Published in final edited form as:

J Proteomics. 2016 June 16; 142: 45–52. doi:10.1016/j.jprot.2016.04.044.

Quantitative proteomics and integrative network analysis identified novel genes and pathways related to osteoporosis

Yong Zeng^{1,2}, Lan Zhang², Wei Zhu^{2,3}, Chao Xu², Hao He², Yu Zhou², Yao-Zhong Liu², Qing Tian², Ji-Gang Zhang², Fei-Yan Deng^{2,4}, Hong-Gang Hu¹, Li-Shu Zhang¹, and Hong-Wen Deng^{1,2}

¹College of Life Sciences and Bioengineering, Beijing Jiao Tong University, Beijing, China, 100044

²Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, USA, 70112

³College of Life Sciences, Hunan Normal University, Changsha, Hunan, China 410081

⁴Laboratory of Proteins and Proteomics, Department of Epidemiology, Soochow University School of Public Health, Suzhou, Jiangsu, China, 205123

Abstract

Osteoporosis is mainly characterized by low bone mineral density (BMD), and can be attributed to excessive bone resorption by osteoclasts. Migration of circulating monocytes from blood to bone is important for subsequent osteoclast differentiation and bone resorption. Identification of those genes and pathways related to osteoclastogenesis and BMD will contribute to a better understanding of the pathophysiological mechanisms of osteoporosis. In this study, we applied the LC-nano-ESI-MS^E (Liquid Chromatograph-nano-Electrospray Ionization-Mass Spectrometry) for quantitative proteomic profiling in 33 female Caucasians with discordant BMD levels, with 16 high vs. 17 low BMD subjects. Protein quantitation was accomplished by label-free measurement of total ion currents collected from MS^E data. Comparison of protein expression in high vs. low BMD subjects showed that ITGA2B ($p=0.0063$) and GSN ($p=0.019$) were up-regulated in the high BMD group. Additionally, our protein-RNA integrative analysis showed that RHOA ($p=0.00062$) differentially expressed between high vs. low BMD groups. Network analysis based on multiple tools revealed two pathways: “Regulation of actin cytoskeleton” ($p=1.13E-5$, FDR=3.34E-4) and “Leukocyte transendothelial migration” ($p=2.76E-4$, FDR=4.71E-3) that are functionally relevant to osteoporosis. Consistently, ITGA2B, GSN and RHOA played crucial roles in these two pathways respectively. All together, our study strongly supported the contribution of the genes ITGA2B, GSN and RHOA and the two pathways to osteoporosis risk.

Corresponding Author: Hong-Wen Deng, Ph.D., School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 2001, New Orleans, LA 70112, hdeng2@tulane.edu.

All authors state that they have no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Graphical Abstract

Keywords

Osteoporosis; Peripheral blood monocytes; Quantitative Proteomics; Pathways; Network analysis

INTRODUCTION

Osteoporosis is a global public health problem with a high heritability, and mainly characterized by low BMD [1]. Osteoporosis has become one of the most serious public health problems around the world, leading to millions of fractures annually [1, 2]. Among fractures at various skeletal sites, hip fracture is the most important owing to its high prevalence, high morbidity and mortality, and excessive therapeutic cost [3, 4].

Monocytes are bone marrow-derived circulating leukocytes that can further differentiate into various cell types like macrophages and dendritic cells [5]. The morphology of mature monocytes in the peripheral circulation is heterogeneous, and these cells constitute 5–10% of peripheral blood leukocytes in humans [6]. Human peripheral blood monocytes (PBMs) may serve as precursors of osteoclasts and produce cytokines important for osteoclast differentiation, activation, and apoptosis as well as acting as one of the most important target cells for sex hormones in bone metabolism [7, 8]. In the bone field, PBMs have already been well-established/accepted as a well working excellent cell model for studying gene/protein expression patterns and their regulation mechanisms in relation to osteoporosis risk in vivo in humans [9–14]. Numerous studies highlighted the feasibility and utility of employing in vivo PBMs to study novel pathophysiological mechanisms during the process of osteoclastogenesis underlying osteoporosis risk [8, 12]. Therefore, PBMs are one major class of cells that are functionally relevant to the pathogenesis of osteoporosis and they have been successfully used for etiology studies in the bone field.

Network analysis can identify the correlation and topology between different proteins. Pathways/modules generated by network analysis may reflect the biological processes more comprehensively and objectively than single protein/gene analysis [15]. Based on the utilization of functional information and topological information, there are quite some approaches to perform gene enrichment analysis, pathway analysis, functional cluster analysis and network reconstruction analysis [16]. For instance, DAVID (Database for Annotation, Visualization and Integrated Discovery) is a popular knowledgebase including multiple online tools which can provide abundant functional information from multiple

databases for a list of genes. However, the topological features generated from DAVID may be quite limited [16, 17]. In contrast, as a frequently-used platform for network analyses, the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database engine can provide more topological information [16, 18]. Furthermore, Cytoscape, an open source software project for data integration, network analysis and visualization [19], is a powerful tool with hundreds of comprehensive and timely-updated applications which can provide multiple functions of network analyses.

With sensitive LC-nano-ESI-MS^E (Liquid Chromatograph-nano-Electrospray Ionization-Mass Spectrometry) based quantitative proteomic analysis, we identified a number of proteins that contribute to the pathogenesis of osteoporosis individually. Different kinds of tools were then used for network analysis to maximize the coverage of significant biological information related to bone metabolism. DAVID was used for the preliminary functional annotation and gene enrichment analysis. STRING and Cytoscape were applied for comprehensive network reconstruction and visualization. Additionally, in order to add more evidence to the results from our quantitative proteomic study, an *in silico* replication study was performed at genome-wide level.

MATERIALS AND METHODS

Human Subjects

Our study was approved by Institutional Review Boards of University of Missouri Kansas City and Tulane University. All the subjects signed consent forms before being enrolled into this study. All subjects were self-identified as European Caucasian females. 33 subjects were recruited in this study. The lumbar spine BMD (g/cm²) and hip BMD (which is the combined value of three regions including femoral neck, trochanter and intertrochanter) were measured by using Hologic 4500 W dual energy X-ray bone densitometer for each subject as we routine do [20, 21].

In order to minimize the influence of environmental factors that may affect bone metabolism, several chronic diseases and conditions were excluded. The detailed exclusion criteria were described in previous studies [12], which mainly include chronic disorders of vital organs, autoimmune-related diseases, metabolic diseases, skeletal diseases, hematopoietic diseases, lymphoreticular diseases and other diseases or any treatment that would affect gene/protein expression.

33 unrelated Caucasian premenopausal women were included in this study, 16 with high-BMD (Z score: 1.32±0.45) and 17 with low-BMD (Z score: -0.96±0.34). The basic clinical information is shown in Table 1.

Peripheral Blood Monocytes (PBMs) Isolation

60 ml of peripheral blood was collected from each subject by a certificated phlebotomist. The fresh blood samples were processed instantly for PBMs isolation. First, peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using density gradient centrifugation with Histopaque-1077 (Sigma, Catalog No. H1077-1). Then, PBMs were isolated from PBMCs using a monocyte negative isolation kit (DynaL Biotech Inc.)

following the manufacturer's recommendation. The kit contains a highly optimized antibody mix, blocking reagent, and Depletion Dynabeads® to deplete T cells, B cells, and natural killer cells from PBMCs, leaving monocytes untouched and free of surface bound antibody and beads. The purity of PBMs isolated using this method was $86\% \pm 3\%$, similar to our earlier experience [8].

Protein Extraction and Sample Preparation

PBMs total proteins were extracted using a complete proteome extraction mammalian kit (Calbiochem Catalog No. 539779). Protein concentration was measured using the Bradford method. Up to 20 µg total protein was precipitated using a protein precipitation kit (Calbiochem Catalog No. 539180). Protein pellets were dissolved in 50 µl 50 mM ammonium bicarbonate with 0.1% RapiGest (Waters), reduced by 5.0 mM dithiothreitol, alkylated by 15 mM indole acetic acid, and then digested by trypsin (Sigma, Catalog No. T6567). Protein digests were concentrated to 20 µl, of which 15 µl was aspirated and brought back to 20 µl, with 0.5% formic acid (FA) and 100 fmol yeast ADH1 (Alcohol Dehydrogenase I) digestion standard (Waters, Catalog No. 186002328).

PBM Proteome Profiling through UPLC-nano-ESI-MS/MS^E Analyses

The proteome of PBM total protein was profiled through nanoAcquity ultra performance liquid chromatography coupled with Synapt High Definition Mass Spectrometry (HDMS) (Waters) and the data acquisition process was controlled by MassLynx 4.1 software (Waters). Briefly, protein digests (~0.5 µg) were injected into a BEH C18 75 µm x 150 mm analytical column, the particle size of the column is 1.7 µm. Subsequently, the protein digests were separated by solvent A (water with 0.1% FA) and solvent B (acetonitrile with 0.1% FA) at a flow rate of 0.3 µl/min using a 2 hours' gradient as follows: 3% B initial, 10% B at 1.0 min, 30% B at 75 min, 40% B at 90 min, 95% B at 91 min, 95% B at 95 min, 3% B at 96 min, equilibrate thereafter till 120 min. The eluate was analyzed by HDMS under positive ion V-mode. The following parameters were set for data acquisition: collision energy: 5 volts for MS and ramp 15–40 volts for MS^E; scan time: 0.6 second per scan. The HDMS machine was calibrated daily to ensure high accuracy (2.0 ppm for lock mass of m/z 785.8426). Then, the acquired MS^E data were processed with ProteinLynx Global Server (PLGS) v2.4 (Waters) using the appropriate parameters recommended by Waters Corporation. Specific information regarding database searching and protein quantification based on PLGS can be found in our previous publications [e.g., 12, 22, 23]. Protein identification was achieved by searching the Human protein database International Protein Index version 3.83 (153,078 protein entries). Triplicate LC-nano-ESI-MS^E experiments and analyses were performed for each sample. For each sample, only proteins identified at least twice in the triplicate LC-nano-ESI-MS^E analyses were reported as truly present. Total ion counts of the top three matched peptides were used to quantify each protein. We added 100 femtomol of ADH1 in each prepared sample as internal references (recommended by Waters Corporation). By comparing the ion counts of each protein with ADH1, protein quantification level was exported in femtomol and nanogram. Mean values from triplicate analyses were used to represent protein expression levels in each PBM sample.

Identify differentially-expressed proteins (DEPs)

In order to detect the significantly DEPs, the proteomics dataset was normalized by the housekeeping protein beta actin first. After this, by using the normalized data, we performed student t-test in subjects with low vs. high BMD. P value less than 0.05 was set as the nominal significant cut off.

Protein-RNA integrative analysis

22 out of 33 samples in our study have both proteomics and RNA expression data. The RNA expression data was generated from microarray analyses using Affymetrix 1.0 ST arrays. Our group published a paper previously which gave a specific introduction to the dataset [24]. In order to synthesize the information obtained from these two datasets for each protein coding gene, we used an algorithm to compute crosstalk scores, which combine the disease-association evidence obtained from individual uni-omics studies through a meta-analysis based approach [25].

First, we determine gene-to-phenotype association scores separately for each type of individual uni-omics data. For gene g , compute scores $S_g^k(X_k, Y)$ ($k=1, \dots, D$, where D is the number of omics data), capturing the relationship between different genomic measurements X_k (e.g., from proteomic and RNA transcriptomics) and a phenotype Y . Here $S_g^k(X_k, Y)$ is a t-test statistic that measures the differentiation between high and low BMD groups for k -th omic data; Second, integrate the gene-to-phenotype association scores

$(S_g^1(X_1, Y), \dots, S_g^D(X_D, Y))$ into a gene-specific score $(S_{g-\text{meta}}(S_g^1, S_g^2, \dots, S_g^k))$ by a meta-analysis based approach [25]: $S_{g-\text{meta}}(S_g^1, S_g^2, \dots, S_g^k) = \sum_{k=1}^D |Z_g^k|$ ($k \in (1, \dots, D)$) (1),

$Z_g^k = \frac{S_g^k(X_k, Y)}{\text{sd}(S_g^k)}$ is the standardized score of gene g in the data set k , and $\text{sd}(S_g^k)$ was derived by permutation under the null hypothesis that there is no association between gene g and Y in k -th omic data. When there is no association between gene g and phenotype Y , $|Z_g^k|$ follows a Half-normal distribution. For two omics data, the cumulative density function of $S_{g-\text{meta}}$ is defined as: $F(s) = \int \int_{-\infty}^{x+y \leq s} f(x, y) dx dy = [2\Phi(\frac{\sqrt{2}}{2}s) - 1]^2$ (2). Finally, a P-value for the combined score was adjusted using false discovery rate (FDR) in order to correct for multiple hypothesis testing [26].

DAVID: genes enrichment analysis and functional annotation

The DAVID Knowledgebase agglomerated tens of millions of gene/protein identifiers from a variety of public genomic resources; these identifiers improve the cross-reference capability, particularly across NCBI and UniProt systems. More than 40 publicly available functional annotation sources have been comprehensively integrated and centralized by the DAVID gene clusters [17, 27]. In order to maximize the coverage of important biological information, we imported all the unique gene symbols into this online tool for functional annotation and enrichment analysis. Comprehensive functional information for all of these genes enriched in specific terms has been fully exposed. The results were classified as “functional annotation clustering”, “functional annotation chart” and “functional annotation table”.

STRING and Cytoscape: pathway analysis and network reconstruction and visualization

In this study, we combined all the DEPs both from proteomic profiling and protein-RNA integrative analysis as the target gene list. Subsequently, the target gene list was imported into STRING and Cytoscape respectively for pathway analysis, network reconstruction and visualization. It's well known that the interactions between proteins represent a crucial component for modern biology such as proteomics study. With a user-friendly and comprehensive manner, the STRING knowledge database has been designed to assemble, evaluate and disseminate protein protein association information [18]. In addition, ClueGO in conjunction with CluePedia, two easy-to-use applications based on the most updated version of Cytoscape (Cytoscape_v3.2.0) strongly improved biological interpretation of large lists of genes, consequently provided comprehensive view from different perspectives [28, 29].

RESULTS

Proteomic profiling

In this study, a total of 2058 proteins were identified in 33 samples by LC-nano-ESI-MS^E. To reduce false proteins detection, 1142 out of 2058 proteins, which were detected in 5 or more subjects were used for further analyses [12] (Supplemental Table 1). Furthermore, a global normalization method was applied for data quality control in order to minimize the variability due to sample preparation or equipment conditions. We applied log₂ transformation of our data for further analyses.

Identification of Differentially-expressed proteins (DEPs)

We applied Student's t-test to detect DEPs in low and high-BMD groups. A total of 30 unique proteins were identified as DEPs ($p < 0.05$). Within these 30 proteins, 24 proteins are up-regulated in high-BMD group and 6 proteins are down-regulated in high-BMD group (Figure 1 and Table 2).

Protein-RNA integrative analysis

In our study, 22 (9 high BMD vs 13 low BMD) samples have both proteome data and RNA expression data. Integrative analysis was applied to evaluate the gene expression level in combined RNA-Protein dataset. 95 genes were differentially expressed in RNA-Protein integrative analysis ($q < 0.05$) (Supplemental Table 2).

Gene enrichment analysis and functional annotation

Corresponding to the total of 1142 proteins, 403 unique gene symbols were imported into DAVID online tool for functional annotation and gene enrichment analysis. A total of 987 terms were generated. From those terms, with an integrated consideration of statistical results and physiological features, We highlighted several significant terms ($p < 0.001$) closely related to osteogenesis, bone metabolism and skeletal physiology, such as “actin cytoskeleton organization”, “generation of precursor metabolites and energy”, “leukocyte migration”, “calcium ion binding” and “calcium”. Individually, the DEPs included in selected terms were pointed out (Table 3).

Pathway analysis, network reconstruction and visualization

As the target gene list, 115 genes including all the DEPs both from single proteomic study and protein-RNA analysis was imported into STRING and Cytoscape for pathway analysis, network reconstruction and visualization.

A total of 118 pathways were generated from STRING analysis. In order to narrow down the list and enhance the reliability of the result, we set a strict statistical threshold ($p < 0.01$) for filtering. 30 pathways passed the threshold (Supplemental Table 3). Among these 30 pathways, we selected two “regulation of actin cytoskeleton” ($p = 1.13\text{E-}5$, $\text{FDR} = 3.34\text{E-}4$) and “leukocyte transendothelial migration” ($p = 2.76\text{E-}4$, $\text{FDR} = 4.71\text{E-}3$) based on their functional relevance to bone metabolism. 8 and 5 genes out of our target gene list were enriched in these two pathways, respectively (Table 4). The multiple interactions of those genes enriched in the selected two pathways were represented by different color lines (Figure 2).

Based on Cytoscape platform, the plug-in ClueGO combined with CluePedia created a dynamic and visualized network. Only significant pathways or terms were presented by setting the statistical threshold ($p < 0.05$). All the pathways were functionally grouped and highlighted with different colors (Figure 3, a). Herein we found that most of the pathways were functionally related with the selected two interesting pathways mentioned above (marked with red rectangle in Figure 3, a). A visualized network for bright green marked pathways and the involved genes was generated (Figure 3, b). We intentionally marked the two interesting pathways in yellow (should be green originally) and the three interesting genes with red rectangles.

DISCUSSION

With its potential to differentiate into osteoclasts and the production of potent pro-inflammatory and anti-inflammatory cytokines, PBMs are likely to play broad and critical roles in bone metabolism [30, 31]. Abnormal behavior in PBM formation, migration, differentiation, functional activity, and apoptosis may be closely related to the development of human skeletal disorders [24]. In particular, it has now been well established by numerous studies that multi-omics and molecular studies using PBMs as a working cell model may identify novel susceptibility genes/pathways and provide novel insights into the pathophysiological mechanisms underlying various skeletal disorders [8, 12, 24].

Bone metabolism is a very complex biological process which can be influenced and regulated by both genetic and environmental factors [32, 33]. The balance of bone metabolism is mainly depends on the dynamic between bone remodeling and bone resorption. Essentially, osteoblast and osteoclast are the most important cells participating in these two physiological processes. Osteoblasts originate from mesenchymal stem cells (osteoprogenitor cells) of the bone marrow stroma and are responsible for bone matrix synthesis and its subsequent mineralization [34, 35]. Commitment of mesenchymal stem cells to the osteoblast lineage requires the canonical Wnt signaling pathway and associated proteins [36, 37]. As the only cells capable of resorbing bone, osteoclasts are formed from hematopoietic precursors of the monocytes and macrophage series. Osteoclast differentiation

is supported and regulated by membrane-bound receptor activator of NF- κ B ligand (RANKL) and macrophage-colony stimulating factor (M-CSF) [38, 39]. Based on the well-established cellular mechanism of bone metabolism and the statistics analysis results of the data, our study highlighted the pathways: “Regulation of actin cytoskeleton” and “Leukocyte transendothelial migration”, which are functionally relevant to bone metabolism.

“Regulation of actin cytoskeleton”, is a term that is defined as the biological process that modulates the frequency, rate or extent of the formation, arrangement of constituent parts, or disassembly of cytoskeletal structures comprising actin filaments and their associated proteins (GO description). Proteins included in this pathway perform diverse effects on cell activity, including changes in cell shape, migration, proliferation, and survival in many cell types [40]. A previous study showed that osteoclast undergoes internal structure changes such as rearrangements of actin cytoskeleton that prepare it to resorb bone [38]. Another study suggested that the status of actin cytoskeleton signaling network determines the size of osteoclast-like cells during cell fusion [41]. As one of the top 10 pathways in our study, 8 genes out of the target gene list were enriched in the pathway “Regulation of actin cytoskeleton”.

Cell migration is required for many biological processes, such as embryonic morphogenesis, immune surveillance, and tissue formation, repair and regeneration [40]. Leukocyte migration from the blood into tissues is vital for immune surveillance and inflammation. In this process, α L/ β 2 integrin activates through Vav1, RhoA, which subsequently activates the kinase p160 ROCK. ROCK activation leads to MLC phosphorylation, resulting in retraction of the actin cytoskeleton [42, 43]. It is well known that there is a close relationship between immune system and bone tissue [44, 45]. Monocyte, one type of leukocyte as well as the potential precursor of osteoclasts, can be activated and then migrate from peripheral blood to bone tissue and differentiate into osteoclasts [6]. In addition, calcium ion, which play a central role in the whole process of this pathway has also been considered to be the most important cation for bone metabolism [46]. All the direct or indirect evidences support our conclusion that the pathway “Leukocyte transendothelial migration” is highly related with bone metabolism.

There is a close and interactive relationship between these two pathways in biological function. Previous studies indicated that regulation of actin cytoskeleton played a significant role in cell migration [40, 47]. Many genes included in these two pathways are involved in bone metabolism. For instance, RHOA, ITGA2B, GSN play crucial roles in specific molecular and biological processes such as calcium binding and transporting, osteogenesis, bone remodeling and resorption. RHOA (ras homolog family member A) is a protein-coding gene that is differentially expressed in the protein-RNA integrative analysis in our study ($p=6.24E-4$). RHOA regulates a signal transduction pathway linking plasma membrane receptors to the assembly of focal adhesions and actin stress fibers. It plays an essential role in many important pathways including “regulation of actin cytoskeleton”, “leukocyte transendothelial migration” and “Wnt signaling pathway” in human body. The first two pathways have already been discussed in the above. As for the “Wnt signaling pathway”, it is well-known as the key pathway required for osteoblastogenesis [36]. Previous studies demonstrated that RHOA is critical for osteoblast differentiation, osteoclast podosome

organization, motility, and bone resorption [48]. A knockdown study identified RHOA as a potential regulator of a number of genes in bone cells [49]. Our study adds further evidence to the previous studies on RHOA, suggesting an essential role of RHOA in bone metabolism. ITGA2B (integrin alpha 2b) encodes integrin alpha chain 2b. Integrin has a significant role in cell adhesion, cell signaling, cell differentiation and cell migration. One previous study indicated that: focal adhesion, actin cytoskeleton and cell-cycle are connected pathways and their genes including ITGA2B are implicated in the pathogenesis of low BMD [50]. Another study in Glanzmann thrombasthenia patients suggested a significant role of ITGA2B in bone defects and metastasis [51, 52]. As a differentially expressed protein in our study ($p=0.00632$), ITGA2B was up-regulated in high BMD group (Table 2). Meanwhile, in gene ontology analysis, ITGA2B was included in terms such as “cytoskeleton organization”, “calcium binding” and “calcium”, which are regarded to be closely associated with bone metabolism (Table 3). Both evidences from individual protein differential expression analysis and entire proteins network analysis supported that ITGA2B may play an important role in bone metabolism including matrix mineralization and bone resorption. Previous study from our group already reported that GSN (Gelsolin) may be significant for hip BMD in female Caucasian [22]. By combining modern proteomics approach with comprehensive network analysis, this study provided novel evidences on whole protein level to support our previous conclusion.

To gain additional support for the genes identified from proteome-based network analysis in our discovery study, we assessed associations with BMD at the population level using a previously published genome-wide association (GWA) meta-analysis study [53]. The study included 27,061 subjects, and independently confirmed 13 previously reported loci important for BMD variation [53].

Seven individual GWAS datasets were included in this GWA meta-analysis. The meta-analysis revealed that RHOA and GSN were associated with hip BMD. 17 SNPs in gene RHOA and 2 SNPs in gene GSN were associated with hip BMD ($p<0.01$). The results from GWAS study add further evidences to our current conclusion.

CONCLUSIONS

In our study, label-free based quantitative proteomics integrated with network analysis strongly supported the contribution of the genes ITGA2B, GSN and RHOA and the pathways “Regulation of actin cytoskeleton” and “Leukocyte transendothelial migration” to osteoporosis risk. The results were further verified in multiple level studies including protein-RNA integrative analysis and genome wide association study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was partially supported by grants from NIH (P50AR055081, R21AG27110, R01AR057049, R01AR050496, and R01AG026564), Franklin D. Dickson/Missouri Endowment and Edward G. Schlieder

Endowment. The study also benefited from the National Natural Science Foundation of China (31371194) and the Fundamental Research Funds from the Central Universities (2013JBM098).

ABBREVIATIONS

BMD	Bone Mineral Density
LC	Liquid Chromatograph
ESI	Electrospray Ionization
MS	Mass Spectrometry
PBMs	Peripheral Blood Monocytes
PBMC	Peripheral Blood Mononuclear Cell
DAVID	Database for Annotation, Visualization and Integrated Discovery
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
DEPs	Differentially Expressed Proteins
FDR	False Discovery Rate
RANKL	Receptor Activator of Nuclear factor Kappa-B Ligand
GO	Gene Ontology
GWA	Genome Wide Association

References

1. Zhao LJ, Liu XG, Liu YZ, Liu YJ, Papasian CJ, Sha BY, et al. Genome-wide association study for femoral neck bone geometry. *J Bone Miner Res*. 2010; 25:320–9. [PubMed: 20175129]
2. Liu YJ, Shen H, Xiao P, Xiong DH, Li LH, Recker RR, et al. Molecular genetic studies of gene identification for osteoporosis: a 2004 update. *J Bone Miner Res*. 2006; 21:1511–35. [PubMed: 16995806]
3. Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet*. 2002; 359:1761–7. [PubMed: 12049882]
4. Cauley JA, Thompson DE, Ensrud KC, Scott JC, Black D. Risk of mortality following clinical fractures. *Osteoporos Int*. 2000; 11:556–61. [PubMed: 11069188]
5. Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells. *Science*. 2010; 327:656–61. [PubMed: 20133564]
6. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol*. 2005; 5:953–64. [PubMed: 16322748]
7. Narni-Mancinelli E, Soudja SM, Crozat K, Dalod M, Gounon P, Geissmann F, et al. Inflammatory monocytes and neutrophils are licensed to kill during memory responses in vivo. *PLoS Pathog*. 2011; 7:e1002457. [PubMed: 22241983]
8. Liu YZ, Dvornyk V, Lu Y, Shen H, Lappe JM, Recker RR, et al. A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes. *J Biol Chem*. 2005; 280:29011–6. [PubMed: 15965235]
9. Bonjour JP, Chevalley T, Ferrari S, Rizzoli R. The importance and relevance of peak bone mass in the prevalence of osteoporosis. *Salud publica de Mexico*. 2009; 51(Suppl 1):S5–17. [PubMed: 19287894]

10. Reynolds LM, Taylor JR, Ding JZ, Lohman K, Johnson C, Siscovick D, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nature Communications*. 2014; 5
11. Talens RP, Christensen K, Putter H, Willemsen G, Christiansen L, Kremer D, et al. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging cell*. 2012; 11:694–703. [PubMed: 22621408]
12. Deng FY, Lei SF, Zhang Y, Zhang YL, Zheng YP, Zhang LS, et al. Peripheral blood monocyte-expressed ANXA2 gene is involved in pathogenesis of osteoporosis in humans. *Mol Cell Proteomics*. 2011; 10:M111011700.
13. Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A, et al. Methylomics of gene expression in human monocytes. *Hum Mol Genet*. 2013; 22:5065–74. [PubMed: 23900078]
14. Ralston SH, Galwey N, MacKay I, Albagha OM, Cardon L, Compston JE, et al. Loci for regulation of bone mineral density in men and women identified by genome wide linkage scan: the FAMOS study. *Hum Mol Genet*. 2005; 14:943–51. [PubMed: 15746152]
15. Swa HL, Shaik AA, Lim LH, Gunaratne J. Mass spectrometry based quantitative proteomics and integrative network analysis accentuates modulating roles of annexin-1 in mammary tumorigenesis. *Proteomics*. 2014
16. Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. *J Theor Biol*. 2014
17. Sherman BT, Huang da W, Tan Q, Guo Y, Bour S, Liu D, et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*. 2007; 8:426. [PubMed: 17980028]
18. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41:D808–15. [PubMed: 23203871]
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504. [PubMed: 14597658]
20. Liu XG, Tan LJ, Lei SF, Liu YJ, Shen H, Wang L, et al. Genome-wide Association and Replication Studies Identified TRHR as an Important Gene for Lean Body Mass. *American Journal of Human Genetics*. 2009; 84:418–23. [PubMed: 19268274]
21. Xiong DH, Liu XG, Guo YF, Tan LJ, Wang L, Sha BY, et al. Genome-wide association and follow-up replication studies identified ADAMTS18 and TGFBR3 as bone mass candidate genes in different ethnic groups. *Am J Hum Genet*. 2009; 84:388–98. [PubMed: 19249006]
22. Deng FY, Zhu W, Zeng Y, Zhang JG, Yu N, Liu YZ, et al. Is GSN significant for hip BMD in female Caucasians? *Bone*. 2014; 63:69–75. [PubMed: 24607942]
23. Zhang L, Liu YZ, Zeng Y, Zhu W, Zhao YC, Zhang JG, et al. Network-based Proteomic Analysis for Postmenopausal Osteoporosis in Caucasian Females. *Proteomics*. 2015
24. Liu YZ, Zhou Y, Zhang L, Li J, Tian Q, Zhang JG, et al. Attenuated monocyte apoptosis, a new mechanism for osteoporosis suggested by a transcriptome-wide expression study of monocytes. *PLoS One*. 2015; 10:e0116792. [PubMed: 25659073]
25. Tyekucheva S, Marchionni L, Karchin R, Parmigiani G. Integrating diverse genomic data using gene sets. *Genome Biol*. 2011; 12:R105. [PubMed: 22018358]
26. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003; 100:9440–5. [PubMed: 12883005]
27. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
28. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25:1091–3. [PubMed: 19237447]
29. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013; 29:661–3. [PubMed: 23325622]
30. Xing L, Schwarz EM, Boyce BF. Osteoclast precursors, RANKL/RANK, and immunology. *Immunol Rev*. 2005; 208:19–29. [PubMed: 16313338]

31. Ross FP, Teitelbaum SL. alphavbeta3 and macrophage colony-stimulating factor: partners in osteoclast biology. *Immunol Rev.* 2005; 208:88–105. [PubMed: 16313343]
32. Kelly PJ, Eisman JA, Sambrook PN. Interaction of genetic and environmental influences on peak bone density. *Osteoporos Int.* 1990; 1:56–60. [PubMed: 2133642]
33. Slemenda CW, Christian JC, Reed T, Reister TK, Williams CJ, Johnston CC Jr. Long-term bone loss in men: effects of genetic and environmental factors. *Ann Intern Med.* 1992; 117:286–91. [PubMed: 1637023]
34. Aubin JE. Regulation of osteoblast formation and function. *Rev Endocr Metab Disord.* 2001; 2:81–94. [PubMed: 11704982]
35. Bruder SP, Fink DJ, Caplan AI. Mesenchymal Stem-Cells in in Bone-Development, Bone Repair, and Skeletal Regeneration Therapy. *J Cell Biochem.* 1994; 56:283–94. [PubMed: 7876320]
36. Bodine PVN, Komm BS. Wnt signaling and osteoblastogenesis. *Rev Endocr Metab Dis.* 2006; 7:33–9.
37. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Bi.* 2004; 20:781–810.
38. Boyle WJ, Simonet WS, Lacey DL. Osteoclast differentiation and activation. *Nature.* 2003; 423:337–42. [PubMed: 12748652]
39. Roodman GD. Regulation of osteoclast differentiation. *Ann N Y Acad Sci.* 2006; 1068:100–9. [PubMed: 16831910]
40. Yamaguchi H, Condeelis J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Bba-Mol Cell Res.* 2007; 1773:642–52.
41. Takito J, Otsuka H, Yanagisawa N, Arai H, Shiga M, Inoue M, et al. Regulation of Osteoclast Multinucleation by the Actin Cytoskeleton Signaling Network. *J Cell Physiol.* 2014
42. van Buul JD, Hordijk PL. Signaling in leukocyte transendothelial migration. *Arterioscl Throm Vas.* 2004; 24:824–33.
43. Wittchen ES, van Buul JD, BurrIDGE K, Worthylake RA. Trading spaces: Rap, Rac, and Rho as architects of transendothelial migration. *Curr Opin Hematol.* 2005; 12:14–21. [PubMed: 15604886]
44. Walsh MC, Kim N, Kadono Y, Rho J, Lee SY, Lorenzo J, et al. Osteoimmunology: Interplay between the immune system and bone metabolism. *Annu Rev Immunol.* 2006; 24:33–63. [PubMed: 16551243]
45. Martin TJ, Sims NA. Osteoclast-derived activity in the coupling of bone formation to resorption. *Trends Mol Med.* 2005; 11:76–81. [PubMed: 15694870]
46. Nordin BE. Calcium and osteoporosis. *Nutrition.* 1997; 13:664–86. [PubMed: 9263260]
47. Le Clairche C, Carlier MF. Regulation of actin assembly associated with protrusion and adhesion in cell migration. *Physiol Rev.* 2008; 88:489–513. [PubMed: 18391171]
48. Chellaiah MA, Soga N, Swanson S, McAllister S, Alvarez U, Wang DM, et al. Rho-A is critical for osteoclast podosome organization, motility, and bone resorption. *Journal of Biological Chemistry.* 2000; 275:11993–2002. [PubMed: 10766830]
49. Mullin BH, Mamotte C, Prince RL, Wilson SG. Influence of ARHGEF3 and RHOA knockdown on ACTA2 and other genes in osteoblasts and osteoclasts. *PLoS One.* 2014; 9:e98116. [PubMed: 24840563]
50. Zintzaras E, Doxani C, Ziogas DC, Mprotsis T, Rodopoulou P, Karachalios T. Bone mineral density and genetic markers involved in three connected pathways (focal adhesion, actin cytoskeleton regulation and cell cycle): the CUMAGAS-BMD information system. *Biomarkers.* 2011; 16:698–708. [PubMed: 22066665]
51. Nurden AT, Fiore M, Nurden P, Heilig R, Pillois X. Are bone defects in rare patients with Glanzmann's thrombasthenia associated with ITGB3 or ITGA2B mutations? *Platelets.* 2011; 22:547–51. [PubMed: 21557682]
52. Nurden AT, Fiore M, Nurden P, Pillois X. Glanzmann thrombasthenia: a review of ITGA2B and ITGB3 defects with emphasis on variants, phenotypic variability, and mouse models. *Blood.* 2011; 118:5996–6005. [PubMed: 21917754]

53. Zhang L, Choi HJ, Estrada K, Leo PJ, Li J, Pei YF, et al. Multistage genome-wide association meta-analyses identified two new loci for bone mineral density. *Hum Mol Genet.* 2014; 23:1923–33. [PubMed: 24249740]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Biological Significance

Mass spectrometry based quantitative proteomics study integrated with network analysis identified novel genes and pathways related to osteoporosis. The results were further verified in multiple level studies including protein-RNA integrative analysis and genome wide association studies.

Highlights

1. A total of 2058 proteins were identified in 33 samples by LC-nano-ESI-MS^E
2. A total of 30 unique proteins were identified as differentially expressed proteins ($p < 0.05$).
3. Differentially expressed genes: ITGA2B, GSN and RHOA played crucial roles in osteoporosis risk.
4. Pathways: “Regulation of actin cytoskeleton” ($p = 1.13 \times 10^{-5}$, FDR = 3.34×10^{-4}) and “Leukocyte transendothelial migration” ($p = 2.76 \times 10^{-4}$, FDR = 4.71×10^{-3}) are functionally relevant to osteoporosis.
5. The results were further verified in multiple level studies including protein-RNA integrative analysis and genome wide association study.

Figure 1.

Statistical distributions of differentially expressed proteins. **Note:** Y axis showed the ratio of protein expression level between high and low BMD groups. X axis showed the p value of each DEP. The black line is the cut off line which means the **Ratio H/L** equals to 1. Proteins above the line are up-regulated in high BMD group, below the line are down-regulated in high BMD group.

Figure 2. Prediction of Protein interactions in the selected two pathways in STRING. **Note:** Multiple interactions between those genes enriched in these two pathways: “regulation of actin cytoskeleton” and “leukocyte transendothelial migration”. 8 and 5 genes out of the target gene list were enriched in these two pathways respectively (see table 4). Different line colors represented the types of evidences for the association.

Figure 3.

Cytoscape platform based ClueGO/CluePedia network analysis and visualization. **Note:** Network analysis and visualization of the target gene list. a) GO/pathway terms specific for differentially expressed genes. The bars represent the number of genes associated with the terms. The percentage of genes per term (the percentage of the genes from the uploaded cluster that were associated with the term, compared with all the genes associated with the term in the database) is shown as bar label. b) A visualized network for pathways and the related genes. The interesting two pathways and three genes were intentionally marked in yellow circles and red rectangles, respectively.

Table 1

Basic characteristics of the subjects

Group	Pre-menopausal women with High BMD	Pre-menopausal women with low BMD
Sample size	16	17
Age (yrs)	51.80(2.27)	50.25(1.86)
Height (cm)	162.44(5.99)	167.45(6.61)
Weight (kg)	73.16(16.76)	65.25(7.74)
Hip BMD	1.32(0.45)	− 0.96(0.34)

Note: Data were presented as Mean (SD). Hip BMD is presented as Z-score. Z-score is defined as the number of standard deviations a subject's BMD from the average value of the reference group (with the same age, gender and ethnicity).

Table 2

Differentially expressed proteins

UniProt ID	Gene symbols	P value	Ratio H/L
Q59G88	CORO1A	0.0008	1.7
P18054	ALOX12	0.0029	2.5
O00194	RAB27B	0.0045	1.9
B0LUF9	ALDH2	0.0048	1.9
Q6LDK5	ITGA2B	0.0063	1.8
P09493	TPM1	0.0087	1.6
-	LOC390956	0.0093	0.7
P08238	HSP90AB1	0.0136	1.2
P16109	SELP	0.0145	2.9
P30086	PEBP1	0.0177	0.6
Q5T0I0	GSN	0.0186	1.8
P14770	GP9	0.0206	2.6
P61018	RAB4B	0.0251	1.6
Q15365	PCBP1	0.0252	0.6
Q5VSR5	PFKP	0.0300	2.7
B4E190	ARF3	0.0311	2.1
P10909	CLU	0.0352	1.5
P35527	KRT9	0.0360	3.3
P40939	HADHA	0.0364	1.6
P14649	MYL6B	0.0372	1.3
Q13418	ILK	0.0392	1.6
Q96DA2	RAB39B	0.0400	1.5
P04264	KRT1	0.0414	1.7
Q9Y536	PPIAL4A	0.0427	0.7
-	ALDOAP2	0.0430	0.8
P20337	RAB3B	0.0446	1.8
P04406	GAPDH	0.0458	0.7
A6NDY9	FLNA	0.0474	1.4
P07951	TPM2	0.0480	1.5
P27105	STOM	0.0491	1.6

Note: With normalized expression data, student t test was applied to identify DEPs. P value less than 0.05 was set as the nominal significant cutoff. 30 proteins reached the nominal significant statistical level. **Ratio H/L** represented the fold change of protein expression level between low and high BMD groups, which reflected that proteins were up-regulated (>1) or down-regulated (<1) in High BMD group.

Table 3

Functional annotation and enrichment analysis in DAVID

Category	Term	P Value	DEPs Included
GOTERM_BP_FAT	GO:0030036~actin cytoskeleton organization	7.56E-16	CORO1A, FLNA, GSN, TPM1
GOTERM_BP_FAT	GO:0006928~cell motion	7.73E-13	ALOX12, GAPDH, HADHA
GOTERM_BP_FAT	GO:0007010~cytoskeleton organization	4.83E-12	GSN, ITGA2B
GOTERM_BP_FAT	GO:0006091~generation of precursor metabolites and energy	6.77E-10	GAPDH, PFKP
GOTERM_BP_FAT	GO:0051592~response to calcium ion	2.40E-04	PEBP1
GOTERM_BP_FAT	GO:0050900~leukocyte migration	3.01E-04	CORO1A, SELP
GOTERM_CC_FAT	GO:0015629~actin cytoskeleton	9.17E-24	CORO1A, FLNA, GSN, ILK, MYL6B, TPM1, TPM2
GOTERM_MF_FAT	GO:0005509~calcium ion binding	1.80E-04	GSN, ITGA2B, MYL6B,
INTERPRO	IPR002048:Calcium-binding EF-hand	5.53E-04	MYL6B
SP_PIR_KEYWORDS	calcium	1.08E-06	GSN, ITGA2B

Note: 10 terms highly related to bone metabolism are presented here. Functional processes significantly enriched in terms such as “actin cytoskeleton organization”, “generation of precursor metabolites and energy”, “leukocyte migration”, “calcium ion binding” and “calcium”. The last column listed the DEPs included in corresponding terms.

Table 4

Genes enriched in selected two pathways in STRING analysis

Pathway	Gene
Regulation of actin cytoskeleton	RAC3, ITGA2B, RAC2, ARPC2, RHOA, CDC42, GSN, ARPC5
Leukocyte transendothelial migration	RHOA, GNAI1, CDC42, RAC2, VASP

Note: The first column listed the names of selected two pathways and the second column showed genes enriched in these two pathways, respectively.