Identification of Candidate Genes for Osteoporosis via Integrated Bioinformatics Analysis

HUAN XIN XIE, LEI CAO, LIN LIN YE, G. X. SHAN, C. J. JIANG AND W. Q. SONG*

Department of Rehabilitation, Xuanwu Hospital, Capital Medical University, Beijing 100053, China

Xie et al.: Identification of Candidate Genes for Osteoporosis

Osteoporosis is a common bone disease, characterized by low bone mass, damaged bone microstructure, increased bone brittleness, and increased susceptibility to fracture. Osteoporosis has no obvious early-stage manifestations; however, it presents with severe symptoms upon detection[1]. Therefore, osteoporosis is also known as a “silent disease.” In recent years, osteoporosis and its complications including deformity, disability, and depression have been a major public health concern among ageing populations worldwide. Race, age, sex, other bone metabolic diseases (endocrine system diseases, rheumatic immune diseases, etc.), drugs, and lifestyle contribute to osteoporosis pathogenesis[2]. The cost of nursing and medical care for osteoporosis individuals is enormous, imposing a heavy burden on the family and society[3].

The potential mechanism underlying osteoporosis pathogenesis involves an imbalance between bone

Osteoporosis is a common bone disease; however, its pathophysiology is yet unclear. This study aimed to investigate candidate genes in osteoporosis and its pathomechanism. Microarray datasets Genomic Spatial Event-7429, 13850, 56815, and 7158, comprising data obtained from blood samples of osteoporosis patients and healthy controls, were downloaded from the Gene Expression Omnibus database. Differentially expressed genes were identified via intersection of the four datasets, using Affy and Limma software packages. Functional and pathway enrichment analysis of differentially expressed genes were conducted using the database for annotation, visualization, and integrated discovery database. Thereafter, protein-protein interactions between the products of differentially expressed genes and key modules were analyzed using the search tool for the retrieval of interacting genes/proteins database and Cytoscape software. Furthermore, a transcriptional regulatory network was established with the differentially expressed genes, using the Web-based gene set analysis toolkit database. In total, 702 differentially expressed genes were filtered from the intersection of the four datasets, which were primarily enriched in functions and pathways associated with glutamate secretion, Adenosine triphosphate binding, extracellular region, and phosphatidylinositol 3-kinase-protein kinase B signaling pathway. Furthermore, 361 nodes and 846 edges were present in the protein-protein interaction network. Two significant modules were obtained, each containing 22 key genes. Transcriptional factor-target gene regulatory network analysis revealed two vital transcription factors, Forkhead Box O4 and LIM Homeobox 3. Genes including endothelin receptor type A, XC chemokine receptor 1, 5-hydroxytryptamine receptor 2a, phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-4, glutamate metabotropic receptor 5, kiss-1 metastasis suppressor, 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-2, cholinergic receptor muscarinic 3, g protein subunit gamma 4, neurexins receptor 1, neurexins U, neurexins B, breast cancer gene 1 associated RING domain 1, exonuclease 1, replication factor C subunit 2, restriction site associated DNA 52 homolog DNA repair protein, DNA topoisomerase III alpha, Hemolytic uremic syndrome 1 checkpoint clamp component, timeless circadian regulator, breast cancer gene 1 DNA repair associated, tumor protein p53 and caspase are potential key genes for osteoporosis and may help elucidate the underlying pathomechanism, and transcription factors Forkhead Box O4 and LIM Homeobox 3 are potential therapeutic targets for osteoporosis.

Key words: Osteoporosis, differentially expressed genes, bioinformatics, pathway, protein-protein interaction network

*Address for correspondence
E-mail: songwq66@126.com
resorption and formation\textsuperscript{[2]}. Branden reported that tumor necrosis factor (TNF) receptor-associated factor 3 (TRAF3) down regulation in bone tissue potentially increases osteoclastogenesis and reduces osteoblastogenesis through the nuclear factor kappa light chain enhancer of activated B cells (NF-Kb) signaling pathway\textsuperscript{[4]}. H3 lysine 36 trimethylation (H3K36me3) mediated by SET-domain-containing 2 (SETD2) potentially regulates the cell fate of bone marrow stromal cells (BMSCs)\textsuperscript{[5]}. Family with sequence similarity 210 member A (FAM210A) reportedly regulates bone structure and function and may impact osteoporosis\textsuperscript{[6]}. Uehara reported that protein kinase N3 promotes osteoclast-mediated bone resorption in response to Wingless-related integration site (Wnt) family member 5A-receptor tyrosine kinase like orphan receptor 2 (Wnt5a-Ror2) signaling\textsuperscript{[7]}. Furthermore, numerous studies have verified that the NF-κB signaling pathway, Notch signaling pathway, Wnt/β-catenin pathway, and phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) (PI3K/AKT) pathway are associated with the regulation of BMSC differentiation during osteoporosis. However, the mechanisms underlying BMSC differentiation remain unclear.

Microarray technology has contributed significantly to advancements in the identification of differentially expressed genes (DEGs), functional genes, and gene regulatory networks associated with disease\textsuperscript{[8]}. It has advantages including small size, rapidity, accuracy, high sensitivity, and parallel detection of large data on the same chip. However, these results are generated via a single queue and are always inconsistent among different studies because of organizational or sample heterogeneity among individual experiments, different technological testing platforms, and various data processing methods\textsuperscript{[11]}. In this study, 4 Messenger Ribonucleic acid (mRNA) microarray datasets from Gene Expression Omnibus (GEO) were downloaded and analyzed to identify DEGs in blood samples obtained from osteoporosis patients and healthy controls. Further, we carried out Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and constructed a protein-protein interaction (PPI) network and interaction network between transcription factors (TFs) and DEGs. This study aimed to identify more reliable potential modulators and pathophysiological mechanisms of osteoporosis.

**MATERIALS AND METHODS**

**Gene datasets:**

We obtained four gene expression datasets, (Genomic spatial event (GSE) (GSE7429, GSE13850, GSE56815, and GSE7158) from the GEO database\textsuperscript{[10]}. Thereafter, we screened the four original microarray datasets and analyzed differential gene expression profiles between blood samples from osteoporosis patients and healthy controls among premenopausal or postmenopausal women. According to the diagnostic criteria recommended by World Health Organization (WHO), bone mineral density (BMD) should be measured via dual energy X-ray absorptiometry (DXA). Patients with a hip or lumber T-score ≤-2.5 were identified to have osteoporosis\textsuperscript{[2]}. All women were divided into low and high BMD groups.

**Data preprocessing and analysis of DEGs:**

The Affy package with the R package (version 3.5.1) was used to preprocess the original data from the four datasets, including quantile normalization, log2 transformation, and the K-nearest neighbors algorithm to supplement missing values\textsuperscript{[11]}. The Limma package in R was used to identify DEGs in the high BMD and low BMD groups\textsuperscript{[12]}. A corrected p-value <0.01 was selected as the threshold for DEG screening. Consequently, data from different datasets were intersected and analyzed to obtain 702 DEGs.

**Gene ontology (GO) and signaling pathway enrichment analysis of DEGs:**

GO and KEGG pathway enrichment analyses were performed using database for annotation, visualization and integrated discovery (DAVID) database to identify functional annotations (molecular function, biological process, and cellular component) and pathways involving DEGs. A p-value of <0.05 was considered the threshold\textsuperscript{[13]}.  

**Construction of the PPI network and module analysis:**

Search tool for the retrieval of interacting genes/proteins (STRING) database (version 11.0), a PPI network, helped to identify core regulatory genes\textsuperscript{[14]}. We used a confidence score of >0.7 as the cut-off criterion. Furthermore, we constructed a visualized PPI network, using Cytoscape software (version 3.7.0), an open source software platform for visualizing complex networks and integrating them with any type of attribute.
data\textsuperscript{[15]}. Molecular Complex Detection (MCODE), a plug-in for Cytoscape, was used to construct functional modules by clustering in a huge protein network and information of each node was determined to obtain the most important visualized model data\textsuperscript{[16]}. The screening requirements were as follows: MCODE scores >5, degree cut-off=2, node score cut-off=0.2, Max depth=100, and k-score=2.

**RESULTS AND DISCUSSION**

In this study, we performed four microarray analyses. After processing the data, we identified 5440 DEGs in GSE7158, 6104 DEGs in GSE7429, 5263 DEGs in GSE13850, and 9260 DEGs in GSE56815 (fig. 1). The intersection of the four datasets yielded 702 genes, as obtained through a Venn diagram (fig. 2) between low BDM and high BDM women.

The PPI network of DEGs and the most significant module were established using Cytoscape software. The PPI network comprised 361 nodes and 846 edges (fig. 3A). Two significant modules (module 1 and 2) were obtained using the MCODE plug-in. Genes in the two modules include endothelin receptor type A (EDNRA),

![Fig. 1: Heat map image showing the expression of DEGs GSE7158, GSE7429, GSE13850 and GSE56815](image-url)
To further elucidate the association between TFs and target genes, we analyzed the WebGestalt database to construct a TF-target gene network, using Cytoscape software. Two most vital TFs, Forkhead Box O4 (FoxO4) and LIM Homeobox 3 (LHX3), and 145 interactions pairs were obtained (fig. 4).

Osteoporosis is a systemic skeletal disease characterized by low BMD, occurring in the aged population and represents increasing burdens on society. In summary, 702 DEGs among the 4 databases, 22 hub genes and 2 TFs were identified, which are potential key genes in osteoporosis.

To identify interactions among DEGs, GO and KEGG enrichment analyses were carried out. Molecular function upon GO enrichment analysis was primarily glutamate binding, protein kinase activity, and protein tyrosine kinase activity. Previous studies have reported that these functions play important roles in osteoporosis progression. The expression of glutamate occurs in both osteoblasts and osteoclasts. Glutamate is vital for bone formation and resorption. Glutamate transporters are potentially applicable for enhancing the osteogenic ability of osteoblasts\(^{[19]}\). BMSC proliferation during osteoporosis was primarily determined on the basis of protein kinase B (Akt)/mammalian target of rapamycin (mTOR) activity\(^{[19]}\). Protein tyrosine kinase activity is essential for the regulation of bone resorption in osteoclasts\(^{[20]}\).

In this study, KEGG analysis revealed that the important pathways are the PI3K-AKT and Wnt signaling pathways. Recent studies reported that these pathways may regulate osteoclast and osteoblast...
activity during osteoporosis. Osteoblast differentiation is reportedly regulated by MicroRNA 216a (miR-216a) in the PI3K-AKT signaling pathway[1]. Similarly, Zhao reported that the Phosphatase and tensin homolog (PTEN)/PI3K-AKT signaling pathway promotes osteoclastogenesis during osteoporosis[21]. Ge reported that the Wnt/β-catenin signaling pathway plays a vital role in promoting osteogenesis[22]. Boyden reported that Wnt signaling is involved in bone metabolism through Low-density lipoprotein receptor-related protein 5 (LRP5)[23]. Furthermore, some studies strongly suggest that BMD is associated with Wnt signaling.

On construction of the PPI network using Cytoscape software, we obtained two significant modules, including 22 key genes (EDNRA, XCR1, HTR2A, PLCB4, GRM5, KISS1, PLCB2, CHRM3, GNG4, NTSR1, NMU, NMB, BARD1, EXO1, RFC2, RAD52, TOP3A, HUS1, TIMELESS, BRCA1, TP53, and CLSPN) potentially significant in osteoporosis. BRCA1 inhibits postnatal osteoblastic osteogenesis through Mitogen-activated protein kinase (MAPK) signaling pathways[24]. A bioinformatics analysis reported that BRCA1 is a potential key gene for bone regeneration and repair during osteopenia in osteoporosis[25]. BRCA1 is a tumor suppressor in osteosarcomas and regulates tumorigenesis[26,27]. Johnson reported that EDNRA inhibits the marked endothelin-1-mediated increase in mineralization. NMU, a small peptide, influences bone rebuilding in conditions of low bone mass, such as osteoporosis, owing to leptin-dependent regulation[28]. Dimitri reported that neuromedin U is potentially involved in the regulation of the bone remodeling[29]. Knockdown of neuromedin B could expedite apoptosis in osteoclast lineage cells and reduce proliferation, thus decreasing bone mass[30]. A previous study suggested that TP53 regulates osteogenic differentiation of dental stem cells; however, the pathomechanism of osteogenesis has yet been unclear[31]. In this study, significant differences were observed in the expression of the aforementioned genes between blood samples of osteoporosis patients and healthy controls. This study shows that these genes are essential markers in bone metabolism. Further studies are required to further elucidate the mechanism underlying osteoporosis.

HUS1 protects against the toxic effects of formaldehyde on mouse bone marrow mesenchymal stem cells[32]. GRM5 expression is reportedly associated with bone cancer pain in a mouse model[33]. TOP3A, a vital microsatellite marker, plays a potentially significant role in high-grade osteosarcoma and influences
tumorigenesis[34], XCR1 is essential in cell proliferation and migration and has been correlated with bone metastasis[35]. XCR1 is a potential diagnostic marker for large B-cell lymphoma demonstrating bone marrow involvement[36]. These four genes are associated with bone tumor and bone mass; however, no direct evidence is available to determine the effect of these genes on osteoblast/osteoclast differentiation and the pathophysiology of osteoporosis.

HTR2A, BARD1, TIMELESS, PLCB2, RFC2, CHRM3, GNG4, PLCB4, CLSPN, EXO1, RAD52, NTSR1, and BRCA1 are key molecular players in osteoporosis in this study. HTR2A, a neuronal marker, is involved in differentiation of bone marrow mesenchymal stem cells into neuronal lineage cells[37]. BARD1 and TIMELESS are potential biomarkers for breast cancer[38,39]. PLCB2 and RFC2 play a potentially important role in osteosarcoma[40,41]. CHRM3, GNG4, and PLCB4 regulate colon cancer, glioblastoma, and gastrointestinal stromal tumors, respectively[42-44]. CLSPN, EXO1, and RAD52 are associated with DNA replication checkpoints, regulation of DNA, and DNA repair, respectively[45-47]. NTSR1 interacts with lipids[48]. No study has reported the association between these genes and osteoporosis. However, our results suggest that they are potential key factors involving in osteoporosis regulation and warrant further clinical and in vitro analyses.

In the present study, FoxO4 and LHX3 were identified as key TFs in the TF-target gene network, playing an important role in osteoporosis. The FoxO TF family, comprising FoxO1, FoxO3a, FoxO4, and FoxO6, is involved in cellular survival, the oxidative stress response, and carcinogenesis[49]. FoxO4 mediates osteoblast differentiation by deacetylating FoxO4 in progenitors[50]. Iyer reported that FoxO4 attenuates osteogenesis during osteoporosis through the Wnt pathway[51]. Furthermore, Ambrogini reported that FoxO4 helps defend against oxidative stress in osteoblasts and regulates bone mass homeostasis in mice[52]. Moreover, FoxO4 participates in the tumorigenic regulation in different cancers including gastric cancer, colorectal cancer, hepatocellular carcinoma, prostate cancer, breast cancer, and lung cancer[49]. Roupe reported that FoxO4 is involved in wound healing and tissue regeneration[53]. LHX3, a typical LIM family TF, is reportedly associated with regulation of pituitary hormones, thyroid function, and the β-subunit of follicle-stimulating hormone (FSHβ)[54-56]. This study predicts that LHX3 is a potentially important TF in monitoring osteoporosis. These results warrant verification through further studies.

This study has two primary limitations. First, hub genes were identified using bioinformatics approaches without experimental verification. Further studies are required to verify these results using in vitro analyses including quantitative real-time reverse transcription polymerase chain reaction (RT-PCR) and western blot analyses. Second, all data were standardized; however, there may be heterogeneity among individual microarray studies using different platforms and clinical samples, such as assets, sex, or geographical regions. In this study, we intersected 4 datasets, with differences among different algorithms.

In conclusion, this study revealed 702 DEGs via bioinformatics analysis. Thereafter, we 22 hub genes and 2 TFs were identified to be involved in osteoporosis pathophysiology and serve as potential therapeutic targets for osteoporosis. However, further studies are required to verify the function of these genes in osteoporosis.

**Author’s contributions:**

Huan Xin Xie and Lei Cao carried out the studies, participated in collecting data, and drafted the manuscript. Lin Lin Ye and Jie Hu performed the statistical analysis and participated in its design. Gui Xiang Shan and Wei Qun Song helped to draft the manuscript. All authors read and approved the final manuscript.

**Acknowledgements:**

We would like to thank all the lab tutors for participating in this study.

**Funding:**

The work was supported by the National Science Foundation of China (No. 303-01-002-0155).

**Conflict of Interests:**

The authors declared no conflict of interest.

**REFERENCES**


27. Engert F, Kovac M, Baumhoer D, Nathrath M, Fulda S. Osteosarcoma cells with genetic signatures of BRCA1/BRCA2 are susceptible to the PARP inhibitor talazoparib alone or in combination with chemotherapeutics. Oncotarget 2017;8;40794.


44. Li CF, Liu TT, Chuang IC, Chen YY, Fang FM, Chan TC, et al. PLCβ4 copy gain and PLCβ4 overexpression in primary gastrointestinal stromal tumors: Integrative characterization of a lipid-catabolizing enzyme associated with worse disease-free survival. Oncotarget 2017;8:19997-20010.


This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

This article was originally published in a special issue, "Evolutionary Strategies in Biomedical Research and Pharmaceutical Sciences" Indian J Pharm Sci 2020:83(3) Spl issue2;6-13