzed in the study are available in the Gene Expression Omnibus database (https://ncbi.nlm.nih.gov/geo/) with accession codes GSE66416, GSE46400, and GSE273620. The ingredients and targets related to DDQLT can be accessed in the TCMSP (https://www.tcmsp-e.com/), HERB (http://herb.ac.cn/), SwissADME database (http://www.swissadme.ch/) and SwissTargetPrediction database (http://www.swisstargetprediction.ch/). The molecular structures of the ingredients and targets are obtained through the PubChem database(https://pubchem.ncbi.nlm.nih.gov/), UniProt database (https://www.uniprot.org/) and PDB database (https://www.rcsb.org/). Molecular docking is carried out based on the CB Dock database (https://cadd.labshare.cn/cb-dock2/) database. All data generated or analysed during this study are included in this published article and its table files.

# References


1. Poorman GW et al (2019) Traumatic Fracture of the Pediatric Cervical Spine: Etiology, Epidemiology, Concurrent Injuries, and an Analysis of Perioperative Outcomes Using the Kids' Inpatient Database. Int J Spine Surg 13:68–78. 10.14444/6009
2. Cong E, Walker MD (2014) The Chinese skeleton: insights into microstructure that help to explain the epidemiology of fracture. Bone Res 2:14009. 10.1038/boneres.2014.9
3. Guzon-Illescas O et al (2019) Mortality after osteoporotic hip fracture: incidence, trends, and associated factors. J Orthop Surg Res 14:203. 10.1186/s13018-019-1226-6
4. Wang X et al (2013) Role of mesenchymal stem cells in bone regeneration and fracture repair: a review. Int Orthop 37:2491–2498. 10.1007/s00264-013-2059-2
5. Mosekilde L, Bak B (1993) The effects of growth hormone on fracture healing in rats: a histological description. Bone 14:19–27. 10.1016/8756-3282(93)90251-5
6. Littenberg B et al (1998) Closed fractures of the tibial shaft. A meta-analysis of three methods of treatment. J Bone Joint Surg Am Vol 80:174–183. 10.2106/00004623-199802000-00004
7. Fazzalari N (2011) Bone fracture and bone fracture repair. *Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 22, 2003–2006. 10.1007/s00198-011-1611-4
8. Yao J, Xin R, Zhao C, Yu C (2024) MicroRNAs in osteoblast differentiation and fracture healing: From pathogenesis to therapeutic implication. Injury 55:111410. 10.1016/j.injury.2024.111410

9. Wang H et al (2024) Advances in the role and mechanism of fibroblasts in fracture healing. Front Endocrinol (Lausanne) 15:1350958. 10.3389/fendo.2024.1350958
10. Liu D, Zhao X, Zhang Q, Zhou F, Tong X (2024) Bone marrow mesenchymal stem cell-derived exosomes promote osteoblast proliferation, migration and inhibit apoptosis by regulating KLF3-AS1/miR-338-3p. BMC Musculoskelet Disord 25:122. 10.1186/s12891-024-07236-0
11. Carter D, Beaupré G, Giori N, Helms J (1998) Mechanobiology of skeletal regeneration. *Clinical orthopaedics and related research*, S41-55. 10.1097/00003086-199810001-00006
12. Claes L et al (1998) Effects of mechanical factors on the fracture healing process. Clin Orthop Relat Res 132–147. 10.1097/00003086-199810001-00015
13. Le A, Miclau T, Hu D, Helms J (2001) Molecular aspects of healing in stabilized and non-stabilized fractures. J Orthop research: official publication Orthop Res Soc 19:78–84. 10.1016/s0736-0266(00)00006-1
14. Dirckx N, Van Hul M, Maes C (2013) Osteoblast recruitment to sites of bone formation in skeletal development, homeostasis, and regeneration. Birth Defects Res C Embryo Today 99:170–191. 10.1002/bdrc.21047
15. Meng W et al (2025) BDNF improves fracture healing through promoting osteoblasts proliferation and migration via trkb/akt regulated NUCKS1 expression. Sci Rep 15:27107. 10.1038/s41598-025-11594-7
16. Xue-Han HONG (2019) Effect of Qili Powder on Postoperative Swelling and Pain in Patients with Patellar Fracture. Med Inform 32(03):162–164. 10.3969/j.issn.1006-1959.2019.03.054
17. Ma Q et al (2018) Effects of Tao-Hong-Si-Wu decoction on acute blood stasis in rats based on a LC-Q/TOF-MS metabolomics and network approach. Biomedical chromatography: BMC 32. 10.1002/bmc.4144
18. Wu X, Zeng LH, Taniguchi T, Xie QM (2007) Activation of PKA and phosphorylation of sodium-dependent vitamin C transporter 2 by prostaglandin E2 promote osteoblast-like differentiation in MC3T3-E1 cells. Cell Death Differ 14:1792–1801. 10.1038/sj.cdd.4402190
19. Zhu L, Xie L, Wang Z, Li KL, Cai W (2025) Mass spectrometry-based metabolomics reveal the effects and potential mechanism of isochlorogenic acid A in MC3T3-E1 cells. Front Mol Biosci 12:1518873. 10.3389/fmolb.2025.1518873
20. Croucher PI, McDonald MM, Martin TJ (2016) Bone metastasis: the importance of the neighbourhood. Nat Rev Cancer 16:373–386. 10.1038/nrc.2016.44
21. Long F (2011) Building strong bones: molecular regulation of the osteoblast lineage. Nat Rev Mol Cell Biol 13:27–38. 10.1038/nrm3254
22. Lamolle SF et al (2009) The effect of hydrofluoric acid treatment of titanium surface on nanostructural and chemical changes and the growth of MC3T3-E1 cells. Biomaterials 30:736–742. 10.1016/j.biomaterials.2008.10.052
23. Zhao Y, Zou B, Shi Z, Wu Q, Chen GQ (2007) The effect of 3-hydroxybutyrate on the in vitro differentiation of murine osteoblast MC3T3-E1 and in vivo bone formation in ovariectomized rats.

Biomaterials 28:3063–3073. 10.1016/j.biomaterials.2007.03.003

24. Isama K, Tsuchiya T (2003) Enhancing effect of poly(L-lactide) on the differentiation of mouse osteoblast-like MC3T3-E1 cells. Biomaterials 24:3303–3309. 10.1016/s0142-9612(03)00216-3
25. Lazáry A et al (2007) Effect of gypsum on proliferation and differentiation of MC3T3-E1 mouse osteoblastic cells. Biomaterials 28:393–399. 10.1016/j.biomaterials.2006.09.002
26. Clough E et al (2024) NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res 52:D138–d144. 10.1093/nar/gkad965
27. Emily C et al (2023) NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res 52. 10.1093/nar/gkad965
28. Wang S et al (2025) Bioinformatics analysis of oxidative phosphorylation-related differentially expressed genes in osteoporosis. Eur J Med Res 30:294. 10.1186/s40001-025-02568-6
29. Kuang Y et al (2023) Weighted gene coexpression network analysis reveals negative regulation of hypertrophic cardiomyopathy by carboxylesterase 1 and cathepsin C. Gen Physiol Biophys 42:361–372. 10.4149/gpb_2023009
30. Xin P, Pei S, Ma N, Xiao L (2025) PCK1 as a potential hub gene in distinguishing lactate metabolism between rheumatoid arthritis and osteoarthritis. PeerJ 13:e19661. 10.7717/peerj.19661
31. Zhu YX, Huang JQ, Ming YY, Zhuang Z, Xia H (2021) Screening of key biomarkers of tendinopathy based on bioinformatics and machine learning algorithms. PLoS ONE 16:e0259475. 10.1371/journal.pone.0259475
32. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118–127. 10.1093/biostatistics/kxj037
33. Haoru D et al (2021) Integrated bioinformatic analysis reveals the underlying molecular mechanism of and potential drugs for pulmonary arterial hypertension. Aging 13. 10.18632/aging.203040
34. Haichao G et al (2024) Identification of key genes and molecular mechanisms of chronic urticaria based on bioinformatics. Skin Res Technol 30. 10.1111/srt.13624
35. Ast T et al (2024) METTL17 is an Fe-S cluster checkpoint for mitochondrial translation. Mol Cell 84:359–374e358. 10.1016/j.molcel.2023.12.016
36. Li N et al (2025) Potent prophylactic cancer vaccines harnessing surface antigens shared by tumour cells and induced pluripotent stem cells. Nat Biomed Eng 9:215–233. 10.1038/s41551-024-01309-0
37. Emil K, David G, Regina Z, Sonia HR, G.-R., Mina R (2022) ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics* 38. 10.1093/bioinformatics/btac409
38. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9. 10.1186/1471-2105-9-559
39. Xiaoqian L et al (2019) Identification of hub genes and key pathways associated with angioimmunoblastic T-cell lymphoma using weighted gene co-expression network analysis. Cancer Manag Res 11. 10.2147/cmar.S185030

40. Zijian L, Juan F, Zhengwei Y (2020) Key Modules and Hub Genes Identified by Coexpression Network Analysis for Revealing Novel Biomarkers for Spina Bifida. Front Genet 11. 10.3389/fgene.2020.583316
41. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4., Article17, 10.2202/1544-6115.1128
42. Horvath S, Dong J (2008) Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4:e1000117. 10.1371/journal.pcbi.1000117
43. Zhang F et al (2020) Identification of Potential Therapeutic Targets of Alzheimer's Disease By Weighted Gene Co-Expression Network Analysis. Chin Med Sci J 35:330–341. 10.24920/003695
44. Ru J et al (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform 6. 10.1186/1758-2946-6-13
45. Xu X et al (2012) A novel chemometric method for the prediction of human oral bioavailability. Int J Mol Sci 13:6964–6982. 10.3390/ijms13066964
46. Tao W et al (2013) Network pharmacology-based prediction of the active ingredients and potential targets of Chinese herbal Radix Curcumae formula for application to cardiovascular disease. J Ethnopharmacol 145:1–10. 10.1016/j.jep.2012.09.051
47. Fang S et al (2021) HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. Nucleic Acids Res 49:D1197–D1206. 10.1093/nar/gkaa1063
48. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. 10.1038/srep42717
49. Cheng T, Pan Y, Hao M, Wang Y, Bryant S (2014) PubChem applications in drug discovery: a bibliometric analysis. Drug Discovery Today 19:1751–1756. 10.1016/j.drudis.2014.08.008
50. Kim S et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49:D1388–D1395. 10.1093/nar/gkaa971
51. Daina A, Michielin O, Zoete V (2019) SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res 47:W357–W364. 10.1093/nar/gkz382
52. Sun S et al (2026) Exploring the mechanism of Dieda Qili Tablet on fracture healing based on network pharmacology combined with machine learning models. Sci Rep. 10.1038/s41598-026-48781-z
53. Nadezhda T, Yassen D, Francisco A S, D., Mario A (2012) Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc 7. 10.1038/nprot.2012.004
54. Sun S et al (2026) Exploring the mechanism of Dieda Qili Tablet on fracture healing based on network pharmacology combined with machine learning models. Sci Rep 16. 10.1038/s41598-026-48781-z

55. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C (2014) jvenn: an interactive Venn diagram viewer. BMC Bioinformatics 15:293. 10.1186/1471-2105-15-293
56. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. 10.1093/nar/28.1.27
57. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–462. 10.1093/nar/gkv1070
58. Zhou Y et al (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10:1523. 10.1038/s41467-019-09234-6
59. Simon J, Maryam W A, A., Daniel J (2019) A Bioinformatics Toolkit: In Silico Tools and Online Resources for Investigating Genetic Variation. Semin Thromb Hemost 45. 10.1055/s-0039-1692978
60. Damian S et al (2022) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51. 10.1093/nar/gkac1000
61. Zhengfei M, Ping Z, Peidong Y, Zhongwu S (2023) Identification of immune-related molecular markers in intracranial aneurysm (IA) based on machine learning and cytoscape-cytohubba plug-in. BMC Genom Data 24. 10.1186/s12863-023-01121-w
62. Ali A, Mohan J, Nadaf TAA, Ravishankar H, Deepa KR (2024) Bioinformatics-Driven Discovery of Signaling Pathways and Genes Influencing Cervical Cancer. SN Comput Sci 5:989. 10.1007/s42979-024-03347-6
63. Beam AL, Kohane IS (2018) Big Data and Machine Learning in Health Care. JAMA 319:1317–1318. 10.1001/jama.2017.18391
64. Sun S et al (2025) Screening of hub genes and immunocytes related to tendon injury based on bioinformatics and machine learning models. Sci Rep 15:31432. 10.1038/s41598-025-15210-6
65. Willi S, Patrick R, Harald B (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med 26. 10.1002/sim.3148
66. Saejoon K (2014) Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms. Int J Data Min Bioinform 10. 10.1504/ijdmb.2014.064889
67. Jan Y et al (2020) Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. BMC Neurol 20. 10.1186/s12883-020-01672-w
68. Gaofeng H et al (2025) Identification of thyroid cancer biomarkers using WGCNA and machine learning. Eur J Med Res 30. 10.1186/s40001-025-02466-x
69. Huang S et al (2018) Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteom 15:41–51. 10.21873/cgp.20063
70. Breiman L, Random Forests (2001) Mach Learn 45:5–32. 10.1023/A:1010933404324
71. Shi X et al (2025) Integrated bioinformatics and experiment validation reveal cuproptosis-related biomarkers and therapeutic targets in sepsis-induced myocardial dysfunction. BMC Infect Dis 25:445. 10.1186/s12879-025-10822-9

72. Ling G, Chong-En X (2023) Integrated bioinformatics and machine learning algorithms reveal the critical cellular senescence-associated genes and immune infiltration in heart failure due to ischemic cardiomyopathy. Front Immunol 14. 10.3389/fimmu.2023.1150304
73. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210. 10.1093/nar/30.1.207
74. Naot D et al (2021) Molecular characterisation of osteoblasts from bone obtained from people of Polynesian and European ancestry undergoing joint replacement surgery. Sci Rep 11:2428. 10.1038/s41598-021-81731-5
75. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193. 10.1093/bioinformatics/19.2.185
76. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28:882–883. 10.1093/bioinformatics/bts034
77. Ritchie ME et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. 10.1093/nar/gkv007
78. Schober P, Vetter TR (2020) Nonparametric Statistical Methods in Medical Research. Anesth Analg 131:1862–1863. 10.1213/ane.0000000000005101
79. Bland JM, Altman DG (2009) Analysis of continuous data from small samples. BMJ 338:a3166. 10.1136/bmj.a3166
80. Mollan KR et al (2020) Precise and accurate power of the rank-sum test for a continuous outcome. J Biopharm Stat 30:639–648. 10.1080/10543406.2020.1730866
81. Amisha B, Disha T, Sanjay K, Subhash C (2023) Network pharmacology, molecular docking, and molecular dynamics simulation to elucidate the mechanism of anti-aging action of Tinospora cordifolia. Mol Divers 28. 10.1007/s11030-023-10684-w
82. Areeba A et al (2024) Silibinins and curcumin as promising ligands against mutant cystic fibrosis transmembrane regulator protein. AMB Express 14. 10.1186/s13568-024-01742-z
83. Yang L et al (2019) CB-Dock: a web server for cavity detection-guided protein-ligand blind docking. Acta Pharmacol Sin 41. 10.1038/s41401-019-0228-6
84. Newman AM et al (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12:453–457. 10.1038/nmeth.3337
85. Gentles AJ et al (2015) The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 21:938–945. 10.1038/nm.3909
86. Zhao L, Huang Y, Li J, Xie R, Li K (2025) Integrating bulk and single-cell RNA sequencing reveals intratumor heterogeneity phenotypes and immune infiltration in pancreatic cancer. Discov Oncol 16:1325. 10.1007/s12672-025-03080-3

87. Niu W, Zhang T, Ma L, Competitive Endogenous RNA (2025) Network Involving Immune Subgroups, Infiltration, and lncRNAs in Prostate Cancer. Genes (Basel) 16. 10.3390/genes16050527
88. An W et al (2024) Identification of crosstalk genes and immune characteristics between Alzheimer's disease and atherosclerosis. Front Immunol 15:1443464. 10.3389/fimmu.2024.1443464
89. Zhou Q et al (2024) Identification of ubiquitination-related hub genes in chronic myeloid leukemia cell by bioinformatics analysis. J Cancer 15:3750–3759. 10.7150/jca.96405
90. Weitao S et al (2022) Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. Imeta 1. 10.1002/imt2.36
91. Cobo T et al (2016) Role of Periostin in Adhesion and Migration of Bone Remodeling Cells. PLoS ONE 11:e0147837. 10.1371/journal.pone.0147837
92. Steppe L, Megafu M, Tschaffon-Müller MEA, Ignatius A, Haffner-Luntzer M (2023) Fracture healing research: Recent insights. Bone Rep 19:101686. 10.1016/j.bonr.2023.101686
93. Chen G, Deng C, Li YP (2012) TGF-β and BMP signaling in osteoblast differentiation and bone formation. Int J Biol Sci 8:272–288. 10.7150/ijbs.2929
94. Buckwalter JA, Glimcher MJ, Cooper RR, Recker R (1996) Bone biology. I: Structure, blood supply, cells, matrix, and mineralization. Instr Course Lect 45:371–386
95. Downey PA, Siegel MI (2006) Bone biology and the clinical implications for osteoporosis. Phys Ther 86:77–91. 10.1093/ptj/86.1.77
96. Ponzetti M, Rucci N (2021) Osteoblast Differentiation and Signaling: Established Concepts and Emerging Topics. Int J Mol Sci 22. 10.3390/ijms22136651
97. Robling AG, Castillo AB, Turner CH (2006) Biomechanical and molecular regulation of bone remodeling. Annu Rev Biomed Eng 8:455–498. 10.1146/annurev.bioeng.8.061505.095721
98. Datta HK, Ng WF, Walker JA, Tuck SP, Varanasi SS (2008) The cell biology of bone metabolism. J Clin Pathol 61:577–587. 10.1136/jcp.2007.048868
99. Sims NA, Gooi JH (2008) Bone remodeling: Multiple cellular interactions required for coupling of bone formation and resorption. Semin Cell Dev Biol 19:444–451. 10.1016/j.semcdb.2008.07.016
100. Matsuo K, Irie N (2008) Osteoclast-osteoblast communication. Arch Biochem Biophys 473:201–209. 10.1016/j.abb.2008.03.027
101. Shi Y et al (2043) Gli1 identifies osteogenic progenitors for bone formation and fracture repair. *Nat Commun* 8. 10.1038/s41467-017-02171-2 (2017)
102. Yoon H et al (2025) Unraveling the dynamics of osteoblast differentiation in MC3T3-E1 cells: Transcriptomic insights into matrix mineralization and cell proliferation. Bone 194:117442. 10.1016/j.bone.2025.117442
103. Qin XK et al (2024) [Expert consensus on clinical application of Dieda Qili Tablets in treatment of acute soft tissue injury]. Zhongguo Zhong Yao Za Zhi 49:5076–5080. 10.19540/j.cnki.cjcmm.20240531.501

104. Kim YJ, Bae YC, Suh KT, Jung JS (2006) Quercetin, a flavonoid, inhibits proliferation and increases osteogenic differentiation in human adipose stromal cells. Biochem Pharmacol 72:1268–1278. 10.1016/j.bcp.2006.08.021
105. Shen J et al (2024) The SRC/NF-κB-AKT/NOS3 axis as a key mediator of Kaempferol's protective effects against oxidative stress-induced osteoclastogenesis. Immun Inflamm Dis 12:e70045. 10.1002/iid3.70045
106. Figueiredo LC et al (2022) Propolis, Aloe Vera, Green Tea, Cranberry, Calendula, Myrrha and Salvia Properties against Periodontal Microorganisms. *Microorganisms* 10. 10.3390/microorganisms10112172
107. E S H EA, O M NR (2003) S. & A, S. Components, therapeutic value and uses of myrrh. Pharmazie 58
108. Li JS et al (2021) [Mechanism of Olibanum-Myrrha in treatment of rheumatoid arthritis based on network pharmacology and molecular docking]. Zhongguo Zhong Yao Za Zhi 46:2371–2379. 10.19540/j.cnki.cjcmm.20201126.401
109. James A (2004) Cell cycle development. Dev Cell 6. 10.1016/s1534-5807(04)00067-x
110. Antonin L, Laurence A-S, Christophe M, Jean-Charles S, Eric D (2011) [Cell cycle, mitosis and therapeutic applications]. Bull Cancer 98. 10.1684/bdc.2011.1382
111. Gould NR et al (2025) Proliferation of Early and Mature Osteoblasts Is Required for Bone Fracture Healing in Mice. bioRxiv. 10.1101/2025.05.27.656371
112. Schafer KA (1998) The cell cycle: a review. Vet Pathol 35:461–478. 10.1177/030098589803500601
113. Israels ED, Israels LG (2001) The cell cycle. Stem Cells 19:88–91. 10.1634/stemcells.19-1-88
114. Ying S, Yang L, Xiaoli M, Hao H (2021) The Influence of Cell Cycle Regulation on Chemotherapy. Int J Mol Sci 22. 10.3390/ijms22136923
115. Shi L et al (1994) Premature p34cdc2 activation required for apoptosis. Science 263:1143–1145. 10.1126/science.8108732
116. Santamaría D et al (2007) Cdk1 is sufficient to drive the mammalian cell cycle. Nature 448:811–815. 10.1038/nature06046
117. Gavet O, Pines J (2010) Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis. Dev Cell 18:533–543. 10.1016/j.devcel.2010.02.013
118. Satyanarayana A, Hilton MB, Kaldis P (2008) p21 Inhibits Cdk1 in the absence of Cdk2 to maintain the G1/S phase DNA damage checkpoint. Mol Biol Cell 19:65–77. 10.1091/mbc.e07-06-0525
119. Solomon MJ, Lee T, Kirschner MW (1992) Role of phosphorylation in p34cdc2 activation: identification of an activating kinase. Mol Biol Cell 3:13–27. 10.1091/mbc.3.1.13
120. Palozola KC et al (2017) Mitotic transcription and waves of gene reactivation during mitotic exit. Science 358:119–122. 10.1126/science.aal4671
121. Tanguay PL, Rodier G, Meloche S (2010) C-terminal domain phosphorylation of ERK3 controlled by Cdk1 and Cdc14 regulates its stability in mitosis. Biochem J 428:103–111. 10.1042/bj20091604

122. Liu R et al (2015) CDK1-Mediated SIRT3 Activation Enhances Mitochondrial Function and Tumor Radioresistance. Mol Cancer Ther 14:2090–2102. 10.1158/1535-7163.Mct-15-0017
123. Wang Z et al (2014) Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. Dev Cell 29:217–232. 10.1016/j.devcel.2014.03.012
124. Kamranvar SA, Rani B, Johansson S (2022) Cell Cycle Regulation by Integrin-Mediated Adhesion. *Cells* 11. 10.3390/cells11162521
125. Garg P et al (2017) Prospective Review of Mesenchymal Stem Cells Differentiation into Osteoblasts. Orthop Surg 9:13–19. 10.1111/os.12304
126. Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics. Nat Rev Genet 16:321–332. 10.1038/nrg3920
127. Ransohoff DF (2007) How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. J Clin Epidemiol 60:1205–1219. 10.1016/j.jclinepi.2007.04.020
128. Ferreira LG, Santos D, Oliva RN, G., Andricopulo AD (2015) Molecular docking and structure-based drug design strategies. Molecules 20:13384–13421. 10.3390/molecules200713384
129. Takayanagi H (2007) Osteoimmunology: shared mechanisms and crosstalk between the immune and bone systems. Nat Rev Immunol 7:292–304. 10.1038/nri2062
130. Lorenzo J, Horowitz M, Choi Y (2008) Osteoimmunology: interactions of the bone and immune system. Endocr Rev 29:403–440. 10.1210/er.2007-0038
131. Pacifici R (2010) The immune system and bone. Arch Biochem Biophys 503:41–53. 10.1016/j.abb.2010.05.027
132. Ono T, Takayanagi H (2017) Osteoimmunology in Bone Fracture Healing. Curr Osteoporos Rep 15:367–375. 10.1007/s11914-017-0381-0
133. Duda GN et al (2023) The decisive early phase of bone regeneration. Nat Rev Rheumatol 19:78–95. 10.1038/s41584-022-00887-0
134. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682–690. 10.1038/nchembio.118

# Tables

Table 1 to 24 are available in the Supplementary Files section.