



## Protein-protein interaction network and significant gene analysis of osteoporosis

X.M. Wu<sup>1\*</sup>, X. Ma<sup>2\*</sup>, C. Tang<sup>1\*</sup>, K.N. Xie<sup>1</sup>, J. Liu<sup>1</sup>, W. Guo<sup>1</sup>, Y.L. Yan<sup>1</sup>,  
G.H. Shen<sup>1\*</sup> and E.P. Luo<sup>1\*</sup>

<sup>1</sup>School of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, China

<sup>2</sup>Department of Orthopaedics, First Affiliated Hospital of Medical School, Xi'an Jiaotong University, Xi'an, Shaanxi, China

\*These authors contributed equally to this study.

Corresponding authors: E.P. Luo / G.H. Shen

E-mail: luoerpingllepp@hotmail.com / ghlulu@fmmu.edu.cn

Genet. Mol. Res. 12 (4): 4751-4759 (2013)

Received March 4, 2013

Accepted August 30, 2013

Published October 18, 2013

DOI <http://dx.doi.org/10.4238/2013.October.18.12>

**ABSTRACT.** This study used DNA microarray data to identify differentially expressed genes of osteoporosis and provide useful information for treatments of the disease. We downloaded gene expression data of Osteoporosis GSE35956 from the Gene Expression Omnibus database, which included five normal and five osteoporosis samples. We then identified the differentially expressed genes between normal and disease samples using the R language software, and constructed the protein interaction network. DAVID was used to perform the biological process enrichment and KEGG pathway cluster analyses. We used the Cytoscape plug-in unit, Cluster ONE, to perform cluster module analysis to find hub proteins of the network module and to analyze their Gene Ontology (GO) functions. A total of 294 genes were found to be differentially expressed between normal and disease samples, which were used to construct the differential gene-protein interaction network. GO function analysis revealed that the genes' functions were mainly

involved in the intracellular signaling cascade. KEGG pathway analysis suggested that the main metabolic pathways of these genes were those of cancer: the neurotrophin/T cell/Fc epsilon RI/B cell/ErbB/p53 signaling pathway, the cell cycle pathway, and the chronic myeloid leukemia pathway. Screening analysis of hub proteins revealed that KRT18 had the highest hub degree. In conclusion, we found differentially expressed genes related to osteoporosis. GO biological process enrichment and KEGG pathway enrichment analyses identified significant osteoporosis genes and their molecular functions. Finally, module analysis of hub proteins in interaction networks showed that cell death was one of the main biological processes of osteoporosis genes.

**Key words:** DEGs; Protein-protein interaction network; Osteoporosis; Function analysis

## INTRODUCTION

Osteoporosis is a general bone disease characterized by the loss of bone mass and the degeneration of bone microstructure, which lead to an increased risk of bone fragility and fracture (Orlic et al., 2007). In western countries, 30% of women and 20% of men have had osteoporotic fractures, resulting in medical expenses in the United Kingdom the United States of up to 1.4 and 1.5 billion dollars, respectively (Cummings and Melton, 2002). As a developing country in Asia, China has a large population base, which is aging at a rapid rate, making osteoporosis and fractures an increasingly severe public health problem (Kung et al., 2007; Wang et al., 2012). Symptoms of osteoporosis patients are often missed, which can only be diagnosed once the bone is fractured. Thus, this disease is known as a “silent epidemic”. Therefore, research of osteoporosis-related genes is urgently needed and relevant.

Multiple factors regulate bone remodeling, such as hormones (estrogen, parathyroid hormone (PTH), vitamin D), interleukins (IL-1, IL-6, IL-11) and other cytokines (tumor necrosis alpha), and growth factors (bone morphogenetic proteins) (Troen, 2003). Nine differentially expressed genes were identified between bones of osteoporotic and non-osteoporotic women (Balla et al., 2008). In a recent study, the Trail pathway was shown to likely be involved in the pathogenesis of osteoporosis (Zhang et al., 2011). To date, most research related to the genetics of osteoporosis has focused on a single gene or a single pathway of bone metabolism, and only limited data are available about gene expression profiles of human osteoporotic bone tissues (Dvornyk et al., 2003).

In this study, DNA microarrays were used to identify differentially expressed genes between osteoporosis and normal bone marrow cells, with the aim of finding potential genes related to the pathogenesis of osteoporosis for future therapy. We performed an enrichment analysis of the metabolic pathway and screened the hub proteins. Together, results of this study will improve understanding of basic bone physiology in healthy and diseased individuals, which will help to develop new targets for therapeutic intervention and to identify challenges for clinical trial designs.

## MATERIAL AND METHODS

### Collection and arrangement of chip data

The Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) (Edgar et al., 2002) database is the largest and most comprehensive gene expression data source that can be easily accessed by users. Gene expression data in the GEO database are mainly provided by labs from around the world, and include Affymatrix chip data, expressed sequence tag (EST) data, serial analysis of gene expression (SAGE) sequence data, and second generation sequence data. We downloaded the gene expression data GSE35956 (Benisch et al., 2012) from the GEO database, which was based on the GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array platform. This data contained five normal senile marrow cell samples and five osteoporosis cell samples.

### Preprocessing of chip data and screening of differential expression genes

After obtaining the original data, they were divided into two groups: disease group (five osteoporosis samples) and control group (five normal samples). We used the R software (v. 2.13.0) (Team, 2004) platform to analyze the chip data. After normalizing the different chips using the Robust Multichip Averaging (RMA) method (Irizarry et al., 2003), we used the linear regression model package limma (Smyth, 2004) to classify chips into each group, and used the Bayes method (Benjamini and Hochberg, 1995) to correct for multiple testing. We chose a P-value  $< 0.05$  and  $|\log_{2}FC| > 2.5$  as significant thresholds to characterize differential expression in the disease group.

### Structure of differential gene-protein interaction network data

To obtain the protein interaction network of differentially expressed genes, we further constructed a protein-protein interaction network of osteoporosis using the following procedures:

(1) Collection of human protein interaction network data. We arranged protein interaction data from the MINT (Ceol et al., 2010), HPRD (Mishra et al., 2006) and BioGRID (Stark et al., 2011) databases using the method of screening the same protein interaction data in any two databases, and ultimately obtained 21,975 pairs of protein interactions.

(2) The differentially expressed osteoporosis genes were mapped into the collected human protein interaction data pairs using Perl programming, and the protein interaction network (PPI) of the differentially expressed genes was obtained. Cytoscape (Smoot et al., 2011) was used to visualize the PPI network.

### Enrichment analysis of GO function

The online tool DAVID (Huang et al., 2007) was used to perform biological process enrichment analysis of all of the genes in the protein interaction network in order to identify changes and to classify their functions at the cellular level. Only GO terms that contained more than two genes with a false discovery rate (FDR) value  $< 1.0E-04$  were chosen.

## Enrichment analysis of KEGG pathway

We focused on biological pathway method to explore changes of marrow cells in osteoporosis patients at the molecular level. We obtained all of the metabolic and non-metabolic pathways from the KEGG pathway database and used the DAVID website to perform KEGG pathway cluster analysis of differentially expressed genes (Huang et al., 2009a,b). In order to determine significant changes in signaling pathways of osteoporosis marrow cells, we only chose pathway terms that contained more than two genes with  $P < 1.0E-03$ .

## Screening analysis of hub proteins

Most of the known biological networks are scale-free networks, in which few hubs have a lot of connections, while most, which constitute the key hub of the network, have only a few connections (Lamb et al., 2006). We analyzed the hubs of the interaction network and used the scale-free property of the protein interaction network to find the key hub proteins. We then used the plug-in unit of Cytoscape (Smoot et al., 2011), Cluster ONE (Nepusz et al., 2012), to perform a cluster module analysis of the protein interaction network to find the hub proteins of the network module together with GO functional analysis of the hub proteins. These analytical results can further validate the molecular mechanism of osteoporosis and its potentially essential genes.

## RESULTS AND DISCUSSION

### Screening of differentially expressed genes

Using the limma package of R software with  $|\logFC| > 2.5$  and  $P < 0.05$  as thresholds, we ultimately obtained 294 differentially expressed genes between the normal and osteoporosis samples, including 265 upregulated differential genes and 29 downregulated differential genes.

### Structure of the differential gene-protein interaction network

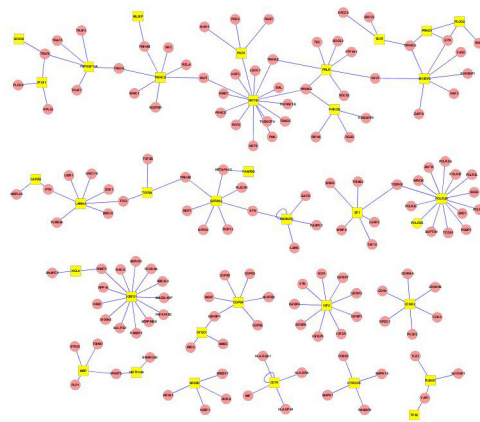
We screened all of the protein interaction pairs, which contained all of the 294 differentially expressed genes using a Perl program. We then filtered the networks with less than five hubs, and finally obtained a protein interaction network that contained 175 nodes and 168 edges, as shown in Figure 1.

### GO functional analysis

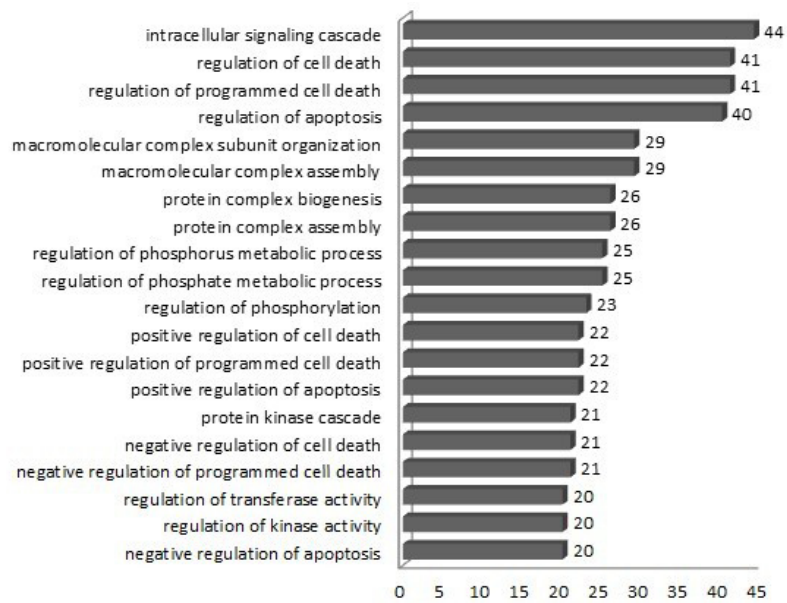
We chose  $FDR < 1.0E-04$  as the significance threshold for GO functional analysis of the 175 genes in the protein interaction network. The biological process (BP) enrichment results of these genes are shown in Figure 2.

We found that the gene functions were mainly enriched in intracellular signaling cascades, such as regulation of cell death/apoptosis, macromolecular/protein complex assembly, and regulation of phosphorylation/phosphate/phosphorus. In particular, the intracellular sig-

naling cascade of mitogen activated protein kinase (MARK) is rapidly initiated by female hormones (Filardo, 2002), which show close associations with bone mineral density (BMD) (Becherini et al., 2000). The differentially expressed estrogen-related receptor beta (*ESRRB*) gene is a female hormone-related gene. Bonewald (2004) found that the bone loss related to osteoporosis was partially due to osteocyte cell death, which was directly validated in the present study. Recent research has also revealed that the phosphor-proteome of human mesenchymal stem cells (hMSCs) shows dynamic changes during the osteogenic differentiation period, which can be potentially helpful for the treatment of osteoporosis or chondrography.



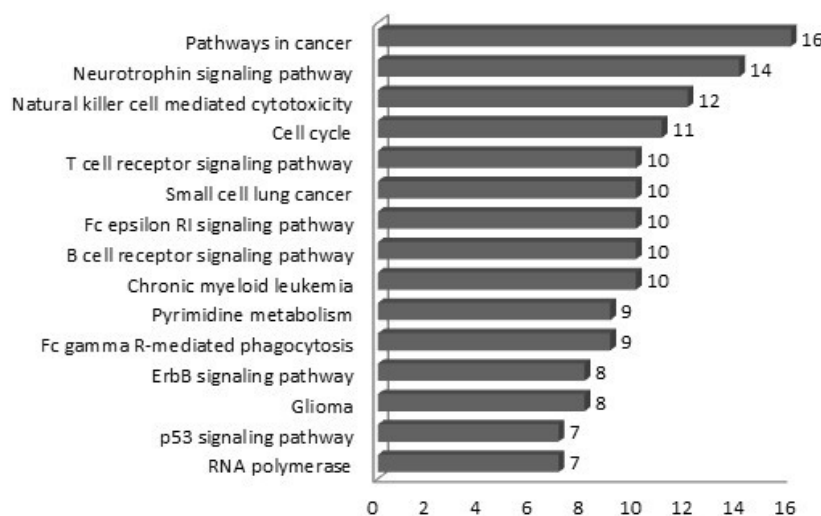
**Figure 1.** Interaction network of differential gene expressed proteins. Yellow plots stand for differential expression genes and pink plots stand for non-differential expression genes.



**Figure 2.** GO biological process enrichment analysis of genes in interaction network (FDR < 1.0E-04).

### Enrichment analysis of metabolic pathways

After analyzing cellular biological processes, we performed the enrichment analysis of metabolic pathways. We used the online tool DAVID to conduct the KEGG pathway analysis of the 175 genes in the protein interaction network, and chose metabolic pathways with P values  $< 1.0E-03$ . These results are presented in Figure 3.



**Figure 3.** KEGG pathway enrichment analysis of genes in interaction network (P value  $< 1.0E-03$ ).

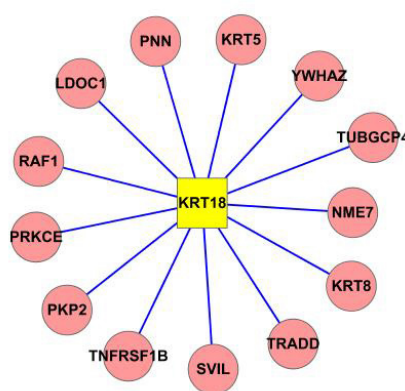
The enrichment results revealed that metabolic pathway-related genes were involved in pathways in cancer, the neurotrophin/T cell/Fc epsilon RI/B cell/ErbB/p53 signaling pathway, the cell cycle pathway, and the chronic myeloid leukemia pathway.

### Screening analysis of hub proteins

After calculating the hub degree of the interaction network formed by the differentially expressed gene products, we found that KRT18 had the highest hub degree (Nepusz et al., 2012). We then used the Cluster ONE plug-in unit of Cytoscape to conduct a cluster module analysis of the protein interaction network. We chose a minimum size of five, and a minimum density of 0.05 as thresholds to find the module in which the hub gene *KRT18* existed, as shown in Figure 4. Subsequently, we performed a GO function analysis of the genes in the module (enriched genes count  $> 2$ ,  $P < 0.05$ ), which revealed that cell death and cytoskeleton-related genes were most significant, as shown in Table 1. Cell death had previously been shown to be related to osteoporosis (Bonewald, 2004), as mentioned above. The cytoskeleton is the bracing structure of the cell, and recent research has revealed that changes in the expression of cytoskeleton-related genes are closely associated with bone loss and osteoporosis (Qian et al., 2009).

Several osteoporosis-related genes have been identified to date, including the osteoprotegerin (*OPG*) gene (Min et al., 2000), female hormone genes, and type I collagen alpha

1 chain (*COL1A1*) genes (Garcia et al., 2002). OPG is a member of the tumor necrosis factor (TNF) receptor superfamily. In the present study, another member of this family, the *TNFRSF1B* gene, was found in the KRT18 interaction network module, which suggests that it may be closely related to osteoporosis. Recently, Garcia et al. (2002) found two new polymorphic sites in the promoter region of the *COL1A1* gene. Oligonucleotides of the new sites were combined with primary osteoblast cell nuclear protein, which indicated that one of the sites was related to BMD. In the present study, the *COL10A1* gene, which belongs to the same family as the *COL1A1* gene, was differentially expressed in osteoporosis. Thus, this gene also has potential as a target for osteoporosis treatment.



**Figure 4.** Interaction network module of KRT18 protein.

**Table 1.** GO functional analysis of 14 genes in KRT18 interaction network (count > 2, P value < 0.05).

Category	Term	Description	Count	P value
GOTERM_BP	GO:0008219	Cell death	4	0.001981
GOTERM_BP	GO:0016265	Death	4	0.002118
GOTERM_CC	GO:0005856	Cytoskeleton	5	0.002456
GOTERM_CC	GO:0044430	Cytoskeletal part	4	0.00865
GOTERM_BP	GO:0007010	Cytoskeleton organization	3	0.017652
GOTERM_BP	GO:0006915	Apoptosis	3	0.022067
GOTERM_BP	GO:0012501	Programmed cell death	3	0.022955
GOTERM_CC	GO:0043232	Intracellular non-membrane-bounded organelle	5	0.029423
GOTERM_CC	GO:0043228	Non-membrane-bounded organelle	5	0.029423

Many of the screened differentially expressed genes found here have previously been found to be closely related with osteoporosis and BMD, such as the *IBSP* (Koller et al., 2010; Duncan et al., 2011), *TNFRSF11A* (Richards et al., 2009; Zupan et al., 2010; Guo et al., 2012), and *WNT4* (Zmuda et al., 2011) genes.

## CONCLUSION

The screening of genes related with osteoporosis has fundamental and applied relevance, as it is currently a common disease. The present study identified 294 differentially expressed genes between senile osteoporosis and normal samples. With information obtained

from three main databases, we constructed the differential expression gene-protein interaction network, which contained 175 genes. After GO biological process enrichment and KEGG pathway enrichment analyses, we identified the significant genes of osteoporosis along with their molecular functions.

GO enrichment function analysis was significantly characterized by intracellular signaling cascades and the regulation of cell death/apoptosis, which supports results of related studies (Filardo, 2002; Bonewald, 2004). The results of KEGG pathway enrichment revealed that cancer and cell signaling pathways were significant metabolic pathways for osteoporosis-related genes.

Finally, the module analysis revealed that the main biological processes of hub proteins in interaction networks were related to cell death and the cytoskeleton of cellular components. This result also supports a previous study (Qian et al., 2009), thus verifying the accuracy of the present study. Furthermore, we here identified the differentially expressed *THFRSF1B*, *ESRRB*, *COL10A1*, *IBSP*, *TNFRSF11A*, and *WNT4* genes as potential key genes of osteoporosis.

## ACKNOWLEDGEMENTS

Research supported by funding of a Preliminary Study on the Complex Impedance Frequency Response of Osteoporosis Bones Tissues, National Science Foundation of China (#50577067).

## REFERENCES

- Balla B, Kosa JP, Kiss J, Borsy A, et al. (2008). Different gene expression patterns in the bone tissue of aging postmenopausal osteoporotic and non-osteoporotic women. *Calcif. Tissue Int.* 82: 12-26.
- Becherini L, Gennari L, Masi L, Mansani R, et al. (2000). Evidence of a linkage disequilibrium between polymorphisms in the human estrogen receptor alpha gene and their relationship to bone mass variation in postmenopausal Italian women. *Hum. Mol. Genet.* 9: 2043-2050.
- Benisch P, Schilling T, Klein-Hitpass L, Frey SP, et al. (2012). The transcriptional profile of mesenchymal stem cell populations in primary osteoporosis is distinct and shows overexpression of osteogenic inhibitors. *PLoS One* 7: e45142.
- Benjamini Y and Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. Series B (Methodological)* 57: 289-300.
- Bonewald LF (2004). Osteocyte biology: its implications for osteoporosis. *J. Musculoskelet. Neuronal. Interact.* 4: 101-104.
- Ceol A, Chatr AA, Licata L, Peluso D, et al. (2010). MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.* 38: D532-D539.
- Cummings SR and Melton LJ (2002). Epidemiology and outcomes of osteoporotic fractures. *Lancet* 359: 1761-1767.
- Duncan EL, Danoy P, Kemp JP, Leo PJ, et al. (2011). Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet.* 7: e1001372.
- Dvornyk V, Recker RR and Deng HW (2003). Gene expression studies of osteoporosis: implications for microarray research. *Osteoporos. Int.* 14: 451-461.
- Edgar R, Domrachev M and Lash AE (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30: 207-210.
- Filardo EJ (2002). Epidermal growth factor receptor (EGFR) transactivation by estrogen via the G-protein-coupled receptor, GPR30: a novel signaling pathway with potential significance for breast cancer. *J. Steroid. Biochem. Mol. Biol.* 80: 231-238.
- Garcia T, Roman-Roman S, Jackson A, Theilhaber J, et al. (2002). Behavior of osteoblast, adipocyte, and myoblast markers in genome-wide expression analysis of mouse calvaria primary osteoblasts *in vitro*. *Bone* 31: 205-211.
- Guo Y, Wang JT, Liu H, Li M, et al. (2012). Are bone mineral density loci associated with hip osteoporotic fractures? A validation study on previously reported genome-wide association loci in a Chinese population. *Genet. Mol. Res.* 11: 202-210.



- Huang da W, Sherman BT, Tan Q, Collins JR, et al. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8: R183.
- Huang da W, Sherman BT and Lempicki RA (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37: 1-13.
- Huang da W, Sherman BT and Lempicki RA (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264.
- Koller DL, Ichikawa S, Lai D, Padgett LR, et al. (2010). Genome-wide association study of bone mineral density in premenopausal European-American women and replication in African-American women. *J. Clin. Endocrinol. Metab.* 95: 1802-1809.
- Kung AW, Lee KK, Ho AY, Tang G, et al. (2007). Ten-year risk of osteoporotic fractures in postmenopausal Chinese women according to clinical risk factors and BMD T-scores: a prospective study. *J. Bone Miner. Res.* 22: 1080-1087.
- Lamb J, Crawford ED, Peck D, Modell JW, et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929-1935.
- Min H, Morony S, Sarosi I, Dunstan CR, et al. (2000). Osteoprotegerin reverses osteoporosis by inhibiting endosteal osteoclasts and prevents vascular calcification by blocking a process resembling osteoclastogenesis. *J. Exp. Med.* 192: 463-474.
- Mishra GR, Suresh M, Kumaran K, Kannabiran N, et al. (2006). Human protein reference database - 2006 update. *Nucleic Acids Res.* 34: D411-D414.
- Nepusz T, Yu H and Paccanaro A (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* 9: 471-472.
- Orlic I, Borovecki F, Simic P and Vukicevic S (2007). Gene expression profiling in bone tissue of osteoporotic mice. *Arh. Hig. Rada Toksikol.* 58: 3-11.
- Qian A, Di S, Gao X, Zhang W, et al. (2009). cDNA microarray reveals the alterations of cytoskeleton-related genes in osteoblast under high magneto-gravitational environment. *Acta Biochim. Biophys. Sin. (Shanghai)* 41: 561-577.
- Richards JB, Kavvoura FK, Rivadeneira F, Stykarsdottir U, et al. (2009). Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern. Med.* 151: 528-537.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432.
- Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article3.
- Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, et al. (2011). The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* 39: D698-D704.
- Team RDC (2004). R: A Language and Environment for Statistical Computing.
- Troen BR (2003). Molecular mechanisms underlying osteoclast formation and activation. *Exp. Gerontol.* 38: 605-614.
- Wang C, Zhang Z, Zhang H, He JW, et al. (2012). Susceptibility genes for osteoporotic fracture in postmenopausal Chinese women. *J. Bone Miner. Res.* 27: 2582-2591.
- Zhang YP, Liu YZ, Guo Y, Liu XG, et al. (2011). Pathway-based association analyses identified TRAIL pathway for osteoporotic fractures. *PLoS One* 6: e21835.
- Zmuda JM, Yerges-Armstrong LM, Moffett SP, Klei L, et al. (2011). Genetic analysis of vertebral trabecular bone density and cross-sectional area in older men. *Osteoporos. Int.* 22: 1079-1090.
- Zupan J, Mencej-Bedrac S, Jurkovic-Mlakar S, Prezelj J, et al. (2010). Gene-gene interactions in RANK/RANKL/OPG system influence bone mineral density in postmenopausal women. *J. Steroid. Biochem. Mol. Biol.* 118: 102-106.