Identification of candidate genes in osteoporosis by integrated microarray analysis

J. J. Li, B. Q. Wang, Q. Fei, Y. Yang, D. Li
Beijing Friendship Hospital, Capital Medical University, Beijing, China

Objectives
In order to screen the altered gene expression profile in peripheral blood mononuclear cells of patients with osteoporosis, we performed an integrated analysis of the online microarray studies of osteoporosis.

Methods
We searched the Gene Expression Omnibus (GEO) database for microarray studies of peripheral blood mononuclear cells in patients with osteoporosis. Subsequently, we integrated gene expression data sets from multiple microarray studies to obtain differentially expressed genes (DEGs) between patients with osteoporosis and normal controls. Gene function analysis was performed to uncover the functions of identified DEGs.

Results
A total of three microarray studies were selected for integrated analysis. In all, 1125 genes were found to be significantly differentially expressed between osteoporosis patients and normal controls, with 373 upregulated and 752 downregulated genes. Positive regulation of the cellular amino metabolic process (gene ontology (GO): 0033240, false discovery rate (FDR) = 1.00E + 00) was significantly enriched under the GO category for biological processes, while for molecular functions, flavin adenine dinucleotide binding (GO: 0050660, FDR = 3.66E-01) and androgen receptor binding (GO: 0050681, FDR = 6.35E-01) were significantly enriched. DEGs were enriched in many osteoporosis-related signalling pathways, including those of mitogen-activated protein kinase (MAPK) and calcium. Protein-protein interaction (PPI) network analysis showed that the significant hub proteins contained ubiquitin specific peptidase 9, X-linked (Degree = 99), ubiquitin specific peptidase 19 (Degree = 57) and ubiquitin conjugating enzyme E2 B (Degree = 57).

Conclusion
Analysis of gene function of identified differentially expressed genes may expand our understanding of fundamental mechanisms leading to osteoporosis. Moreover, significantly enriched pathways, such as MAPK and calcium, may involve in osteoporosis through osteoblastic differentiation and bone formation.

Cite this article: Bone Joint Res 2016;5:594–601

Keywords: Integrated analysis, Microarray, Osteoporosis, Peripheral blood mononuclear cells, Differentially expressed genes

Article focus
- To screen the altered gene expression profile in peripheral blood mononuclear cells of patients with osteoporosis.
- To interpret the biological roles of the identified differentially expressed genes in patients with osteoporosis.
- To identify the potential candidate genes in osteoporosis.

Key messages
- A total of 1125 genes were found to be significantly differentially expressed in peripheral blood monocytes (PBMs) between osteoporosis patients and normal controls, with 373 upregulated and 752 downregulated genes.
- Some genes, such as nuclear receptor interacting protein 1, interleukin 1...
The significantly enriched pathways, such as the mitogen-activated protein kinase and the calcium signalling pathways, may involve in osteoporosis through osteoblast differentiation and bone formation.

**Strengths and limitations**

- This study revealed the gene expression profiles in PBMs between osteoporosis patients and normal controls with large statistical power.
- The gene function analysis of identified differentially expressed genes may expand our understanding of fundamental mechanisms leading to osteoporosis.
- However, the differentially expressed genes in our study are predicted without experimental evidence, such as reverse transcription polymerase chain reaction and Western blot. Therefore, further research is needed to uncover gene functions in osteoporosis pathogenesis.

**Introduction**

Osteoporosis is one of the most common and serious metabolic bone diseases, which affects 200 million people worldwide, among which 80% are women aged 60 years or older. Osteoporosis manifests clinically by reduced bone mass, a decreased amount of normally mineralised bone and microarchitectural deterioration of bone tissues, accounting for an increase in susceptibility to fracture. Genetic predisposition, combined with advanced age, gender, immobilisation, and other risk factors, contribute to the development of osteoporosis. Genome-wide association studies have been performed to search for an association between bone mineral density (BMD) and osteoporosis in twin and family studies. A meta-analysis of genome-wide association studies identified 56 low BMD-associated loci, and 14 loci which were associated with fracture. The association of vitamin D (1,25-dihydroxyvitamin D3) receptor, oestrogen receptor 1 (ER) and LDL receptor related protein 5 with osteoporosis, has been most widely studied as candidate genetic genes.

The imbalance between bone resorption (by osteoclasts) and bone formation (by osteoblasts) is a key pathological mechanism of osteoporosis as a consequence of pathologically prolonging the life span of osteoclasts and early apoptosis of osteoblasts and osteocytes. Circulating monocytes play important roles in this process, as they differentiate into osteoclasts to be involved in bone resorption, and produce a wide variety of factors for osteoclastogenesis and bone resorption. Osteoclasts from circulating mononuclear cells of osteoporotic patients demonstrated increased bone resorption activity compared with normal controls. Primary osteoporosis is related either to the postmenopausal state or advanced age (senile osteoporosis) due to the cellular changes associated with oestrogen deficiency and ageing. For postmenopausal osteoporosis, oestrogen deficiency leads to upregulation of receptor activator of nuclear factor-kappa ligand (RANKL) and tumour necrosis factor-α, resulting in excessive bone resorption and rapid bone loss. For senile osteoporosis, there is also a significant reduction in bone formation due to osteoblastic dysfunction. Osteoblasts and adipocytes both differentiate from mesenchymal stem cells (MSCs), and there is a shift from osteoblastogenesis to adipogenesis in the bone marrow with increasing age, causing reduction in the life span of osteoblasts.

Recently, a number of genes and pathophysiological mechanisms have been identified for osteoporosis through microarray studies in peripheral blood monocytes (PBMs) of osteoporotic patients. However, there are inconsistencies among these studies due to the use of different sample sources, platforms and analysis techniques. Towards this end, we integrated gene expression data sets from multiple microarray studies of osteoporosis to overcome these limitations of individual studies. The identification of differentially expressed genes (DEGs) between osteoporosis and normal controls and their biological function would expand our understanding of fundamental mechanisms leading to osteoporosis, and may be meaningful for further diagnosis and therapy for this disease.

**Materials and Methods**

**Eligible gene expression profiles of PBMs in osteoporosis.** The Gene Expression Omnibus (GEO) database was searched to obtain gene expression profiling studies of osteoporosis subjects in PBMs. The following key search terms were used: “osteoporosis”, “peripheral blood monocytes”, “gene expression”, “microarray”. We included original microarray studies which analysed the differential gene expression profiles of PBMs between patients with osteoporosis and normal controls (Fig. 1).

**Statistical analysis.** The existence of heterogeneity among multiple microarray studies arising from different microarray platforms, gene nomenclature and clinical samples, makes it infeasible to compare the gene expression data directly. Therefore, normalisation is necessary to minimise the heterogeneity. Consequently, we preprocessed the raw microarray data of each study by quantile normalisation and log2 transformation to obtain the z-score for the purpose of global normalisation.

The MATLAB laboratory (MATLAB) software (MathWorks, Natick, Massachusetts) was used to identify the differently expressed probe sets in PBMs between osteoporosis patients and normal controls. Gene specific t-tests were carried out, and p-values were calculated. Multiple testing adjustment was performed. The genes with a p-value < 0.05 were selected as DEGs. Heat map analysis was performed using the “heatmap.2” function of the R/Bioconductor
Functional annotation of DEGs. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed to detect the biological functions and potential pathway of DEGs. Go-rilla,23 a web-accessible programme was used for GO enrichment analysis. The online-based software GeNecoDIS was used24 in pathway enrichment analysis based on the KEGG pathway database which is a recognised and comprehensive database including an extensive array of biochemistry pathways.25

Protein-protein interaction (PPI) network construction. The PPIs analysis is expected to explore the functions of proteins at the molecular level. PPIs may provide novel insights into the regulatory mechanisms of cellular growth, development, metabolism, differentiation and apoptosis.26 The Biological General Repository for Interaction Datasets (BioGRID),27 which is an online interaction repository, was used to construct the PPI network. The PPI network was visualised using Cytoscape software (Institute of Systems Biology, Seattle, Washington).28 Nodes denote proteins and edges denote interactions between two proteins.

**Results**

Gene expression changes in PBMs between osteoporosis and NC. In this study, a total of three microarray studies (GSE7429, GSE2208 and GSE7158) were available by...
searching the GEO database, all of which analysed the differential gene expression profiles of PBM between osteoporosis patients and normal controls. The included blood samples were from unrelated pre- and postmenopausal women. Patients with a hip T-score ≤ -2.5 were defined as having osteoporosis according to the World Health

Table II. The top ten most significantly up- or downregulated differentially expressed genes

| Gene ID | Gene symbol | Official full name | p-value    | Fold change |
|---------|-------------|--------------------|------------|-------------|
| 51098   | IFT52       | Intraflagellar transport 52 homolog (Chlamydomonas) | 3.90×10⁻⁵  | 1.8521      |
| 57226   | LYRM2       | Lyr motif containing 2 | 1.32×10⁻⁴  | 1.6957      |
| 7320    | UEZEB       | Ubiquitin-conjugating enzyme E2B (RAD6 homolog) | 2.30×10⁻⁴  | 1.5391      |
| 11213   | IRAK3       | Interleukin-1 receptor-associated kinase 3 | 3.16×10⁻⁴  | 1.6481      |
| 10026   | PICK        | Phosphatidylinositol glycan anchor biosynthesis class K | 3.41×10⁻⁴  | 1.7252      |
| 8204    | NRP1        | Nuclear receptor interacting protein 1 | 3.46×10⁻⁴  | 1.7031      |
| 52297   | CCDC91      | Coiled-coil domain containing 91 | 3.98×10⁻⁴  | 1.7437      |
| 8565    | YARS        | Tyrosyl-tRNA synthetase | 4.77×10⁻⁴  | 1.6445      |
| 8239    | USP9X       | Ubiquitin-specific peptidase 9 X-linked | 4.93×10⁻⁴  | 1.5406      |
| 10944   | C11orf58    | Chromosome 11 open reading frame 58 | 5.12×10⁻⁴  | 1.5487      |

Upregulated genes

| Gene ID | Gene symbol | Official full name | p-value    | Fold change |
|---------|-------------|--------------------|------------|-------------|
| 81577   | GFOD2       | Glucose-fructose oxidoreductase domain containing 2 | 5.33×10⁻⁵  | -1.6809     |
| 79574   | EPS8L3      | EPS8-like 3 | 8.17×10⁻⁵  | -2.1028     |
| 6943    | TCF21       | Transcription factor 21 | 2.23×10⁻⁴  | -1.741      |
| 1104112 | TXNRD3      | Thioredoxin reductase 3 | 2.66×10⁻⁴  | -1.6261     |
| 730227  | LINCO1136   | Long intergenic non-protein coding RNA 1136 | 3.49×10⁻⁴  | -1.7914     |
| 10505   | SEMA4F      | Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4F | 3.66×10⁻⁴  | -1.5667     |
| 7706    | TRIM25      | Tripartite motif-containing 25 | 4.20×10⁻⁴  | -1.6876     |
| 2627    | GATA6       | GATA binding protein 6 | 4.24×10⁻⁴  | -1.7631     |
| 10869   | USP19       | Ubiquitin-specific peptidase 19 | 5.49×10⁻⁴  | -1.6535     |
| 51267   | CLEC1A      | C-type lectin domain family 1, member A | 6.09×10⁻⁴  | -1.7525     |

Downregulated genes

Heat map image showing the expression of the top 100 significantly upregulated or downregulated genes in patients with osteoporosis.
Organization criteria. The details of the included microarray studies are presented in Table I. According to the selection criteria (p-value < 0.05), 1125 genes were found to be significantly differentially expressed in PBMs between osteoporosis patients and normal controls, with 373 upregulated and 752 downregulated genes. The top ten most significantly up- or downregulated genes are listed in Table II. The upregulated gene with the lowest p-value was intraflagellar transport 52 (IFT52), which is a core component of the IFT particle. IFT is responsible for the formation and maintenance of flagella and cilia, and is involved in bone development by regulating cell differentiation in osteoblasts, osteocytes, chondrocytes, and MSCs. The downregulated gene with the lowest p-value was glucose-fructose oxidoreductase domain containing 2 (GFOD2), which has been found to be associated with blood lipid levels. The full list of the differentially genes is provided as supplementary material. The top 100 most significant DeGs are displayed in a heat map across different microarray studies of osteoporosis (Fig. 2).

**Functional annotation.** For the identified DeGs, we performed enrichment analysis in GO categories including biological process and molecular function. Positive regulation of the cellular amine metabolic process (GO: 0033240, false discovery rate (FDR) = 1.00E + 00) and neural tube formation (GO: 0050681, FDR = 1.00E + 00) were significantly enriched under the GO category of biological processes (Fig. 3a), while for molecular functions, flavin adenine dinucleotide binding (GO: 0050660, FDR = 3.66E-01) and androgen receptor binding (GO: 0050681, FDR = 6.35E-01) were significantly enriched (Fig. 3b).

Furthermore, KEGG pathway enrichment analysis was also conducted. Hypergeometric tests with p-value < 0.05 were identified as the criteria for significantly enriched pathways. The most significantly enriched pathway was measles (FDR = 6.53E-06). Furthermore, neuroactive ligand-receptor interaction (FDR = 5.08E-05) and the mitogen-activated protein kinase (MAPK) signalling pathway (FDR = 5.78E-05) are also highly enriched. Interestingly, the calcium signalling pathway (FDR = 5.83E-05) was found to be significantly enriched.

**PPI network construction.** PPI networks of the top ten upregulated and downregulated DEGs were established.
by Cytoscape software, consisting of 398 nodes and 433 edges. In the PPI network, the nodes with ‘high degree’ are defined as ‘hub proteins’, and degrees are defined by the number of neighbours directly connected to the node. The significant hub proteins contained USP9X (Degree = 99), USP19 (Degree = 57) and UBE2B (Degree = 57) (Fig. 4).

**Discussion**
Recent studies have advanced our understanding of pathophysiological mechanisms for osteoporosis. Several key pathways such as the RANK-RANKL-osteo protegerin pathway, MSC differentiation, endochondral ossification and the Wnt signalling pathways, have been found to be involved in the regulation of bone mineral homeostasis. Recent studies have advanced our understanding of pathophysiological mechanisms for osteoporosis. Several key pathways such as the RANK-RANKL-osteo protegerin pathway, MSC differentiation, endochondral ossification and the Wnt signalling pathways, have been found to be involved in the regulation of bone mineral homeostasis. Based on the published evidence, novel therapeutic targets for osteoporosis have been identified and applied in a clinical setting, such as anti-RANKL receptor activator of NF-κB ligand. In this study, we integrated three microarray studies, of which gene expression profiles of PBMs between osteoporosis patients and normal controls were compared with the aim of identifying new candidates to target for osteoporosis drug discovery.

We identified 1125 DEGs in PBMs between osteoporosis patients and normal controls, which included 373 upregulated and 752 downregulated genes. Direct connections have been found between some of these genes and osteoporosis, such as nuclear receptor interacting protein 1 (NRIP1), interleukin 1 receptor associated kinase 3 (IRAK3), bone morphogenetic protein 7 (BMP7), SMAD family member 1 (SMAD1), and MAPK3. The single-nucleotide polymorphisms (SNPs) of the NRIP1 gene which modulates transcriptional activity of the oestrogen
receptor, were correlative with postmenopausal osteoporosis.36,37 Chen and Xia38 showed that there was a close relationship between NRP1 and BMD in B cells in a microarray study. The expression of IRAK3 is predominately in peripheral blood leukocytes and induced strongly upon the maturation of macrophages.39 Li et al40 reported that the IRAK3 plays a critical role as a negative regulator of toll-like receptor/interleukin (IL)-1 receptor (TLR/IL-1R) signalling in osteoclast differentiation and activation and, concomitantly, in the development of osteoporosis.

In this study, we also found several other genes which have been implicated in bone and skeletal muscle development, such as IFT52, GATA binding protein 6 (GATA6), tripartite motif containing 25 (TRIM25) and ubiquitin specific peptidase 19 (USP19), but the correlation with osteoporosis has never been reported. IFT52 is a core component of the IFT particle, which is involved in bone development by regulating cell differentiation in osteoblasts, osteocytes, chondrocytes, and MSCs.33 GATA6, as a direct target of micro-RNA miR-181a, plays a key role during BMP-induced osteoblastic differentiation of C2C12 (a mouse myoblast cell line) and MC3T3 cells (a clonal osteoblast-like mouse calvarial cell line).41 TRIM25, as one of many oestrogen-responsive genes, functions in the general maintenance of healthy bone.42

Furthermore, the results from PPI network analysis of the top ten upregulated and downregulated DEGs indicated that the most significant hub protein was USP9X. USP9X, located on the X-chromosome, plays a pivotal role in development and cancer by regulating transforming growth factor (TGF) signalling, Notch, epidermal growth factor, Wnt and mechanistic target of rapamycin signalling pathways. USP9X has been shown to regulate the apoptotic signalling networks.43 Osteoporosis is associated with dysregulated apoptosis based on the evidence that osteoporosis occurs mostly as a result of oestrogen deficiency which accelerates osteoblast apoptosis and susceptibility to osteoporotic fractures.44

KEGG pathway enrichment analysis showed that DEGs were enriched in many osteoporosis-related signalling pathways including the MAPK signalling pathway and the calcium signalling pathway. MAPK signalling pathways contribute greatly to osteoblast differentiation and bone formation via TGF-β and BMP signalling pathways.45,46 Vitamin D deficiency and reduced calcium absorption in the elderly population may be a risk factor of osteoporosis. An increase in calcium intake is the most commonly recommended preventive measure for osteoporosis.47

However, the DEGs in our study are predicted without experimental evidence such as reverse transcription polymerase chain reaction and Western blot. Therefore, further research is needed to uncover gene functions in osteoporosis pathogenesis. Overall, in our study a number of DEGs were identified and some of them may play important roles in the pathogenesis of osteoporosis. These findings may be useful in identifying the molecular mechanisms of osteoporosis and advancing therapy development.

The present study has some limitations. First, although global normalisation was performed, there may exist heterogeneity among the individual microarray studies arising from different platforms and clinical samples (such as severity, gender, or geographical regions). Secondly, results from different software algorithms may be slightly different. In the process of data analysis, we selected software algorithms (e.g. GO-riilla, GENEODIS, BioGRID) that are commonly used based on relevant references.23,24,27

In conclusion, we used an integrated analysis approach to integrate three gene expression data sets of osteoporosis, and identified DEGs in PBMs between osteoporosis patients and normal controls. The gene function analysis showed that these DEGs were highly correlated with the development of osteoporosis, expanding our understanding of fundamental mechanisms leading to osteoporosis, and may be meaningful for further diagnosis and therapy for this disease.

### Supplementary material

Table showing the top 15 most significantly enriched KEGG pathway of DEGs is available alongside the online version of this article at www.bjr.boneandjoint.org.uk

### References

1. Lewiecki EM, Silverman SL. Redefining osteoporosis treatment: who to treat and how long to treat. Arq Bras Endocrinol Metabol 2006;50:694-704.
2. Kamel HK. Postmenopausal osteoporosis: etiology, current diagnostic strategies, and nonprescription interventions. J Manag Care Pharm 2006;12(Suppl A):S4-S9.
3. Benisch P, Schilling T, Klein-Hitpass L, et al. The transcriptional profile of mesenchymal stem cell populations in primary osteoporosis is distinct and shows overexpression of osteogenic inhibitors. PLoS One 2012;7:e45142.
4. Richards JB, Rivadeneira F, Inouye M, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 2008;371:1505-1512.
5. Athanasiadis G, Malouf J, Hernandez-Sosa N, et al. Linkage and association analyses using families identified a locus affecting an osteoporosis-related trait. Bone 2014;60:98-103.
6. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet 2012;44:491-501.
7. Boyce BF, Hughes DE, Wright KR, Xing L, Dai A. Recent advances in bone biology provide insight into the pathogenesis of bone diseases. Lab Invest 1998;79:83-94.
8. Manolagas SC, Kousteni S, Jilka RL. Sex steroids and bone. Recent Prog Horm Res 2002;57:385-409.
9. Tomkinson A, Gevers EF, Wit JM, Reeve J, Noble BS. The role of estrogen in the control of rat osteocyte apoptosis. J Bone Miner Res 1998;13:1243-1250.
10. Fujikawa Y, Quinn JM, Sabokbar A, McGee JO, Athanasou NA. The human osteoclast precursor circulates in the monocyte fraction. Endocrinology 1998;137:4058-4068.
11. Higuchi S, Tabata N, Tajima M, et al. Induction of human osteoclast-like cells by treatment of blood monocytes with anti-fusion regulatory protein-1/CD98 monoclonal antibodies. J Bone Miner Res 1998;13:44-49.
12. Pacicci R. Estrogen, cytokines, and pathogenesis of postmenopausal osteoporosis. J Bone Miner Res 1996;11:1043-1051.
13. Jeon M, Hirayama T, Brown MA, et al. Osteoclast formation from circulating precursors in osteoporosis. Scand J Rheumatol 2003;32:95-100.
14. Stuss M, Sewerynek E, Król I, Stepien-Ktos W, Jędrzejczyk S. Assessment of OPG, RANKL, bone turnover markers serum levels and BMD after treatment with strontium ranelate and ibandronate in patients with postmenopausal osteoporosis. Endokrynol Pol 2016;67:174-184.

15. Gimble JM, Zvonic S, Floyd ZE, Kassem M, Nuttall ME. Identification of candidate genes in osteoporosis by integrated microarray analysis. Bone & Joint Research 2013;532:211-215.

16. Chan GK, Duque G. Age-related bone loss: old bone, new facts. Gerontology 2002;49:52-71.

17. Liu YZ, Dvornyk V, Lu Y, et al. A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes. J Bone Miner Res 2008;23:461-470.

18. Lei SF, Wu S, Li LM, et al. In vivo genome wide gene expression study of circulating monocytes suggested GBP1, STAT1 and CXCL10 as novel risk genes for the differentiation of peak bone mass. Bone 2009;44:1016-1014.

19. Siddiqui AS, Delaney AD, Schnerr C, et al. Sequence biases in large scale gene expression profiling data. Nucleic Acids Res 2008;36:e83-e93.

20. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 2013;41:D991-D995.

21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-5121.

22. Reimers M, Carey VJ. Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 2006;411:119-134.

23. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 2009;10:48.

24. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res 2012;40:W478-483.

25. Altermann E, Klaenhammer TR. Studying Biological Interactions in Yeast. Cold Spring Harb Protoc 2016 January 04. 2002;13:83-88.

26. Guevara-Cruz M, Lai CQ, Richardson K, et al. Effect of a GFOD2 variant on responses in total and LDL cholesterol in Mexican subjects with hypercholesterolemia after soy protein and soluble fiber supplementation. Gene 2013;522:211-215.

Funding Statement
This study was supported by a grant from Beijing High-level health technical personnel training plan (No: 2015-3-009).

Author Contribution
J. J. Li: Designed the study, Devised the methodology, Drafted the manuscript.
B. Q. Wang: Developed the search strategy, Screened the data sets for analysis, Assisted with manuscript preparation.
Q. Fei: Designed the study, Devised the methodology, Drafted the manuscript.
Y. Yang: Conducted data extraction and transformation.
D. Li: Performed data analysis.

ICMJE conflict of interest
None declared.

© 2016 Fei et al. This is an open-access article distributed under the terms of the Creative Commons Attribution license (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.