DOI 10.1002/pmic.201500005 *Proteomics* 2016, *16*, 12–28

RESEARCH ARTICLE

Network-based proteomic analysis for postmenopausal osteoporosis in Caucasian females

Lan Zhang[1], Yao-Zhong Liu[1], Yong Zeng[1,2], Wei Zhu[1], Ying-Chun Zhao[1], Ji-Gang Zhang[1], Jia-Qiang Zhu[1], Hao He[1], Hui Shen[1], Qing Tian[1], Fei-Yan Deng[1,3], Christopher J. Papasian[4] and Hong-Wen Deng[1,2]

[1] Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA
[2] College of Life Sciences and Bioengineering, Beijing Jiaotong University, Beijing, P. R. China
[3] Laboratory of Proteins and Proteomics, Department of Epidemiology, Soochow University School of Public Health, Suzhou, P. R. China
[4] Department of Basic Medical Sciences, School of Medicine, University of Missouri – Kansas City, MO, USA

Menopause is one of the crucial physiological events during the life of a woman. Transition of menopause status is accompanied by increased risks of various health problems such as osteoporosis. Peripheral blood monocytes can differentiate into osteoclasts and produce cytokines important for osteoclast activity. With quantitative proteomics LC-nano-ESI-MS^E (where MS^E is elevated-energy MS), we performed protein expression profiling of peripheral blood monocytes in 42 postmenopausal women with discordant bone mineral density (BMD) levels. Traditional comparative analysis showed proteins encoded by four genes (LOC654188, PPIA, TAGLN2, YWHAB) and three genes (LMNB1, ANXA2P2, ANXA2) were significantly down- and up-regulated, respectively, in extremely low- versus high-BMD subjects. To study functionally orchestrating groups of detected proteins in the form of networks, we performed weighted gene coexpression network analysis and gene set enrichment analysis. Weighted gene coexpression network analysis showed that the module including the annexin gene family was most significantly correlated with low BMD, and the lipid-binding related GO terms were enriched in this identified module. Gene set enrichment analysis revealed that two significantly enriched gene sets may be involved in postmenopausal BMD variation by regulating pro-inflammatory cytokines activities. To gain more insights into the proteomics data generated, we performed integrative analyses of the datasets available to us at the genome (DNA level), transcriptome (RNA level), and proteome levels jointly.

Received: January 5, 2015
Revised: September 6, 2015
Accepted: October 28, 2015

Keywords:
Bioinformatics / GSEA / LC-MS / Menopause / Peripheral blood monocyte / WGCNA

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Osteoporosis, characterized by low bone mineral density (BMD) and microarchitectural deterioration of bone tissue that leads to an increased risk of fracture [1], is a major

Correspondence: Dr. Hong-Wen Deng, Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 2001, New Orleans, LA 70112, USA
E-mail: hdeng2@tulane.edu
Fax: +1-504-988-1706

Abbreviations: **BMD**, bone mineral density; **ES**, enrichment score; **FN-BMD**, femoral neck bone mineral density; **GEFOS2**, Genetic Factors of Osteoporosis Consortium; **GSEA**, gene set enrichment analysis; **GWA**, genome-wide association; **HDMS**, high-definition MS; **MS^E**, elevated-energy MS; **NES**, normalized enrichment score; **PBM**, peripheral blood monocyte; **RAP1A**, Ras-related protein Rap-1A; **TOM**, topological overlap matrix; **WGCNA**, weighted gene coexpression network analysis

public health concern in females [2]. Osteoporosis develops because of an imbalance between the two counteracting processes of bone resorption and bone formation. Risk factors for osteoporosis in females include age, exercise, menopause, duration of fertility, and low-calcium diet [3, 4].

Monocytes are circulating, bone marrow derived, leukocytes that can differentiate into various cell types including tissue macrophages, dendritic cells, and osteoclasts [5]. In humans, mature peripheral blood monocytes (PBMs) constitute 5–10% of peripheral blood leukocytes [6]. PBMs are morphologically heterogeneous and can be further divided into three (classic, nonclassic, and intermediate) subpopulations based on the surface expressions of CD14 and CD16 [7]. The classic circulating monocyte is characterized by high-level CD14 expression and lack of CD16 expression, and this subset of PBM can serve as potential osteoclast precursors [8, 9]. PBMs produce cytokines important for osteoclast differentiation, activation, and apoptosis and are one of the most important target cells for sex hormones in bone metabolism [10, 11]. Therefore, PBMs represent a class of cells that are of high functional relevance to the pathogenesis of osteoporosis, and have been well substantiated and accepted for studies of pathophysiology in the bone field [12–14], as reviewed comprehensively recently [8].

The menopause transition occurs between 40 and 58 years of age for most females [15, 16], and is associated with an increased risk of a number of disorders such as cardiovascular disease and osteoporosis [17, 18]. Natural menopause is due to a decline in ovarian function, which is characterized by decreased estrogen production and increased levels of follicle-stimulating hormone [19]. Several recent studies have demonstrated the influence of menopause on monocyte. For example, the number of monocytes increases during menopause [20], and differential estrogen receptor expression was observed in monocyte populations isolated from pre- and postmenopausal females [21]. In view of the hormonal changes associated with menopause described above, it is significant that follicle-stimulating hormone has been shown to enhance receptor activator of nuclear factor kappa-B ligand (RANKL)-stimulated osteoclastogenesis from human PBMs [22], while estrogen inhibited RANKL-stimulated osteoclastic differentiation of monocytes [23].

Proteomics studies allow comprehensive detection and quantification of proteins in complex biological systems. The application of quantitative proteomics to osteoporosis research has resulted in the identification of biomarkers of osteoblasts and osteoclasts and characterization of protein signaling during bone remodeling [24–26]. A comparative protein expression study of PBMs in Chinese premenopausal high- and low-BMD females identified a total of 38 differentially expressed proteins; 5 of these 38 proteins were subjected to and confirmed with Western blotting experiments [27]. Another MS-based proteomic study revealed increased expression of ANXA2 in PBMs in Caucasians with extremely low BMD; ANXA2 has been shown to contribute to monocyte trans-endothelium migration into the bone microenvironment [10].

Genes and their functional products such as proteins do not act in isolation; rather they work together interactively in forms of pathways and functional modules. Methods for inferring gene interactions from expression datasets have become an active research area of systems biology. Network-based analysis strategies such as gene set enrichment analysis (GSEA) and weighted gene coexpression network analysis (WGCNA) are now extensively used in various functional genomics and bioinformatics studies. GSEA is a computational method to explore the biological similarity within groups of gene sets, and it can be applied to proteome datasets as well [28–30]. Recently, this method was modified into protein set enrichment analysis (PSEA) to study the differential protein expression based on spectral index from label-free quantitative proteomics research [31]. WGCNA can be used to study modules based on coexpression networks instead of individual genes, thus capturing orchestrating gene actions and interactions. By examining eigengenes (the first principle component, capturing the major effects of groups of correlated and interacting genes/proteins) and intramodular hub genes (genes that are highly correlated with the module eigengene, often the most central functional elements of modules associated with the trait) of each network, the relation between biological traits and modules can be further explored in more detail and in more comprehensive and systematic manner [32]. This promising novel approach has been successfully applied to identify candidate biomarkers and potential therapeutic targets [33].

With sensitive LC-nano-ESI-MS^E (where MS^E is elevated-energy MS) based quantitative proteomic strategy, the first aim of this study was to identify individual proteins that contribute to the pathogenesis of postmenopausal osteoporosis. By applying an advanced network analysis strategy, the second aim of the study was to detect highly correlated protein clusters important for postmenopausal BMD variation. To explore additional evidence for the results from our quantitative proteomic discovery study, integrating analyses of identified target genes were performed at DNA, RNA, and protein levels.

2 Materials and methods

2.1 Human subjects

This study was approved by the involved Institutional Review Boards, and all subjects signed consent forms before being enrolled into this study. All subjects were self-identified as European Caucasian females. Postmenopausal status was defined as no menses for at least 1 year after their last menses. Forty-two postmenopausal subjects were recruited in this study. The lumbar spine BMD (g/cm^2) and hip BMD (which is the combined value of three regions including femoral neck, trochanter, and interchochanter) were measured by a

Table 1. Characteristics of the study sample

Category	Postmenopausal women with high BMD	Postmenopausal women with low BMD
Sample size	21	21
Age (years)	63.95 (8.39)	62.43 (9.30)
Height (cm)	161.86 (6.44)	162.62 (9.28)
Weight (kg)	80.62 (12.15)	61.06 (9.75)
Hip BMD	1.02 (0.12)	−0.76 (0.14)

Data are presented as mean (SD). BMD values are presented as *Z*-score, which is defined as the number of SDs a subject's BMD differs from the average value of the reference group (same age, gender, and ethnicity).

Hologic 4500-W dual-energy X-ray bone densitometer for each subject.

In order to minimize the influence of environmental factors that may affect bone metabolism, several chronic diseases and conditions were excluded. Detailed exclusion criteria were described in previous studies by our group [10], which mainly include vital organs' chronic disorders, autoimmune-related diseases, metabolic diseases, skeletal diseases, hematopoietic diseases, lymphoreticular disease, and other diseases or any treatment that would potentially affect gene expression in PBMs.

The study sample included 42 unrelated postmenopausal Caucasian women, 21 with high BMD (*Z*-score: 1.02 ± 0.12) and 21 with low BMD (*Z*-score: –0.76 ± 0.14). The *Z*-score represents how extreme the subjects selected from the BMD distribution in our study population are, and it is defined as the number of SDs a subject's BMD differs from the average BMD of their age-, gender-, and ethnicity-matched population. Detailed clinical information is provided in Table 1.

2.2 PBM isolation

Sixty-milliliter peripheral blood was collected from each subject by a phlebotomist, and fresh blood samples were processed immediately for PBM isolation. PBMCs were first isolated from whole blood by density gradient centrifugation with Histopaque-1077 (Sigma, H1077-1). To minimize the potential cell-type heterogeneity, we used the Monocyte Isolation Kit II (MiltenyiBiotec, Auburn, CA, USA) to specifically isolate the classic PBM subset ($CD14^{++}CD16^{-}$). Nonmonocytes were indirectly magnetically labeled with a cocktail of biotin-conjugated monoclonal antibodies (against CD3, CD7, CD16, CD19, CD56, CD123, and glycophorin A), as primary labeling reagent, and anti-biotin monoclonal antibodies conjugated to MicroBeads, as secondary labeling reagent. The magnetically labeled nonmonocytes were depleted by retaining them on a MACS Column in the magnetic field of a MACS Separator, while the unlabeled monocytes passed through the column. With these procedures, the classical $CD14^{++}CD16^{-}$ monocytes that represent 90–95% of the total monocytes in a healthy person were negatively isolated, untouched, and unaffected by antibodies from fresh blood samples [34], best representing the in vivo expression profiles in human subjects.

2.3 Complete proteome extraction

We used ProteoExtract Complete Mammalian Proteome Extraction Kit (Calbiochem, San Diego, CA, USA) to isolate virtually all proteins from PBMs. Protein concentrations were measured using Bradford's method. With the protein precipitation kit (Calbiochem, no. 539180), up to 20 μg total protein was precipitated. Protein pellets were dissolved in 50-μL ammonium bicarbonate (50 mM) with 0.1% RapiGest (Waters, Milford, MA, USA), reduced by dithiothreitol (5.0 mM), alkylated by iodoacetamide (15 mM), and then digested by 1-μg trypsin (Sigma, T6567). Protein digests were concentrated to less than 5 μL and brought to 20 μL with 200-mM ammonium formate (pH 10). We used 100 fmol yeast alcohol dehydrogenase I digestion as the internal standard (Waters, no. 186002328) for protein quantification.

2.4 Proteomics profiling with LC-nano-ESI-MS^E

The proteome of PBM total protein was profiled through nanoAcquity ultraperformance LC coupled with Synapt High Definition Mass Spectrometry (HDMS, Waters) and the data acquisition process was controlled by MassLynx 4.1 software (Waters).

As described in our previous research [10], protein digests (~0.5 μg) were injected into a BEH C18 (75 μm × 150 mm) analytical column, and separation procedure was achieved by changing the percentage of solvent A (water with 0.1% formic acid) and solvent B (ACN with 0.1% formic acid). The 2-h separation gradient at a flow rate of 0.3 μL/min was set as follows: 3% B original, 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 sensitivity mode. Data acquisition process was conducted with the following parameters, the collision energy was set 5 V for MS and 15–40 V for MS^E; the scan time was set 0.6 s per scan. The HDMS machine was calibrated daily to ensure high accuracy (2.0 ppm for lock mass of m/z 785.8426).

Triplicate LC-nano-ESI-MS^E datasets were acquired for each PBM proteome digest sample, and the MS^E data were processed with Protein Lynx Global Server version 2.3 (Waters) using the default parameter setting. Properties of each ion such as m/z and retention time were determined based on the alternating low- and elevated-energy nature of MS^E, and a list of all precursor and product ions was produced. Particularly, the ion's intensity was derived from the areas of both the chromatographic and mass spectrometric peaks, and the precursor ion intensity threshold was set to be above 1000 counts to remove noisy signal. Protein identification was achieved by searching the Human protein database International Protein

Index version 3.56 (153 078 protein entries) with the following parameters: enzyme specificity (trypsin), fixed modification (carbamidomethyl C), variable modifications (acetyl N-TERM, deamidation N, deamidation Q, and oxidation M), number of missed cleavages permitted (1), mass tolerance for precursor ions (15 ppm), mass tolerance for product ions (30 ppm), minimum peptide matches per protein (1), minimum fragment ion matches per protein (7), minimum fragment ion matches per peptide (2), and false-positive rate (4% per randomized database searching).

For each sample, only proteins identified at least twice in the triplicate LC-nano-ESI-MS^E analyses were reported as truly present to ensure reproducibility. Total ion counts of the top-three matched peptides were used to quantify each protein. With the inner standard alcohol dehydrogenase I as reference, 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.

2.5 Differential protein expression analysis

In order to adjust the variability due to sample preparation or equipment conditions, the global normalization method was applied to adjust the potential systematic bias. The assumption of global normalization is that though some protein abundance may change, the expression levels of conserved proteins in each subject are approximately equal. In this study, raw protein expression levels were normalized against the internal control beta-actin, since it is a widely recognized housekeeping protein. Mann–Whitney *U* test was used to detect significantly differentially expressed proteins in subjects with low versus high BMD. Since weight was significantly different between the two BMD groups, it was adjusted before conducting the test; and *p*-value less than 0.05 was set as the significant cutoff.

2.6 Gene set enrichment analysis

Proteins were ranked based on the correlation between their expression and BMD variation. By applying GSEA, biologically informative protein subsets were studied in various functional categories. The molecular signatures database Msigdb.v4.0 was used to build the annotation sets and the enrichment score (ES) was calculated through running sum statistics as originally described [29]. The ES was defined as the maximum deviance (positive or negative) of the running statistics from zero. The empirical distribution of the ES was calculated based on 2000 phenotype-based permutations, and the statistical significance (nominal *p*-value) of the observed ES was determined according to the null distribution. The ES will be normalized for each gene set to account for the size of the set, yielding a normalized ES (NES). NES was defined as actual ES of a gene set divided by the mean of the ES against all permutations of the dataset [29]. The FDR corresponding to each NES was calculated to estimate the probability of a given score developed from a false-positive finding. As previously implemented, gene sets with FDR less than 0.25 were considered as significantly enriched [29]. Often it is useful to identify the core members of gene sets that contribute to the ES, which was defined as "leading edge" subsets [29]. The "leading edge" gene subsets were computed for each significant gene set by selecting the genes with more-extreme rank than the rank corresponding to the ES value.

2.7 WGCNA and GO analysis

WGCNA includes gene coexpression, network construction, module identification, module-phenotype correlation recognition, and key driver gene identification [33]. Intensive tutorials of this method can be found at http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/. In the current study, we applied this method to construct the correlation patterns among our protein samples. The nodes of network refer to protein expression profiles, and edges between proteins were determined by the pairwise correlations between their expressions. Since LC-MS^E data have inevitably missing values, it is necessary to filter data before subsequently analyses to guarantee the reliability of the de novo network construction. For data preprocessing, we used "goodSamplesGenes" function in WGCNA R package to iteratively remove samples and proteins with more than 50% missing entries (default). Eventually 486 proteins passed the filtering and entered subsequent analysis. The unsigned coexpression similarity, defined as the absolute value of correlation coefficient between two nodes, was used in our analysis. By raising the unsigned coexpression similarity to a power $\beta \geq 1$ (soft thresholding), the correlation matrix was converted into the adjacency matrix, where the adjacency between *i*th and *j*th nodes was defined as $\alpha_{ij} = |cor(x_i, x_j)|^{\beta}$. The power β was chosen based on the criterion of approximate scale-free topology [32], which means the node degree distribution following the power law [35] and it is the key parameter for adjacency matrix construction. Network connectivity k_i of the *i*th gene was calculated as the sum of the connection strengths with all other network genes ($k_i = \sum_{i \neq j} \alpha_{ij}$). The topological overlap matrix (TOM) between *i*th and *j*th protein expressions was taken as $\text{TOM} = \frac{\sum_{u \neq i,j} \alpha_{iu}\alpha_{uj} + \alpha_{ij}}{\min\{k_i, k_j\} + 1 - \alpha_{ij}}$, where $\sum_{u \neq i,j} \alpha_{iu}\alpha_{uj}$ denotes the number of nodes to which both *i* and *j* are connected; *u* indexes the other nodes of the network. Hierarchical cluster and dynamic tree cut were utilized to identify protein modules based on the corresponding dissimilarity (1-TOM), and the minimum module size chosen was 30. We further checked the correlations between eigenprotein (the first principle component of module) from different modules to ensure the reliability of the cluster process. If the correlations between eigenproteins of different modules were greater than 0.75, they could be merged into one.

To relate protein modules to phenotypes in our dataset, we calculated Pearson's correlation between eigenprotein and clinical traits, and modules with *p*-value less than 0.05 were taken as target modules. Two methods were used to process the GO enrichment analysis in interesting modules: the function carried by WGCNA package ("org.Hs.eg.db" as annotation database) and DAVID software [36]. All proteins entered into WGCNA served as the genetic background, and proteins in selected modules were imported as target list. The Bonferroni-adjusted *p*-value (<0.05) and FDR (<0.10) were used as thresholds to detect the significant functional annotation clusters. Interacting patterns for all proteins in the identified module were searched through STRING version 9.1 (a search tool for the retrieval of interacting genes/proteins) [37], a comprehensive database designed to collect and evaluate protein–protein association information. The combined score (combining the probabilities of protein–protein associations based on different lines of evidence, i.e., gene fusion, phylogenetic cooccurrence, large-scale experiments, etc.) [38] for each pair of proteins was recorded. Based on previous research, combined scores between 0.4 and 0.7 indicate medium confidence for predicting protein interactions, combined scores greater than 0.7 represent high confidence [38].

2.8 Corroboration with transcriptome-wide gene expression studies

In order to explore the significance (beyond that at the protein level) of the identified proteins/genes in differential expression analysis, GSEA, and WGCNA at transcriptional level, two archived datasets of independent microarray studies for osteoporosis from our group were used [39, 40]. The first study contains 73 Caucasian females, 42 subjects with high hip BMD (1.38 ± 0.49) and 31 subjects with low hip BMD (–1.05 ± 0.51). The second study contains 80 Caucasian females, 40 subjects with high hip BMD (1.45 ± 0.67) and 40 subjects with low hip BMD (–1.05 ± 0.44). Total RNA from PBM was extracted using the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA), and the mRNA expression level was obtained with Affymetrix Human Genome U133A Array. Robust multiarray average method [41] was used to normalize the array signals and Student's *t*-test was conducted to perform differential expression analysis for the identified target genes through the Bioconductor's LIMMA (linear models for microarray data) package [42].

2.9 Corroboration with genome-wide association (GWA) meta-analysis studies

To assess additional significance for proteins identified from differential expression analysis and genes identified from network analyses in our discovery study, we compared our findings with genes previously found to be associated with BMD at the population level in two GWA meta-analysis studies [43–45]. The first study included 27 061 subjects, and it independently confirmed 13 previously reported loci important for BMD variation [43]. Seven GWAS datasets were included in this GWA meta-analysis, and we assessed the genetic association between SNPs and hip BMD in samples from females only, and in samples from males and females combined. The second GWA meta-study GEFOS2 (Genetic Factors of Osteoporosis Consortium) is the largest meta-analysis performed to date in the bone field [44, 45]. GEFOS2 included 32 961 individuals from European and East Asian populations in the discovery stage, and 50933 individuals in the replication stage; a total of 56 BMD loci were identified using these cohorts. Since GEFOS2 did not use hip BMD as a phenotype, we checked the genetic association between SNPs and femoral neck (part of the total hip) BMD (FN-BMD) in both female samples and the pooled samples. The significant level of SNPs was studied to further explore the importance of relevant genes to postmenopausal BMD variation.

2.10 Integrative analyses for important genes with DNA, RNA, and protein-level data

To further evaluate the overall significance of the genes to hip BMD found in proteomic analyses, we performed integrative analyses using the method based on multiple genomic features [46] to combine evidence from SNP genotype, RNA expression, and protein expression datasets. In addition to the protein expression data used in the discovery stage, three genomic SNP genotype datasets and two transcriptomic mRNA datasets were involved in the analyses. Since explicit genotypes and phenotypes for each subject were required in the calculation, three SNP genotype datasets generated by our own group [43] were used, and only female subjects were included in the analyses. Detailed features of the three GWAS datasets are shown in Table 2. The mRNA datasets were from two transcriptome-wide expression studies introduced in the earlier part, which include 73 and 80 Caucasian females, respectively [39, 40]. Test statistics were calculated

Table 2. Characteristics of the three GWAS samples involved in the integrative analyses

Study name	Population	*N*	Female (%)	Hip BMD (g/cm^2)
Chinese osteoporosis study (COS)	Han Chinese	1547	50.7	0.92 (0.13)
Kansas city osteoporosis study (KCOS)	Caucasian	2250	75.9	0.97 (0.17)
Omaha osteoporosis study (OOS)	Caucasian	987	49.6	0.97 (0.16)

Hip BMD data are presented as mean (SD).

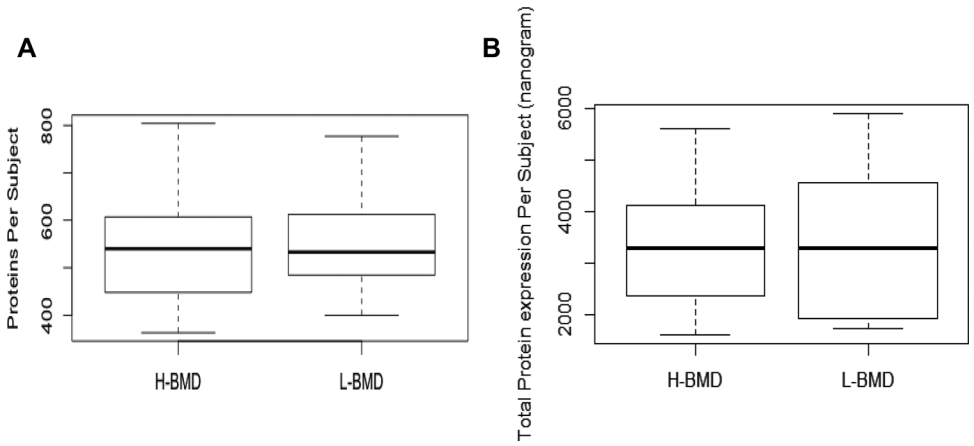

Figure 1. Boxplot for basic descriptive statistics of proteomic profiling. (A) Boxplot for the distribution of average number of proteins identified per subject in high- and low-BMD groups. (B) Boxplot for the distribution of total protein expression level per subject for the 1093 proteins across the high- and low-BMD groups.

independently among these six datasets to study the association between genomic features and hip BMD. Then, a single score was combined based on the test statistics of the six datasets according to the following formula developed earlier [40].

$$\mathrm{TS}\,(g) = \omega_{\mathrm{n}} \times ts\,(X_{\mathrm{n}}, \mathrm{BMD})\ \ n \in \{1, \ldots, 6\}$$

X_{n} represents the nth genomic-level data, and $ts\,(X_{\mathrm{n}}, \mathrm{BMD})$ is the test statistic for independent association analysis between genomic feature "g" and hip BMD at each level. $ts(g)$ denotes the integrative test statistics for genomic feature, and ω_{n} is the weight of the nth data. In this study, we assume that the total contribution of each type of data (such as from DNA, mRNA, and proteins) is the same with weight being 1. The weight values of each of the three genotype datasets and each of the two mRNA datasets were set as 1/3 and 1/2, respectively; the weight of the only proteomic expression data was set as 1. A permutation-based procedure was used to test the statistical significance of the integrated score [46].

3 Results

3.1 Proteomic profiling

Of the 42 samples, 2018 PBM proteins were detected in total by LC-nano-ESI-MS$^{\mathrm{E}}$. To verify the performance of the proteomic profiling procedure, 1093 of 2018 proteins, which were detected in more than 10% of the cohort (5/42), were used for further analysis. There was an overlap of 99.63% of these 1093 identified proteins between high- and low-BMD subject groups. Boxplots for the distribution of the average number of proteins expressed and total protein expression levels per subject in low- versus high-BMD groups are shown in Fig. 1.

3.2 Protein differential expression analysis

Mann–Whitney U test was applied to detect differentially expressed proteins in low- versus high-BMD groups. Four proteins (LOC654188, PPIA, TAGLN2, and YWHAB) were significantly downregulated in the postmenopausal low-BMD group, and three proteins (LMNB1, ANXA2P2, and ANXA2) were significantly upregulated in the postmenopausal low-BMD group. Further details are shown in Table 3.

3.3 Gene set enrichment analysis

Based on the same data preprocessing method used in WGCNA analysis, proteins with more than 50% missing entries were removed; and 486 proteins mapped to 179 genes entered the GSEA analysis. Among these 179 genes, 96 were up-

Table 3. Up- and downregulated proteins in low- versus high-BMD postmenopausal Caucasians

IPI ID	Gene symbol	Protein name	logFC (L/H)	*p*-Value
Proteins downregulated in low-BMD group				
IPI00419585.9	PPIA	Peptidylprolyl isomerase A	−0.7369	0.0028
IPI00887678.1	LOC654188	Similar to peptidylprolyl isomerase A isoform 1	−0.8142	0.0031
IPI00216318.5	YWHAB	Isoform Long of 14-3-3 protein beta/alpha	−0.4973	0.0120
IPI00550363.3	TAGLN2	Transgelin 2	−0.6007	0.0442
Proteins upregulated in low-BMD group				
IPI00455315.4	ANXA2	Annexin A2	0.5368	0.0172
IPI00790831.1	LMNB1	Lamin B1	0.6100	0.0305
IPI00334627.3	ANXA2P2	Annexin A2-like protein	0.5282	0.0456

Table 4. Significant gene set found in GSEA analysis based on BMD variation

Gene set	Platelet activation signaling and aggregation	Hemostasis
ES	−0.8034	−0.7592
NES	−1.8549	−1.8261
Nominal *p*-value	0.0069	0.0099
FDR *q*-value	0.0329	0.0258
Familywise error rate (FWER) *p*-Value	0.0261	0.0405

regulated and 83 were downregulated in the postmenopausal low-BMD group. Thirty-five gene sets in MSigDB matched our dataset with at least 25 genes. Among these identified gene sets, 21 were downregulated in the postmenopausal low-BMD group, and two were significantly enriched at FDR less than 25%, with *p*-values less than 0.05. The enrichment results are shown in Table 4 and Fig. 2. The most significant gene set was related to "platelet activation, signaling, and aggregation," with NES equal to –1.85. Among this gene set, 25 genes in our expression dataset were identified and 16 of them were considered as core enrichment genes ("leading edge" subsets). Another significant gene set was "hemostasis," with NES equal to –1.83. Twenty-eight genes in our dataset were included in this gene set, and seventeen of them were detected as core enrichment genes. Most members in the identified significant gene sets were found upregulated in postmenopausal low-BMD group (NES < 0). Detailed information about the above two significant gene sets is provided in the Supporting Information.

3.4 Weighted gene coexpression network analysis

3.4.1 Module containing annexin family genes was significantly correlated with postmenopausal BMD variation

The scale-free network topologies have node degree distributions following the power law [35] and soft thresholding power (β) was used in scaling the adjacency matrix [32]. In this study, the scale-free fitting index R^2 reached its peak when $\beta = 10$. Thus, $\beta = 10$ was used to construct the adjacency matrix where the adjacency between *i*th and *j*th nodes was defined as $\alpha_{ij} = |\mathrm{cor}(x_i, x_j)|^{10}$.

As shown in Fig. 3A, eight modules were detected by the dynamic tree cut based on 1-TOM. No modules could be merged after checking their eigenprotein coexpression, which confirmed the reliability of the clearly divided modules.

As the first principal component of the detected module, eigenprotein represents the major fraction of variation of the expression profile in the corresponding module, and its correlation with phenotypes can help identify potential biomarkers. By investigating the correlation between eigenproteins and clinical traits, we found some interesting modules worth further exploration. As shown in Fig. 3B, significant correlations between module eigenproteins and phenotypes were observed. Since our study focus was on postmenopausal BMD variation, the yellow module, which showed negative correlation with BMD group ($p = 0.04$) was our target protein cluster. The BMD group was defined as a categorical variable, with 0 representing low BMD and one representing high BMD. In addition, given that the eigenprotein in this module was not significantly correlated with other phenotypes, our findings might suggest that protein expression profiles in this module represent only protein clusters associated with BMD variation. As shown in Fig. 3C, the protein significance and module membership exhibited a significantly positive correlation ($r = 0.6$) in this module. Since protein significance is defined as the correlation of a protein expression profile with a sample trait, and module membership (eigenprotein-based connectivity) measures the correlation between an individual protein and the module eigenprotein [33], our findings further suggest that intramodular hub proteins (proteins greatly correlated with the module eigenprotein) in the yellow module were highly associated with BMD group. The annexin family genes including ANXA1, ANXA2, ANXA5, and ANXA6 were found in this module, and they were all upregulated in the low-BMD group.

3.4.2 Lipid binding related GO term enriched in the target module

The results from DAVID and WGCNA packages were overlapped, and three GO terms were significantly enriched in the BMD group related module based on adjusted *p*-values and FDR. The three GO terms including "lipid binding," "phospholipid binding," and "calcium-dependent phospholipid binding" were identified with adjusted *p*-values less than 0.01. Detailed information about these three GO terms is provided in Table 5. All three GO terms are relevant to lipid-binding processes, and include proteins of the annexin family.

3.4.3 Interacting network involving annexin and ubiquitin protein families were identified in the target module

Isoforms of 16 proteins (ANXA1, ANXA2, ANXA2P2, ANXA5, ANXA6, CD14, EIF4A1, HNRNPK, LCP1, LGALS1, LOC388720, MPO, RPS27A, UBA52, UBB, VCP) were identified in the target module. When constructing the interacting network, STRING treated ANXA2P2 the same as ANXA2, and this database does not have information of LOC388720.

Based on computational predictions by STRING software, two subsets protein interacting networks were detected, as

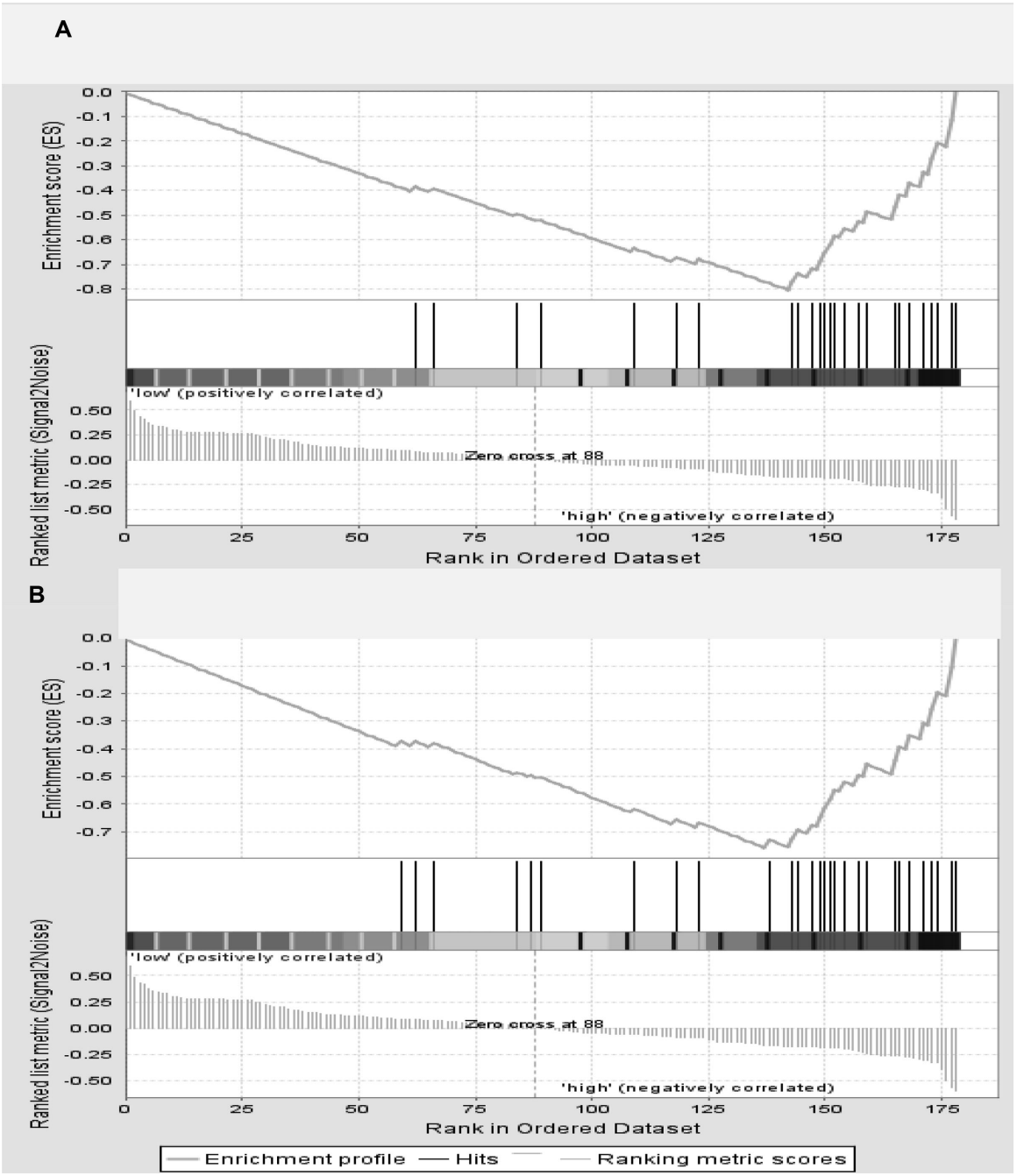

Figure 2. Results of GSEA for postmenopausal BMD variation. (A) Enrichment plot of "platelet activation signaling and aggregation" gene set. (B) Enrichment plot of "hemostasis" gene set. The ES for the gene set is computed as the minimum of the running sum statistic that is depicted as the curve line in the upper panel. The black vertical lines in middle panels indicate the hit genes in the detected gene set. The bottom portion of the plot shows the value of the ranking metric as moving down the list of ranked genes. The ranking metric measures the gene's correlation with BMD group.

shown in Fig. 4. The first coexpression subnetwork contained ANXA1, ANXA2, ANXA5, ANXA6, and LGALS1, and the second contained ubiquitin-related proteins RPS27A, UBB, UBA52, EIF4A1, LCP1, and VCP. The detailed pairwise combined scores for protein interacting networks and the supporting evidence are shown in Table 6. For the annexin-related protein interacting network, the combined prediction score for protein–protein pairwise association between ANXA5 and LGAS1 reached the level of high confidence (combined prediction score >0.7). In the ubiquitin-related subset protein

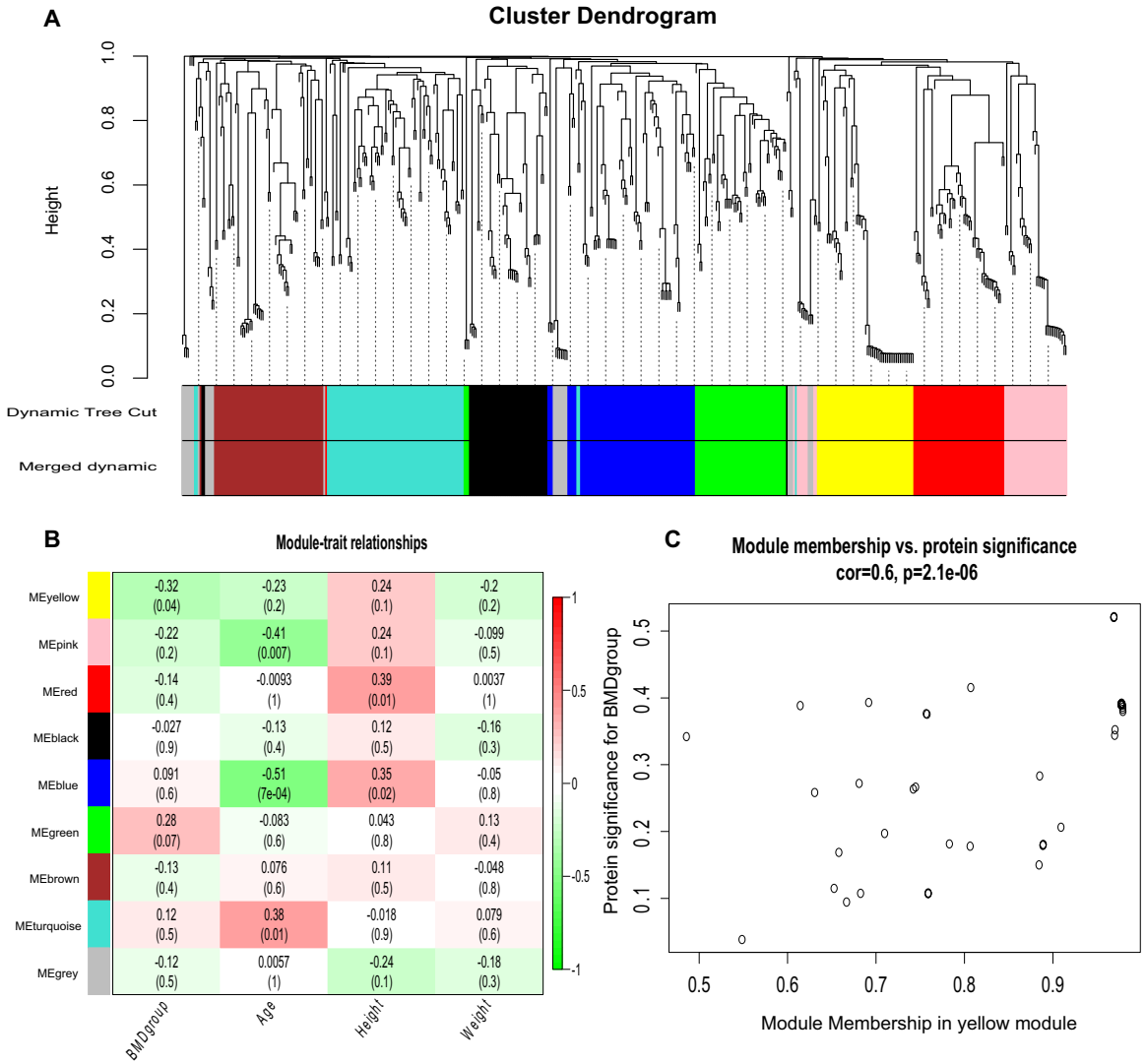

Figure 3. Results of WGCNA for postmenopausal BMD variation. (A) Protein dendrogram depicted by average linkage hierarchical clustering method. The *y*-axis "height" indicates dissimilarity based on topological overlap. The different color rows underneath the dendrogram display module assignment by dynamic tree cut strategy. (B) Heat map of module-trait association. Each column corresponds to a phenotype trait, including BMD group (0, low hip BMD; 1, high hip BMD), age (years), height (cm), and weight (kg). Each row corresponds to eigenprotein of the module. Each cell presents the correlation and relevant *p*-value. (C) Scatter plot of protein significance for BMD group and module membership in the yellow module. This figure shows significant positive correlation between protein significance and module membership. Since highly connected intramodular hub proteins tend to have high module membership values, the significant positive correlation illustrating that proteins highly significantly associated with BMD group were the most essential elements of the yellow module.

interacting network, except for the UBB–LCP1, the combined predicted score for other pairwise association between proteins all reached the level of high confidence.

3.4.4 Significant genes identified in differential expression analysis and network analyses were replicated at mRNA level

Seven genes were found significantly differentially expressed according to proteomic profiling, among which three of them (PPIA, ANXA2, and ANXA2P2) reached the significance level of 0.05 in 80 Caucasian female microarray dataset. None of these genes were found significant in 73 Caucasian female microarray dataset. The detailed information is provided in Table 7.

Based on the network analysis of protein expression data, ACTN2 and CAPZA1 were two core enrichment genes in the GSEA result, and they were upregulated in the low-BMD group based on the ES. The VCP gene was found in the module to be significantly related to BMD variation by WGCNA, and it was upregulated in the postmenopausal low-BMD

Table 5. Significant GO term in the enrichment analysis in the target module

Term	*p*-Value	Fold of enrichment	Bonferroni-corrected *p*-value	Benjamini-corrected *p*-value	FDR
GO:0008289	2.06×10^{-5}	10.3516	0.0016	0.0016	0.0218
GO:0005543	9.74×10^{-5}	12.0769	0.0075	0.0037	0.1027
GO:0005544	9.74×10^{-5}	12.0769	0.0075	0.0037	0.1027

group. Consistent with those findings at the protein level, mRNA expression of ACTN2, CAPZA1, and VCP was negatively associated with BMD with adjusted *p*-values less than 0.05. Further detail is provided in Table 8.

3.4.5 Significant genes identified in differential expression analysis and network analyses were replicated at DNA level

For the seven significantly differential expressed genes identified in our proteomics study, SNPs of YWHAB, ANXA2, and ANXA2P2 were found significantly associated with hip BMD in female sample of the first GWA meta-analysis [43]. SNPs of ANXA2P2 were found significantly associated with FN-BMD in GEFOS2 female sample, while SNPs of YWHAB, LMNB1, and ANXA2P2 were found significantly associated with FN-BMD in GEFOS2 pooled (with both sexes) sample [43]. The detailed information is provided in Table 7.

For the important genes identified in GSEA and WGCNA, SNPs of ACTN2, EIF4A1, HNRNPK, and MPO were found to be associated with hip BMD ($p < 5 \times 10^{-3}$) in the first GWA meta-analysis as shown in Table 9. By using the GEFOS-2 meta-analysis dataset summary results, SNPs of the ACTN4, CAPZA1, HNRNPK, and RAP1A genes were found to be significantly associated with FNBMD ($p < 5 \times 10^{-3}$), as shown in Table 10. One SNP, rs167203 of the HNRNPK gene, was found to be significant ($p < 0.001$) in both of these two GWA meta-analysis datasets.

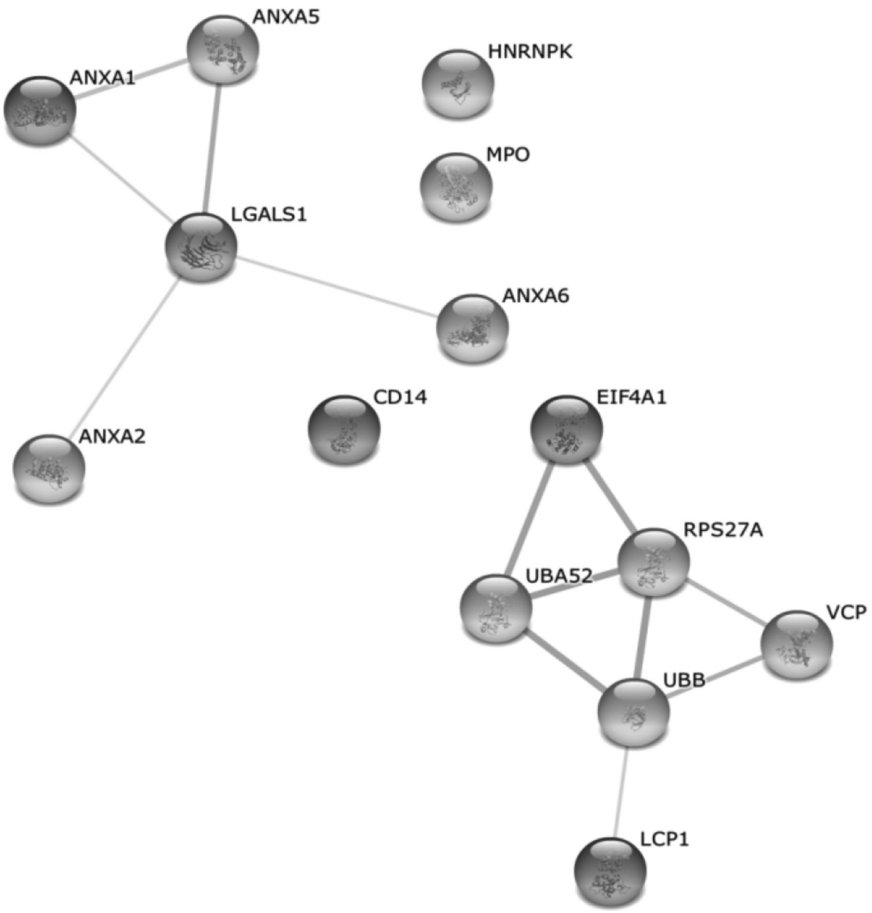

Figure 4. Network view of predicted associations for annexin and ubiquitin protein families in target BMD-related module. The network nodes represent proteins, and lines represent the existence of evidence used in predicting the associations. Darker connecting lines between nodes represent higher combined score.

Table 6. Protein–protein interacting subnetwork identified in the BMD-related module

Pairwise association	Score	Evidence
Annexin-related protein interacting subnetwork		
ANXA1–ANXA5	0.632	Comentioned in PubMed abstracts coexpression experimental/biochemical data
ANXA1–LGALS1	0.516	Comentioned in PubMed abstracts coexpression
ANXA2–LGALS1	0.460	Comentioned in PubMed abstracts coexpression
ANXA5–LGAS1	0.851	Comentioned in PubMed abstracts coexpression
LGALS1–ANXA6	0.471	Co-mentioned in PubMed abstracts
Ubiquitin-related protein interacting subnetwork		
RPS27A–UBA52	0.977	Comentioned in PubMed abstracts coexpression experimental/biochemical data association in curated databases cooccurence across genomes
RPS27A–UBB	0.971	Comentioned in PubMed abstracts coexpression experimental/biochemical data
RPS27A–VCP	0.755	Comentioned in PubMed abstracts experimental/biochemical data
RPS27A–EIF4A1	0.922	Comentioned in PubMed abstracts coexpression experimental/biochemical data association in curated databases
UBA52–UBB	0.942	Comentioned in PubMed abstracts coexpression experimental/biochemical data association in curated databases cooccurence across genomes
UBA52–EIF4A1	0.926	Comentioned in PubMed abstracts coexpression experimental/biochemical data association in curated databases
UBB–LCP1	0.504	Experimental/biochemical data
UBB–VCP	0.819	Comentioned in PubMed abstracts experimental/biochemical data

Score indicates the confidence of protein interacting pattern based on known evidence and predicted methodology (low confidence: score < 0.4; medium: 0.4–0.7; high: >0.7)

3.4.6 CAPZA1 and ACTN2 were significantly associated with hip BMD in integrative analysis

By integrating information from the six independent studies at the DNA, RNA, and protein levels, CAPZA1 ($p < 0.0001$) and ACTN2 ($p = 0.0080$), found in GSEA, were significantly associated with hip BMD level.

4 Discussion

The present study applied traditional differential expression analysis and state-of-the-art network based analysis to proteomics data. To the best of our knowledge, this is the first study using both GSEA and WGCNA strategies to explore the PBM proteome for studying the pathogenesis of postmenopausal osteoporosis.

We used the Miltenyi Monocyte Isolation Kit II to specifically isolate the classic PBM subset ($CD14^{++}CD16^{-}$), which contains the osteoclast precursors [8, 9]. Though additional markers such as $RANK^{+}$ may be used to further characterize osteoclast-committed precursors in PBM subsets, these subpopulations did not significantly induce more cumulative resorption than unfractionated $CD14^{++}$ cells [47], indicating that mature osteoclasts with similar functionality and lifespan can be derived from the unfractionated PBMs. Therefore, the classic PBMs used in this study were "homogeneous" with respect to their differentiability to mature osteoclasts. We successfully applied the quantitative proteomics methodology to profile, and compare, proteomes in high- and low-BMD groups. Since the overlap of identified proteins across different groups was quite high, and the distribution of the number of detected proteins across high- and low-BMD samples varied only at narrow ranges, the stability and reliability of our proteomic experimental workflow was substantiated.

Differential expression analysis showed that TAGLN2 was significantly downregulated in the postmenopausal low-BMD group. As a marker of smooth muscle cell differentiation, transgelin-2 has been reported to regulate angiogenesis, which is closely related to bone formation and resorption [48]. In addition, knock down of the transgelin-2 gene in mice significantly increased production of bone marrow mesenchymal stem cells [49]. Since the interaction between mesenchymal stem cells and monocyte-related osteoclast progenitors plays a vital role in osteoclast differentiation [50], our results suggest that decreased levels of transgelin-2 protein could be a candidate marker for low BMD in postmenopausal females. For the other three proteins that were downregulated in the low-BMD group in our study (encoded by PPIA, LOC654118, and YWHAB), no previous reports have provided experimental verification of their significance to bone pathophysiology. Proteins like annexin A2, lamin B1, and annexin A2 were upregulated in the low-BMD group. The annexin A2 protein is known to stimulate monocyte trans-endothelial migration, and the up-regulation of this protein in subjects with low BMD has been verified previously by Western blot [10].

In the GSEA analysis, the two most significantly enriched gene sets were "platelet activation, signaling, and aggregation" and "hemostasis." The core enrichment genes in these two gene sets include ACTN, ACTN2, ACTN4, ALB, ALDOA, CAPZA1, FGA, FGB, FGG, LOC654188, PFN1, PPBP, PPIA, RAP1A, RAP1B, and VCL. Based on previous evidence, many

Table 7. Comparison of *p*-values for identified differentially expressed genes in low- versus high-BMD at different omics levels

Protein	Microarray analysis		GWAS meta-analyses			
	73 Caucasian	80 Caucasian	7-meta (female)	7-meta (pooled)	GEFOS2 (female)	GEFOS2 (pooled)
PPIA	0.3767	0.0108	0.1037	0.1054	0.4434	0.6368
LOC654188	–	–	–	–	–	–
YWHAB	0.2917	0.5599	0.0128	0.6890	0.0854	0.0492
TAGLN2	0.2039	0.0592	0.1746	0.2529	0.3975	0.2154
ANXA2	0.2099	0.0305	0.0351	0.1182	0.1288	0.0516
LMNB1	0.6451	0.0718	0.1937	0.1813	0.1656	0.0467
ANXA2P2	0.4897	0.0303	0.0055	0.0544	0.0432	0.0101

The GWAS meta-analyses summarize *p*-values for the most significant SNP within the specified gene. "7-meta" represents meta-*p*-values from genetic association between SNPs and hip BMD in female sample and pooled samples, respectively, from GWA meta-analysis including seven GWAS datasets [43]. GEFOS2 represents meta-*p*-values from genetic association between SNPs and FN-BMD in female sample and pooled samples, respectively, in GEFOS2 study [45]. "–" indicates the gene information cannot be found in the dataset.

of the core enrichment genes were involved in regulating bone resorption activities, which lend strong support to our GSEA analyses. For instance, PPIA was found to be downregulated in PBM in low-BMD subjects, and this was consistent with a previous study [10]. RAP1A and RAP1B are two isoforms of the Rap1 protein, and they are critical for osteoclast function [51]. Recently, Zou et al. generated osteoclast-specific Rap1 deletion mice, demonstrating that Rap1 is an essential factor for osteoclast function and important for pathological bone loss [51]. Moreover, an integration of human GWAS and gene expression profiling performed by Hsu et al. revealed chromosome 1p13.2 (where RAP1A is located) is significantly associated with osteoporosis-related traits ($p = 3.6 \times 10^{-8}$) [52]. VCL (vinculin) is an actin-binding protein localized on the cytoplasmic face of integrin-containing podosomes in osteoclasts. It has been shown that insufficient vinculin in osteoclasts could impair osteoclast function, leading to increased bone mass in vivo [53]. Correspondingly, increased vinculin can organize the osteoclast cytoskeleton and promote bone resorption during osteoclastogenesis [53]. PPBP (proplatelet basic protein), also named CXCL7, is a platelet-derived growth factor that belongs to the CXC chemokine family. It has been demonstrated that CXCL7 can work with SDF1 to promote the formation of giant osteoclasts in mouse, which would contribute to the bone resorption process [54]. The three fibrinogen genes (FGG, FGA, FGB) and three actin genes (ACTN1, ACTN2, ACTN4) were also found important in modulating monocyte activities. Several studies have shown that fibrinogen can bind to monocytes and regulate monocyte functions, such as adhesion processes and the production of cytokines relevant to postmenopausal osteoporosis (e.g., IL-8, IL-6, and TNF-α) [55, 56]. With respect to actin-related genes, earlier studies suggested that α-actins are associated with transcription factor NF-κB, which could regulate monocyte maturation

Table 8. Significant probeset identified at mRNA level

Gene	Probeset	Fold change (H vs. L)	Raw *p*	Adjusted *p*
ACTN2	203861_s_at	–0.1305	0.0008	0.0402
CAPZA1	217392_at	–0.0603	0.0002	0.0225
VCP	208649_s_at	–0.1390	0.0009	0.0422

H vs. L means high hip BMD versus low hip BMD group. Adjust *p*-value is obtained from multiple comparison.

Table 9. Significant SNPs in ACTN2, EIF4A1, HNRNPK, and MPO associated with hip BMD in the first GWA meta-analysis

Gene symbol	Marker name	Physical position	Allele	*p*-Value
ACTN2	rs1366990*	Chr1:234948326	t/g	0.0019
EIF4A1	rs12942088	Chr17:7423503	c/t	0.0024
HNRNPK	rs167203	Chr9:85783134	c/g	0.0009
	rs296887	Chr9:85784890	t/c	0.0014
	rs696825	Chr9:85772896	t/c	0.0002
	rs796004	Chr9:85784618	t/c	0.0010
MPO	rs11575868*	Chr17:53707882	t/c	0.0021
	rs7208693*	Chr17:53712817	a/c	0.0005

SNPs with asterisk indicate the results from the female samples; otherwise the results are from the pooled samples of both males and females. Allele before "/" is the minor allele.

Table 10. Significant SNPs in ACTN4, CAPZA1, RAPA1, and HNRNPK associated with FNBMD in GEFOS2 dataset

Gene symbol	Marker name	Physical position	Allele	*p*-Value
ACTN4	rs2368475*	Chr19:43887575	t/c	0.0044
CAPZA1	rs12130243	Chr1:112971190	g/t	0.0001
	rs12120956	Chr1:113004094	a/g	0.0002
	rs3790598	Chr1:112998419	a/g	0.0004
	rs12756815	Chr1:113008766	a/c	0.0005
	rs11584092	Chr1:112977785	a/t	0.0044
HNRNPK	rs167203	Chr9:85783134	c/g	0.0002
	rs296887	Chr9:85784890	t/c	0.0012
RAP1A	rs1886498*	Chr1:112038561	a/c	0.0033
	rs3767595*	Chr1:112046153	t/c	0.0036
	rs4838920*	Chr1:112046582	a/g	0.0049
	rs10489469*	Chr1:112035607	t/g	0.0050

SNPs with asterisk indicate the results from the GEFOS2 female samples; otherwise the results are from the pooled samples of both sex. Allele before "/" is the minor allele.

and cytokine production [57]. In addition, based on previous research on cytoskeletal reorganization during early attachment of monocytes, conspicuous activities of actin and vinculin were detected [58]. The identification of actin-related genes as core enrichment genes in the current study supports the concept that cytoskeletal organization influences monocyte migration, which could influence the rate of bone resorption in postmenopausal low-BMD subjects.

Our finding that the hemostasis gene set was enriched in the current analysis is relevant. The hemostatic balance of various physiologic and pathologic conditions is influenced substantially by levels of pro-inflammatory cytokines, many of which have also been associated with osteoporosis in postmenopausal osteoporosis women [59]. Therefore, our GSEA of the PBM proteome not only identified several important genes that may be overlooked by traditional comparative analyses, but also suggested the potential influence of pro-inflammatory cytokines and cytoskeleton-related processes within PBM's on BMD in postmenopausal women.

According to the WGCNA analysis, the module containing annexin family genes was significantly associated with low BMD. Genes including ANXA1, ANXA2, ANXA5, and ANXA6 were found in this module, and they were consistently upregulated in the low-BMD group. Previous studies have discovered the essential contributions of the annexin protein family to bone mineralization and regulation of osteoclast activity [60–62], which is consistent with our findings by the WGCNA method. The ANXA2 protein and putative annexin A2-like protein were most significantly ($p < 0.01$) associated with BMD group (defined by the BMD extreme values) in this module. In the GO enrichment analysis of the target module, the GO terms "lipid binding," "phospholipid binding," and "calcium-dependent phospholipid binding" were identified. All three of these GO terms are relevant to lipid-binding processes and include proteins of the ANXA family. ANXAs, which are Ca^{2+}- and lipid-binding proteins, have previously been shown to contribute to homeostasis in bone cells and extracellular matrix vesicles [63], and it has been suggested that ANXA2 stimulates osteoclast formation [64]. These previous studies of ANXAs, which implicate them in various bone-related molecular processes, lend support for the results of our current findings based on network analysis, that lipid-binding GO terms were significantly enriched in the postmenopausal low-BMD group.

Most proteomics studies have been focused on the function of individual differentially expressed proteins instead of analyzing interacting and functionally orchestrating protein groups. With the WGCNA method, we generated some notable information regarding closely coexpressed protein groups that would not have been found using traditional differential expression analyses. According to the results given by STRING software, we found two interacting protein subgroups in the target BMD-related module with comparatively strong known or predicted functional associations. The first protein–protein interaction subnetwork contained ANXA1, ANXA2, ANXA5, ANXA6, and LGALS1. Interactions between ANXA1 and ANXA5 have been detected by colocalization assays in the apoptosis process [65], and the highly combined prediction score between them in the identified module suggests that they interact with each other and function together in PBM in a manner that contributes to the pathogenesis of postmenopausal osteoporosis. LGALS1 (galectin-1) can stimulate monocyte migration in a dose-dependent manner and has been strongly implicated in monocyte/macrophage physiology [66]. According to this evidence, identification of the first protein interacting subgroup suggests that annexin family proteins interact with each other; their coexpression pattern with LGASL1 might indicate that these proteins collaborate to impact PBM function. The second highly connected protein subgroup identified in the target module contained UBB, UBA52, RPS27A, LCP1, EIF4A1, and VCP. The first three proteins belong to ubiquitin family, and VCP can function as ubiquitin segregase. Ubiquitins can attach to various substrate proteins, and the ubiquitination of proteins is well known for controlling cytoskeletal dynamics, cell adhesion, and cell migration [67]. Since cytoskeletal

dynamics influence cell-surface expression of monocyte receptors, and thus affect monocyte migration to bone surfaces [68], our findings indicate that ubiquitin-related biological function may influence postmenopausal BMD variation by altering PBM functionality.

With differential expression analysis and network-based analyses, we found several important proteins and corresponding genes that may be closely related to variations in postmenopausal BMD. Significant regulation of PPIA, ANXA2, ANXA2P2, ACTN2, CAPZA1, and VCP was confirmed at the mRNA level in independent cohort study of postmenopausal Caucasian women. These findings further confirmed the functional relevance of these six detected genes in our proteomic discovery study. From the results of the two largest published GWA meta-analyses, at the DNA level, significant SNPs of YWHAB, ANXA2, LMNB1, ANXA2P2, ACTN2, ACTN4, CAPZA1, EIF4A1, MPO, HNRNPK, and RAP1A were associated with BMD in general populations, lending further support for the biological and population importance of these identified genes to BMD. CAPZA1 (F-actin-capping protein subunit alpha-1) and ACTN2 (actinin, alpha 2) were core enrichment genes in the GSEA analysis, and were found to be significantly associated with hip BMD in the multilevel integrative analyses. Based on previous studies, the major function of capping protein is to control the length of actin filaments, thus maintaining their stability [69]. It has been previously reported that actin ring formation is a prerequisite for osteoclast bone resorption, and that the actin ring is the specific cytoskeletal structure required for osteoclast functionality [70, 71]. Consequently, the findings from our comprehensive study provide further support for the concept that actin-related genes (e.g., CAPZA1 and ACTN2) contribute to the pathogenesis of osteoporosis through their specific actions in PBMs.

Integrative network based analysis with proteomics data holds great promise for illuminating the molecular and genetic basis of complex diseases. Though our experimental sample was relatively small (but still the largest and the first in the bone field), the concordant findings among proteomic, transcriptomic, and genomic information provides substantial support for the reliability and power of this study approach. The present study demonstrated that both GSEA and WGCNA were useful for finding postmenopausal osteoporosis-related genes and biologically meaningful gene groups. Consistent results between protein differential expression analysis and network-based analysis further showed the utility of network analysis and the importance of the ANXA gene family in postmenopausal osteoporosis. In addition, both GSEA and WGCNA analyses revealed some noteworthy proteins and protein groups, which would escape detection by regular differential expression analysis. Since proteins do not usually work independently, the significant coexpression pattern or enrichment of proteins groups can help reveal how proteins function within PBMs based on protein networks. The abnormal behavior of PBM activities such as enhanced differentiation or migration could contribute to the development of osteoporosis; the network-based analysis thus may provide us with much more meaningful and comprehensive, systematic biological insight for PBM activities and significance. This integrative analysis strategy not only helps to reveal disease-associated genes through proteomic data, but also sheds light on potential therapeutic interventions for postmenopausal osteoporosis.

We would like to thank all staff who made contribution to this study. 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 Framingham Heart Study (FHS) datasets used for the analysis described in this article were acquired from dbGap (http://www.ncbi.nlm.nih.gov/gap/) with accession number phs000007.v14.p6. The IFS datasets used for the analyses were obtained from dbGap with accession number phs000138.v2.p1. The Women's Health Initiative (WHI) datasets used for the analyses were attained from dbGaP with accession number phs000200.v6.p2. This article was not prepared in cooperation with researchers of the FHS and does not represent the views of the FHS, Boston University, or National Heart, Lung, and Blood Institute (NHLBI). In addition, this article was not prepared in collaboration with investigators of the WHI, and does not reflect their opinions.

The FHS is supported and conducted by the NHLBI in collaboration with Boston University (contract no. N01-HC-25195). Affymetrix genotyping was supported by NHLBI (contract no. N02-HL-64278) and Illumina genotyping was performed with an agreement between Illumina and Boston University. In addition, Framingham Bone Mineral Density datasets were supported by NIH grant R01AR/AG 41398. The Genetic Determinants of Bone Fragility (Indiana Fragility Study) was supported through NIA Division of Geriatrics and Clinical Gerontology (DGCG). Genetic Determinants of Bone Fragility is a GWA study funded as part of the NIA DGCG. Collection of datasets and samples were provided by the parent grant, Genetic Determinants of Bone Fragility (P01-AG018397). The NIH and NIA supported genotyping performed at the Johns Hopkins University Center for Inherited Diseases Research. The WHI program is supported by the NHLBI, NIH, U.S. Department of Health and Human Services, (contract no. N01WH22110, 24152, 32100–2, 32105–6, 32108–9, 32111–13, 32115, 32118, 32119, 32122, 42107–26, 42129–32, and 44221). WHI Population Architecture Using Genomics and Epidemiology (PAGE) is funded through the NHGRI network (grant no. U01HG004790).All authors approved the final content of the article.

The authors have declared no conflict of interest.

5 References

[1] Lerner, U. H., Bone remodeling in post-menopausal osteoporosis. *J. Dent. Res.* 2006, *85*, 584–595.

[2] Melton, L. J., 3rd, Chrischilles, E. A., Cooper, C., Lane, A. W., Riggs, B. L., Perspective. How many women have osteoporosis? *J. Bone Miner. Res.* 1992, *7*, 1005–1010.

[3] Aggarwal, N., Raveendran, A., Khandelwal, N., Sen, R. K. et al., Prevalence and related risk factors of osteoporosis in peri- and postmenopausal Indian women. *J. Midlife Health* 2011, *2*, 81–85.

[4] Sioka, C., Fotopoulos, A., Georgiou, A., Xourgia, X. et al., Age at menarche, age at menopause and duration of fertility as risk factors for osteoporosis. *Climacteric* 2010, *13*, 63–71.

[5] Geissmann, F., Manz, M. G., Jung, S., Sieweke, M. H. et al., Development of monocytes, macrophages, and dendritic cells. *Science* 2010, *327*, 656–661.

[6] Soudja, S. M., Ruiz, A. L., Marie, J. C., Lauvau, G., Inflammatory monocytes activate memory CD8(+) T and innate NK lymphocytes independent of cognate antigen during microbial pathogen invasion. *Immunity* 2012, *37*, 549–562.

[7] Auffray, C., Sieweke, M. H., Geissmann, F., Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu. Rev. Immunol.* 2009, *27*, 669–692.

[8] Zhou, Y., Deng, H. W., Shen, H., Circulating monocytes: an appropriate model for bone-related study. *Osteoporos. Int.* 2015, *26*, 2561–2572.

[9] Komano, Y., Nanki, T., Hayashida, K., Taniguchi, K., Miyasaka, N., Identification of a human peripheral blood monocyte subset that differentiates into osteoclasts. *Arthritis Res. Ther.* 2006, *8*, R152.

[10] Deng, F. Y., Lei, S. F., Zhang, Y., Zhang, Y. L. et al., Peripheral blood monocyte-expressed ANXA2 gene is involved in pathogenesis of osteoporosis in humans. *Mol. Cell. Proteomics* 2011, *10*, M111.011700.

[11] Liu, Y. Z., Dvornyk, V., Lu, Y., Shen, H. et al., A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes. *J. Biol. Chem.* 2005, *280*, 29011–29016.

[12] Mormann, M., Thederan, M., Nackchbandi, I., Giese, T. et al., Lipopolysaccharides (LPS) induce the differentiation of human monocytes to osteoclasts in a tumour necrosis factor (TNF) alpha-dependent manner: a link between infection and pathological bone resorption. *Mol. Immunol.* 2008, *45*, 3330–3337.

[13] Lari, R., Kitchener, P. D., Hamilton, J. A., The proliferative human monocyte subpopulation contains osteoclast precursors. *Arthritis Res. Ther.* 2009, *11*, R23.

[14] Hirayama, T., Danks, L., Sabokbar, A., Athanasou, N. A., Osteoclast formation and activity in the pathogenesis of osteoporosis in rheumatoid arthritis. *Rheumatology (Oxford)* 2002, *41*, 1232–1239.

[15] Dratva, J., Gomez Real, F., Schindler, C., Ackermann-Liebrich, U. et al., Is age at menopause increasing across Europe? Results on age at menopause and determinants from two population-based studies. *Menopause* 2009, *16*, 385–394.

[16] Kim, O. Y., Chae, J. S., Paik, J. K., Seo, H. S. et al., Effects of aging and menopause on serum interleukin-6 levels and peripheral blood mononuclear cell cytokine production in healthy nonobese women. *Age* 2012, *34*, 415–425.

[17] Gohlke-Barwolf, C., Coronary artery disease – is menopause a risk factor? *Basic Res. Cardiol.* 2000, *95* Suppl 1, I77–I83.

[18] Qiu, C., Chen, H., Wen, J., Zhu, P. et al., Associations between age at menarche and menopause with cardiovascular disease, diabetes, and osteoporosis in Chinese women. *J. Clin. Endocrinol. Metab.* 2013, *98*, 1612–1621.

[19] Faienza, M. F., Ventura, A., Marzano, F., Cavallo, L., Postmenopausal osteoporosis: the role of immune system cells. *Clin. Dev. Immunol.* 2013, *2013*, 575936.

[20] Ben-Hur, H., Mor, G., Insler, V., Blickstein, I. et al., Menopause is associated with a significant increase in blood monocyte number and a relative decrease in the expression of estrogen receptors in human peripheral monocytes. *Am. J. Reprod. Immunol.* 1995, *34*, 363–369.

[21] Phiel, K. L., Henderson, R. A., Adelman, S. J., Elloso, M. M., Differential estrogen receptor gene expression in human peripheral blood mononuclear cell populations. *Immunol. Lett.* 2005, *97*, 107–113.

[22] Perrien, D. S., Achenbach, S. J., Bledsoe, S. E., Walser, B. et al., Bone turnover across the menopause transition: correlations with inhibins and follicle-stimulating hormone. *J. Clin. Endocrinol. Metab.* 2006, *91*, 1848–1854.

[23] Robinson, L. J., Yaroslavskiy, B. B., Griswold, R. D., Zadorozny, E. V. et al., Estrogen inhibits RANKL-stimulated osteoclastic differentiation of human monocytes through estrogen and RANKL-regulated interaction of estrogen receptor-alpha with BCAR1 and Traf6. *Exp. Cell Res.* 2009, *315*, 1287–1301.

[24] Hong, D., Chen, H. X., Yu, H. Q., Liang, Y. et al., Morphological and proteomic analysis of early stage of osteoblast differentiation in osteoblastic progenitor cells. *Exp. Cell Res.* 2010, *316*, 2291–2300.

[25] Kim, W. K., Bae, K. H., Choi, H. R., Kim, D. H. et al., Leukocyte common antigen-related (LAR) tyrosine phosphatase positively regulates osteoblast differentiation by modulating extracellular signal-regulated kinase (ERK) activation. *Mol. Cells* 2010, *30*, 335–340.

[26] Lee, J. H., Cho, J. Y., Proteomics approaches for the studies of bone metabolism. *BMB Rep.* 2014, *47*, 141–148.

[27] Deng, F. Y., Liu, Y. Z., Li, L. M., Jiang, C. et al., Proteomic analysis of circulating monocytes in Chinese premenopausal females with extremely discordant bone mineral density. *Proteomics* 2008, *8*, 4259–4272.

[28] Isserlin, R., Merico, D., Alikhani-Koupaei, R., Gramolini, A. et al., Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps. *Proteomics* 2010, *10*, 1316–1327.

[29] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S. et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 15545–15550.

[30] Clutterbuck, A. L., Smith, J. R., Allaway, D., Harris, P. et al., High throughput proteomic analysis of the secretome in an explant model of articular cartilage inflammation. *J. Proteomics* 2011, *74*, 704–715.

[31] Cha, S., Imielinski, M. B., Rejtar, T., Richardson, E. A. et al., In situ proteomic analysis of human breast cancer epithelial

cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology. *Mol. Cell. Proteomics* 2010, *9*, 2529–2544.

[32] Zhang, B., Horvath, S., A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 2005, *4*, Article17.

[33] Langfelder, P., Horvath, S., WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, *9*, 559.

[34] Strauss-Ayali, D., Conrad, S. M., Mosser, D. M., Monocyte subpopulations and their differentiation patterns during infection. *J Leukoc Biol* 2007, *82*, 244–252.

[35] Albert, R., Scale-free networks in cell biology. *J Cell Sci* 2005, *118*, 4947–4957.

[36] Huang da, W., Sherman, B. T., Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009, *4*, 44–57.

[37] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M. et al., STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013, *41*, D808–D815.

[38] von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D. et al., STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 2005, *33*, D433–D437.

[39] Chen, X. D., Xiao, P., Lei, S. F., Liu, Y. Z. et al., Gene expression profiling in monocytes and SNP association suggest the importance of the STAT1 gene for osteoporosis in both Chinese and Caucasians. *J. Bone Miner. Res.* 2010, *25*, 339–355.

[40] Deng, F. Y., Zhu, W., Zeng, Y., Zhang, J. G. et al., Is GSN significant for hip BMD in female Caucasians? *Bone* 2014, *63*, 69–75.

[41] Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M. et al., Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 2003, *31*, e15.

[42] Kooperberg, C., Aragaki, A., Strand, A. D., Olson, J. M., Significance testing for small microarray experiments. *Stat. Med.* 2005, *24*, 2281–2298.

[43] Zhang, L., Choi, H. J., Estrada, K., Leo, P. J. et al., Multistage genome-wide association meta-analyses identified two new loci for bone mineral density. *Hum. Mol. Genet.* 2014, *23*, 1923–1933.

[44] Rivadeneira, F., Styrkarsdottir, U., Estrada, K., Halldorsson, B. V. et al., Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat. Genet.* 2009, *41*, 1199–1206.

[45] Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y. H. 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.

[46] Tyekucheva, S., Marchionni, L., Karchin, R., Parmigiani, G., Integrating diverse genomic data using gene sets. *Genome Biol.* 2011, *12*, R105.

[47] Atkins, G. J., Kostakis, P., Vincent, C., Farrugia, A. N. et al., RANK expression as a cell surface marker of human osteoclast precursors in peripheral blood, bone marrow, and giant cell tumors of bone. *J. Bone Miner. Res.* 2006, *21*, 1339–1349.

[48] Ma, J., Wu, Y., Zhang, W., Smales, R. J. et al., Up-regulation of multiple proteins and biological processes during maxillary expansion in rats. *BMC Musculoskelet. Disord.* 2008, *9*, 37.

[49] Kuo, H. C., Chiu, C. C., Chang, W. C., Sheen, J. M. et al., Use of proteomic differential displays to assess functional discrepancies and adjustments of human bone marrow- and Wharton jelly-derived mesenchymal stem cells. *J. Proteome Res.* 2011, *10*, 1305–1315.

[50] Mbalaviele, G., Jaiswal, N., Meng, A., Cheng, L. et al., Human mesenchymal stem cells promote human osteoclast differentiation from $CD34^+$ bone marrow hematopoietic progenitors. *Endocrinology* 1999, *140*, 3736–3743.

[51] Zou, W., Izawa, T., Zhu, T., Chappel, J. et al., Talin1 and Rap1 are critical for osteoclast function. *Mol. Cell Biol.* 2013, *33*, 830–844.

[52] Hsu, Y. H., Zillikens, M. C., Wilson, S. G., Farber, C. R. et al., An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet.* 2010, *6*, e1000977.

[53] Fukunaga, T., Zou, W., Warren, J. T., Teitelbaum, S. L., Vinculin regulates osteoclast function. *J. Biol. Chem.* 2014, *289*, 13554–13564.

[54] Nakao, K., Aoyama, M., Fukuoka, H., Fujita, M. et al., IGF2 modulates the microenvironment for osteoclastogenesis. *Biochem. Biophys. Res. Commun.* 2009, *378*, 462–466.

[55] Kuhns, D. B., Priel, D. A., Gallin, J. I., Induction of human monocyte interleukin (IL)-8 by fibrinogen through the toll-like receptor pathway. *Inflammation* 2007, *30*, 178–188.

[56] Babakov, V. N., Petukhova, O. A., Turoverova, L. V., Kropacheva, I. V. et al., RelA/NF-kappaB transcription factor associates with alpha-actinin-4. *Exp. Cell Res.* 2008, *314*, 1030–1038.

[57] Inada, M., Miyaura, C., [Cytokines in bone diseases. Cytokine and postmenopausal osteoporosis]. *Clin. Calcium* 2010, *20*, 1467–1472.

[58] Lehto, V. P., Hovi, T., Vartio, T., Badley, R. A., Virtanen, I., Reorganization of cytoskeletal and contractile elements during transition of human monocytes into adherent macrophages. *Lab. Invest.* 1982, *47*, 391–399.

[59] Grignani, G., Maiolo, A., Cytokines and hemostasis. *Haematologica* 2000, *85*, 967–972.

[60] Kim, T. H., Hong, J. M., Shin, E. S., Kim, H. J. et al., Polymorphisms in the annexin gene family and the risk of osteonecrosis of the femoral head in the Korean population. *Bone* 2009, *45*, 125–131.

[61] Orimo, H., The mechanism of mineralization and the role of alkaline phosphatase in health and disease. *J. Nippon Med. Sch.* 2010, *77*, 4–12.

[62] Roodman, G. D., Regulation of osteoclast differentiation. *Ann. NY Acad. Sci.* 2006, *1068*, 100–109.

[63] Balcerzak, M., Hamade, E., Zhang, L., Pikula, S. et al., The roles of annexins and alkaline phosphatase in mineralization process. *Acta Biochim. Pol.* 2003, *50*, 1019–1038.

[64] Menaa, C., Devlin, R. D., Reddy, S. V., Gazitt, Y. et al., Annexin II increases osteoclast formation by stimulating the proliferation of osteoclast precursors in human marrow cultures. *J. Clin. Invest.* 1999, *103*, 1605–1613.

[65] Arur, S., Uche, U. E., Rezaul, K., Fong, M. et al., Annexin I is an endogenous ligand that mediates apoptotic cell engulfment. *Dev. Cell* 2003, *4*, 587–598.

[66] Malik, R. K., Ghurye, R. R., Lawrence-Watt, D. J., Stewart, H. J., Galectin-1 stimulates monocyte chemotaxis via the p44/42 MAP kinase pathway and a pertussis toxin-sensitive pathway. *Glycobiology* 2009, *19*, 1402–1407.

[67] Schaefer, A., Nethe, M., Hordijk, P. L., Ubiquitin links to cytoskeletal dynamics, cell adhesion and migration. *Biochem. J.* 2012, *442*, 13–25.

[68] Wojciak-Stothard, B., Williams, L., Ridley, A. J., Monocyte adhesion and spreading on human endothelial cells is dependent on Rho-regulated receptor clustering. *J. Cell Biol.* 1999, *145*, 1293–1307.

[69] Eierman, D. F., Johnson, C. E., Haskill, J. S., Human monocyte inflammatory mediator gene expression is selectively regulated by adherence substrates. *J. Immunol.* 1989, *142*, 1970–1976.

[70] Woo, J. T., Kasai, S., Stern, P. H., Nagai, K., Compactin suppresses bone resorption by inhibiting the fusion of prefusion osteoclasts and disrupting the actin ring in osteoclasts. *J. Bone Miner. Res.* 2000, *15*, 650–662.

[71] Chellaiah, M. A., Regulation of actin ring formation by rho GTPases in osteoclasts. *J. Biol. Chem.* 2005, *280*, 32930–32943.

