Published in final edited form as:

Osteoporos Int. 2018 December ; 29(12): 2645–2658. doi:10.1007/s00198-018-4654-y.

Gene-based GWAS Analysis for Consecutive Studies of GEFOS

Wei Zhu^{1,2}, Chao Xu², Ji-Gang Zhang², Hao He², Ke-Hao Wu², Lan Zhang², Yong Zeng^{2,3}, Yu Zhou², Kuan-Jui Su², and Hong-Wen Deng^{1,2}

¹College of Life Sciences, Hunan Normal University, Changsha, Hunan, China 410081

²Center for Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA 70112

³College of Life Sciences and Bioengineering, Beijing Jiaotong University, Beijing, China 100044

Abstract

Purpose—Single-nucleotide polymorphism (SNP)-based genome-wide association studies (GWASs) have already identified about 100 loci associated with bone mineral density (BMD), but these loci only explain a small proportion of heritability to osteoporosis risk. In the present study, we performed a gene-based analysis of the largest GWASs in the bone field to identify additional BMD-associated genes.

Methods—BMD-associated genes were identified by combining the summary statistic *P* values of SNPs across individual genes in the two consecutive meta-analyses of GWASs from the Genetic Factors for Osteoporosis (GEFOS) studies. The potential functionality of these genes to bone was partially assessed by differential gene expression analysis. Additionally, the consistency of the identification of potential BMD-associated variants were evaluated by estimating the correlation of the *P* values of the same SNPs/genes between the two consecutive GEFOS studies with largely overlapping samples.

Results—Compared to the SNP-based analysis, the gene-based strategy identified additional BMD-associated genes with genome-wide significance and increased their mutual replication between the two GEFOS datasets. Among these BMD-associated genes, 3 novel genes (*UBTF*, *AAAS*, and *C11orf58*) were partially validated at the gene expression level. The correlation analysis presented a moderately high between-study consistency of potential BMD-associated variants.

Conclusions—Gene-based analysis as a supplementary strategy to SNP-based GWAS, when applied here, is shown that it helped identify some novel BMD-associated genes. In addition to its empirically increased statistical power, gene-based analysis also provides a higher testing stability for identification of BMD genes.

Corresponding Author Hong-Wen Deng, Ph.D., Professor, Director, Center for Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 1610, New Orleans, LA 70112, Tel: 504-988-1310
hdeng2@tulane.edu.

Author Contributions

All authors have given approval to the final version of the manuscript.

Conflict of Interest Disclosure:

Wei Zhu, Chao Xu, Ji-Gang Zhang, Hao He, Ke-Hao Wu, Lan Zhang, Yong Zeng, Yu Zhou, Kuan-Jui Su and Hong-Wen Deng declare that they have no conflict of interest.

Mini Abstract

By integrating the multilevel biological evidence and bioinformatics analyses, the present study represents a systemic endeavor to identify BMD-associated genes and their roles in skeletal metabolism.

Keywords

BMD; osteoporosis; GEFOS; gene-based analysis

Introduction

Osteoporosis is the most common chronic skeletal metabolic disease, characterized by a reduction in bone mass and micro-architectural deterioration in bone tissues that leads to increased fracture risk. Measurement of bone mineral density (BMD) has been used extensively for the clinical diagnosis of osteoporosis and prediction of osteoporotic fractures [1]. Variations in BMD at the prevalent sites of osteoporotic fracture can be partially explained by genetic effects [2]. In recent years, genome-wide association studies (GWASs) have focused on exploring BMD-associated single-nucleotide polymorphisms (SNPs) at a genome-wide significance level (i.e., $p < 5 \times 10^{-8}$), and they have already identified ~100 candidate genes/loci [3]. However, those identified genetic variants only explain a small proportion (~6%) of BMD heritability [4].

Generally, independent GWAS studies provided a relative low efficiency for discovering candidate variants of complex diseases due to poor determining power. Meta-analysis, a statistical synthesis of information from multiple independent studies, has been widely applied to search for missing heritability of complex diseases by increasing testing power [5]. However, the replicability of associated variants between two individual meta-analyses was generally low either with or without overlapping study populations [6]. This poor replication may be due in large part to between-study heterogeneity [7], that is, differences of study designs and analytical approaches, linkage disequilibrium (LD) patterns, and allele frequencies and/or effect sizes [5, 8], or the independent re-sampling of the sample populations. In the bone field, the Genetic Factors for Osteoporosis (GEFOS) Consortium has developed consecutive meta-analyses of GWASs with the largest sample sizes, with the aim to identify genetic variants related to BMD [9–11]. The two meta-analyses of GEFOS2 and GEFOS-seq were performed by the same consortium based on largely overlapping populations and particularly overlapping samples, so the control of the heterogeneity may improve the between-study replication of genetic variants with BMD for consecutive studies within the same GWAS consortium.

It is difficult to understand disease risk only based on GWAS results because the vast majority of disease-associated SNPs identified by GWASs are located in non-coding regions of the genome [12], and they are more possible to influence the diseases by regulating the gene expression levels rather than changing protein structure and function. Because the disease-associated SNPs are frequently enriched in regions implicated in gene regulation (e.g., *cis*-regulatory modules), epigenetic mechanisms are a possible explanation of the

limited impact of GWAS-identified BMD-associated loci. Additionally, the GWAS-identified SNPs serve only as representatives for all the SNPs in the linkage disequilibrium blocks, so other SNPs in high LD with the GWAS-identified SNPs are also possibly causal for the disease. Therefore, the most significantly associated SNPs identified in GWASs may not intuitively participate in the regulation of the diseases/traits' variation. Currently, two approaches, LD calculation with the genomic reference panels to identify SNPs in high LD with GWAS-identified SNPs [13] and fine mapping [14], have been commonly used to analyze SNPs other than the identified lead SNPs in GWASs. In the present study, we used the gene-based analysis to evaluate the overall association of genes with phenotypes by integrating the information of SNPs across individual genes, thus limiting our specific study scope here by alleviating the necessity at this step to determine the certainty of causal SNPs for BMDs.

As a follow-up scenario and alternative analysis for SNP-based GWASs, gene-based analysis has several advantages for further mining values of GWASs and expanding their application in subsequent bioinformatics and biological analyses. Firstly, gene-based analysis can improve the power of identifying associated genes by overcoming the potential allelic heterogeneity in SNP-based GWASs [15], when multiple alleles act through one gene to influence a trait [16]. Secondly, the gene-based method is capable of reducing the interference from cohorts' heterogeneity in LD patterns caused by various arrangements of tagging different SNPs to the same causal variants [10]. Thirdly, gene-based analysis also practically reduces the burden of multiple testing for traditional GWASs. Furthermore, the gene is the functional unit of the human genome with high consistency across populations, which is the major target used by most of the subsequent bioinformatics analyses following GWAS (pathway, gene ontology (GO), protein-protein interaction (PPI)) and bench experiments (*in vivo* and/or *in vitro*).

In the present study, we identified BMD-associated genes by gene-based analysis based on the largest GWAS meta-analysis datasets in the bone field, and partially validated the functionality of several identified associated genes using transcriptomic meta-analysis. Furthermore, we conducted a series of correlation analyses to assess the consistency of significantly associated variants between the two consecutive GEFOS studies. Depending on the results of gene-based analysis, the correlation analysis lend support for the higher reliability of discoveries by the gene-based analysis (than the prevailing SNP-based analysis).

Materials and Methods

GWAS dataset

The first dataset used is the GEFOS2 study with discovery samples composed of 17 GWAS cohorts (n=32,961) [10]. The second is the GEFOS-seq study with 9 GWAS cohorts (n=32,965) including a whole-genome sequencing in the UK10K project (n=2,882), a whole-exome sequencing study with five cohorts (n=3,549), and a deep imputation of genotyped samples using a combined UK10K/1000 Genomes reference panel (n=26,534) [11]. The two discovery studies included ~18,400 overlapped individuals from five studies, including the Australasian Osteoporosis Genetics Consortium (AOGC), Framingham Heart

Study (FHS), Rotterdam Study-I (RS-I), Erasmus Rucphen Family Study (ERF) and TwinsUK (TWINsUK). The BMDs at femoral neck (FN) and lumbar spine (LS) were measured in all cohorts, and the forearm (FA) BMD was measured in the samples of AOGC and TWINsUK. All of the BMDs were measured by using dual-energy X-ray absorptiometry scanners as indicated in the original publications.

Gene-based Analysis

By using the Extended Simes procedure method (GATES) [17], we performed gene-based analysis in Knowledge-based mining system for Genome-wide Genetic studies (KGG) (<http://bioinfo.hku.hk/kggweb/>) to identify the BMD-associated genes harboring multiple association SNP signals. This method defines $P_{(1)}, \dots, P_{(m)}$ as the ascending P values of m SNPs of single gene, m_e is the effective number of independent P values among the m SNPs, and $m_e(j)$ is the effective number of independent P values among the top j SNPs with the smallest P values. When all SNPs are in strong LD, the effective number of P values will approach 1. The P values of SNPs of single gene were integrated into an overall P value to represent the association of the gene with a given trait as follows:

$$P_{Gene} = \text{Min}\left(\frac{m_e P_{(j)}}{m_e(j)}\right) \quad (\text{Equation 1}).$$

The null hypothesis of GATES is that no SNP within the gene is associated with the trait, and the alternative is that at least one SNP in the gene is associated with the trait. The input data used in the present study included the P values of ~2.5 million SNPs for FN-BMD and LS-BMD in GEFOS2, and the P values of ~10 million SNPs in GEFOS-seq for FA-BMD, FN-BMD, and LS-BMD, respectively (<http://www.gefos.org/>). The SNPs in the region from 5-kb upstream to 5-kb downstream of each gene were assigned to the gene to calculate the integrated P values, and the physical positions of genes were annotated according to the human genome (hg19) coordinates by KGG. LD was based on the reference of 1000 Genomes Project for European samples, and the reference file (1000G Phase3 v5 Shapeit2 Reference (EUR)) was downloaded from the KGG website (<http://grass.cas.hku.hk/limx/kgg/phasedgty.html>). The Bonferroni correction was applied to correct the multiple testing problem in the genome-wide gene-based analysis ($P_{\text{threshold}}=0.05/N_{\text{genes}}$), where $P_{\text{threshold}}$ is the genome-wide significant threshold and N_{genes} is the number of genes under testing in the genome. The major advantage of this method is that it requires only publically accessible SNP-based summary statistic P values rather than individual-level genotypic data, and the correlation between different markers caused by LD can be estimated based on a reference panel even without original raw genotype information. GATES used the best-SNP picking strategy, where only one SNP-based P value after the multiple testing adjustment is used to produce the set-based P value representing the association of genes with phenotype. Compared with the all-SNP aggregating tests, due to the weak sensitivity of the strategy to the dilution of SNPs with no effect in the set, the analysis of GATES may provide higher testing sensitivity that is probably more suitable for identifying the dominant significance of loci for traits. The Quantile-Quantile (QQ) plots of the gene-based scores were used to illustrate degree of genomic inflation, that is, how well the observed fits the theoretical χ^2 -

distribution [18]. Additionally, the genomic inflation factors (λ) for the BMDs of GEFOS studies were calculated by using R package of GenABEL as numeric indicator for the inflation.

Gene-pair Based Association Analysis

Based on the STRING database, gene-pair based association analysis was conducted in KGG using the *hybrid set-based test* (HYST) [19] to detect significantly connected protein-protein interaction (PPI) pairs. The purpose of gene-pair based association analysis is to identify significant PPI pairs where both genes are statistically associated with the disease and expose potential disease-susceptibility genes not identified in the individual SNP-based and/or individual gene-based association analysis. The STRING database has covered 60% of the protein-coding genes, and all PPI pairs included are demonstrated by at least one piece of experimental or computational evidence for human proteins [20]. Here, we assumed that genes 1 and 2 of a PPI pair have m_1 and m_2 SNPs with association P values, which can be partitioned into n_1 and n_2 LD blocks (which display weak LD), respectively. For each i th block ($i=1 \dots n_1 + n_2$), GATES is used to calculate the block-based P value (P_i) and to mark the key SNP from which P_i was derived. The scaled chi-square test is then employed to combine the $n_1 + n_2$ block-based P values into a single test statistic, accounting for LD between the $n_1 + n_2$ key SNPs:

$$X = -2 \sum_{i=1}^{n_1} \ln P_i - 2 \sum_{i=1}^{n_2} \ln P_i \quad (\text{Equation 2}).$$

When genes 1 and 2 are on different chromosomes or far apart on the same chromosome, the genomic blocks between these two genes can be assumed to be independent. The null hypothesis of HYST for the PPI-based analysis is that neither gene in a PPI pair is associated with the trait. The alternative hypothesis is that at least one gene involved in a PPI pair is associated with the disease. A heterogeneity test was applied using Cochran's Q [21] (i.e., $Q = \{\Phi^{-1}(1-P_A) - \Phi^{-1}(1-P_B)\}^2$, where Φ^{-1} denotes the inverse normal distribution function) and/or Higgins's I^2 [22], (i.e., $I^2 = (Q - 1)/Q$ if $Q \geq 1$, 0 otherwise) to exclude the PPI pairs whose significance is overwhelmingly dominated by only one gene ($I^2 > 0.5$).

Gene Enrichment Analysis

Across the results of gene-based analysis and gene-pair analysis, we collected the lists of BMD-associated genes at each of skeletal sites in each of the GEFOS studies, which were used as input information for subsequent gene enrichment analysis to assess their potential biological roles in the osteoporosis pathophysiology. The lists of the combined BMD-associated genes included not only the genome-wide significant genes identified by gene-based analysis but also the nominally significant genes from significant gene-pairs. The ClueGO plug-in [23] on the Cytoscape platform was used to retrieve and visualize the gene ontology terms and pathways enriched with disease-susceptibility genes in a biological network from the comprehensive GO database and pathway databases (KEGG and WikiPathways). Here, we chose 0.05 as the Q value threshold to determine whether the input

genes are enriched in GO terms or pathways than random chance after Benjamini-Hochberg multiple-testing FDR (false discovery rate) correction. The clustering functional networks created based on kappa statistics deciphered the similarity or correlation of biological activities in terms of their shared associated genes. A term-gene matrix containing the selected terms and their associated genes is created, then a term-term similarity matrix is calculated using kappa statistics to identify the association strength between the terms (kappa score). The detailed parameters of gene enrichment analysis were selected as follows: 0.6 for kappa score, 3 to 8 levels for GO tree interval, and at least 3 enriched genes or 5% enriched genes for terms and pathways. To further predict the interactive mechanisms of trait-associated genes and their regulatory roles in terms or pathways, PPI of involved genes were analyzed and visualized by CluePedia plug-in [24] on the Cytoscape platform based on the STRING database.

Transcriptomic Evaluation

The meta-analysis strategy combining multiple gene expression datasets may provide a robust, reproducible and accurate predictions for the association of widely expressed genes across tissues/cells relevant to the phenotypes of interest [25]. This meta-analysis calculated the combined *P* values of genes across 6 independent gene expression studies (Accession: GSE56815 [26], GSE7158 [27], GSE2208 [28], E-MEXP-1618 [29], GSE7429 [30], and GSE13850 (No publication)) using Fisher's method to assess the association of gene expressions with BMD. The combined *P* values were further corrected for multiple hypothesis testing using the Benjamini-Hochberg FDR. These microarray datasets were identified and downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) and ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) websites. These expression data met the following criteria: (1) case-control studies for discordant BMD (low or high BMD) in human subjects, (2) available raw data of gene expression, and (3) available phenotype data of subjects. Although the quantile or z-score for each study's inclusion criteria differs, the low BMD group and high BMD group in all six studies can be classified as high risk for osteopenia/osteoporosis or normal/healthy respectively, according to their discordant BMD values and the World Health Organization (WHO)'s definition [31]. Since the results derived from GWASs can originate from different cells and tissues in humans, it is reasonable to partially validate the association of the identified genes with BMD using the transcriptomics data from various cells/tissues related to bone metabolism. These gene expression data are derived from multiple bone-related cell sources (e.g., peripheral blood monocytes [32], osteocytes [33], and circulating B cells [34]) of females with discordant BMD. The basic characteristics of the subjects are shown in Supplementary Table 1. The methodology of the meta-analysis was detailed in our previous publication [35]. The differentially expressed genes (DEGs) with significant *Q* value ($q < 0.05$) were considered statistically significant across low vs. high BMD groups.

Correlation analysis of the *P* values of same variants across the two GEFOS studies

Since the GEFOS2 study does not provide the summary statistics of meta-analysis for FA-BMD, the same SNPs/genes between the GEFOS2 and GEFOS-seq studies were applied to between-study correlation analysis for FN-BMD and LS-BMD, respectively. We tested the Pearson correlation coefficients using R language to show the correlation of the *P* values of

the same SNPs/genes across the two GEFOS studies, and the correlation coefficients were estimated based on $-\log_{10}(P\text{ value})$ of these variants. The null hypothesis of the correlation test is that pairwise SNPs/genes in the two GEFOS studies are unrelated, and a significant P value ($p < 0.05$) of correlation analysis would reject the null hypothesis. Specifically, we first tested the overall correlation of the P values of the same variants by using genome-wide matched SNPs or genes across the two GEFOS studies. Second, we identified the significantly associated variants in each of the two GEFOS studies, then calculated the between-study correlation of their P values. The correlation analyses for associated variants were performed based on different significant thresholds to the P values of SNPs/genes, ranging from nominal significance ($P < 0.05$) to genome-wide significance ($P < 5 \times 10^{-8}$ for SNPs, $P < 2 \times 10^{-6}$ for genes), to investigate the between-study stability of P values in detail. Furthermore, we also estimated the correlation of the same variants with non-significance ($P > 0.05$) across the two GEFOS studies to represent a correlation baseline due to overlapping samples. Theoretically, the P values of the non-significant variants from two independent GWAS studies should be independently uniform and randomly distributed, so the correlation of non-significant variants between any two independent GWAS studies should approach 0. Using the nominal significance P value threshold ($P < 0.05$), the false negative findings of the associated variants were excluded from the non-significant variants as much as possible, so that the correlation of variants with non-significance might be attributed to overlapping samples. Taken together, due to lack of direct access to the original data of independent samples, we estimated the similarity of P values of the associated variants by empirically referring the correlation baseline originated from these overlapping samples. Additionally, we calculated the ratios of between-study replicated associated variants in the discovery of associated variants in each of the two GEFOS studies, to assess the between-study mutual replicability of significant associated variants when the samples used are not all the same.

Results

Gene-based Association Analysis

Compared with the SNP-based GWAS, the gene-based analysis enhanced the discovery of genome-wide significant genes for BMD ($P < 2 \times 10^{-6}$) in both GEFOS2 (Fig. 1A) and GEFOS-seq (Fig. 1B) studies. Notably, the number of common BMD-associated genes identified in both GEFOS studies was larger in our gene-based analysis than in the original SNP-based studies (Fig. 1C, Fig. 1D). Specifically, in the gene based analyses, a total of 50 and 48 associated genes were identified with FN-BMD and LS-BMD individually at genome-wide level in the GEFOS2 (Fig. 1E); 32, 43 and 3 candidate genes with genome-wide significant P value were identified for FN-BMD, LS-BMD and FA-BMD in the GEFOS-seq, respectively (Fig. 1F). In both of the GEFOS studies, the gene-based method identified the replication of some candidate genes for BMDs from the SNP-based GWAS, e.g., *WNT16* and *WLS* for FN-BMD and LS-BMD in GEFOS2, and *GALNT3* for FN-BMD, *TNFRSF11A* for LS-BMD, and *CPED1* for FA-BMD in the GEFOS-seq. Additionally, the gene-based method specifically identified prominent functional genes for BMD, including *HSAC5* and *FAM3C* for FN-BMD, and *ESR1* and *WNT16* for LS-BMD in both the two GEFOS studies. Supplementary Fig. 1 shown that the points in the QQ plots depart upward from the straight

line, indicating some observed P values (plotted on the vertical axis) are clearly more significant than expected under the null hypothesis (plotted on the vertical axis). Additionally, the genomic inflation factors (λ) for the BMDs of GEFOS-seq (Supplementary Fig. 1C, Fig. 1D, and Fig. 1E) are closer to 1 than those of GEFOS2 (Supplementary Fig. 1A and Fig. 1B), and there are earlier separations of the expected values from the observed in the GEFOS2 than those of GEFOS-seq. These results may indicate that, comparatively, many moderately significant P values are more significant than expected under the null hypothesis in the GEFOS2. Compared to thousands of true positives, this result is more reasonably explained by systematic differences in allele frequencies between subpopulations of enrolled subjects: a higher population stratification in the GEFOS2 (Asians and Caucasians) than that of GEFOS-seq (Caucasians only). Detailed information on these associated genes in GEFOS studies can be found in Supplementary Tables 2 and 3 corresponding to their association with specific skeletal sites of BMD.

Gene-pair Based Association Analysis

The gene-pair based association analysis identified some significant PPI pairs with genome-wide significant genes and/or nominally significant genes for the FN-BMD and LS-BMD in both GEFOS studies (Supplementary Fig. 2). Actually, a number of genes in the significant PPI pairs that did not reach genome-wide significance in the gene-based test have been previously reported as candidate BMD susceptibility genes (with at least one SNP having a $P < 5 \times 10^{-8}$ in the National Human Genome Research Institute (NHGRI) Catalog of Published GWAS Catalog [36] for BMD by October 23, 2016 (<http://www.ebi.ac.uk/gwas/>)). Between the two GEFOS studies, 7 same significant gene pairs were detected on FN-BMD, as were 10 same gene pairs for LS-BMD (Supplementary Table 4). Among these gene pairs, *SOST-SP7* and *AXIN1-LRP5* were simultaneously significantly associated with FN-BMD and LS-BMD, whose genes were already reported to play important roles in regulating BMDs on more than one skeletal sites *in silico* and/or *in vivo*, and in clinical studies [32, 37–39]. The significant PPI pairs and involved genes from GEFOS2 and GEFOS-seq studies are detailed in Supplementary Tables 5 and 6, respectively.

Gene Enrichment Analysis

We used the combined lists of BMD-associated genes from the gene-based analysis and gene-pair analysis for each of the skeletal sites to identify GO terms and pathways related to FN-BMD and LS-BMD of each of the two GEFOS studies, respectively. In the two GEFOS studies, several known bone-related functional terms simultaneously achieved significant enrichments for BMD after Benjamini-Hochberg multiple testing correction ($Q < 0.05$), e.g., the Wnt signaling pathway for FN-BMD, and RANKL/RANK (Receptor activator of NF- κ B (ligand)) signaling pathway and regulation of Wnt/B-catenin signaling by small molecule compounds for LS-BMD (Table 1). Additionally, the gene enrichment analysis also identified some potential pathways that may provide new insights into the pathological mechanisms of osteoporosis (Fig. 2A and 2B), including the axis specification, basal cell carcinoma, and breast cancer. Additional enriched GO terms and pathways are detailed in Supplementary Tables 7 and 8.

Differential Gene Expression Analysis

In the analysis, 7 BMD-associated genes identified by the gene-based method were significantly differentially expressed between case-control subjects with discordant FN-BMD after multiple testing correction by FDR ($Q < 0.05$), and 6 identified associated genes were supported for their association with LS-BMD (Table 2). In previous studies, 6 of these associated genes with FN-BMD or LS-BMD have been identified as candidate genes for BMD variations by *in vitro* and/or *in vivo* studies, including (*RPS6KA5*, *MEPE*, *SHFM1*, *FAM3C*, *HDAC5*, and *ESR1*). More importantly, 3 novel associated genes were differentially expressed between the case-control samples with significant Q values: *UBTF* ($Q = 2.09E-02$), *AAAS* ($Q = 4.05E-02$), and *C11orf58* ($Q = 4.19E-03$). Thus, we focused on elaborating (in Discussion section) on the functional roles of the three novel candidate genes in their potential pathophysiological mechanism in bone metabolism.

Correlation analysis of the P values of same variants across the two GEFOS studies

Between the two GEFOS studies, a total of ~2.3 million same SNPs and ~25,000 same genes were matched for FN-BMD and LS-BMD, respectively. Here, we denote C_{OG}/C_{OS} as the overall correlation of P values of same genes/SNPs, C_{SG}/C_{SS} as the correlation of P values of significant associated genes/SNPs, and C_{NG}/C_{NS} as the correlation of P values of non-significant associated genes/SNPs. The r value shows the correlation coefficient of same variants across the two GEFOS studies. In the correlation analyses for FN-BMD (Fig.3A), C_{OG} ($r = 0.48$) appeared to be higher than C_{OS} ($r = 0.38$), and the average C_{SG} ($r_{GEFOS2} = 0.82$, $r_{GEFOS-seq} = 0.83$) appeared to be higher compared to the average C_{SS} ($r_{GEFOS2} = 0.78$, $r_{GEFOS-seq} = 0.83$) across different significance thresholds. The r_{GEFOS2} and $r_{AEFOS-seq}$ indicate the correlation of matched variants with significant P values in GEFOS2 and GEFOS-seq, respectively. Conversely, C_{NG} ($r = 0.09$) was lower than C_{NS} ($r = 0.11$). Therefore, using the C_{NG} or C_{NS} as a reference to exclude the between-study shared-sample effects, the correlation for either overall or significant genes was greater than that of SNPs. We also performed the correlation analysis for the same variants on LS-BMD using the same strategy (Fig.3B), and the correlation coefficients of the P values of the same variants were close to those of FN-BMD (C_{OG} ($r = 0.50$), C_{OS} ($r = 0.39$), average C_{SG} ($r_{GEFOS2} = 0.82$, $r_{GEFOS-seq} = 0.78$), average C_{SS} ($r_{GEFOS2} = 0.72$, $r_{GEFOS-seq} = 0.68$), C_{NG} ($r = 0.09$), C_{NS} ($r = 0.11$)). Furthermore, the significantly associated variants in the GEFOS2 and GEFOS-seq were respectively identified with high between-study mutual replicabilities across almost all significant thresholds (e.g., for FN-BMD, $P < 2E-06$, $Ratio_{GEFOS2} = 50\%$, $Ratio_{GEFOS-seq} = 77\%$) (Table 3). Together, we observed that, with largely overlapping samples across the two GEFOS studies, the significant associated variants identified in each of the two GEFOS studies were moderately consistently detected in the other study with correlated P values for their association with BMDs.

Discussion

As we expected, the gene-based analysis not only successfully replicated a series of known candidate loci from the original SNP-based GWASs, but also identified additional BMD-associated genes corresponding to multiple skeletal sites. It is worth noting that, between the two GEFOS studies, the number of overlapping candidate genes identified by the gene-based

tests exceeded those of the SNP-based GWASs, indicating that the gene-based method may provide higher testing stability and robustness for the identification of genetic variants associated with the phenotypes. The gene-pair analysis reasonably extended the range of potential disease-susceptibility genes not identified in the SNP-based and/or gene-based association analysis. Subsequently, leveraging the extended associated gene lists, the gene enrichment analysis not only identified some well-known pathophysiological mechanisms for BMD regulation, but also identified some unexpected associations of BMD-associated genes with skeletal development or other diseases/traits.

Notably, three BMD-associated genes (*UBTF*, *AAAS* and *C11orf58*) with partial validation from the transcriptomic meta-analysis have not yet been discovered or mentioned as candidate genes in previous publications. However, a literature search offered some exciting clues potentially underlying their functional association with bone metabolism. *Upstream binding transcription factor (UBTF)* functions in ribosomal RNA (rRNA) polymerase (Pol I and II)-specific transcription of the ribosomal genes to maintain genome stability, and the inhibition of *UBTF* leads to DNA damage and genomic instability independent of Pol I transcription [40]. Bone cell proliferation and differentiation could be regulated by suppressed rRNA gene transcription due to *UBTF* acetylation modulated by RUNX2–HDAC1 interaction [41]. The mutation of *Aladin WD repeat nucleoporin (AAAS)* gene causes Triple A syndrome (alacrima, achalasia, adrenal failure, progressive neurodegenerative disease), which leads to osteoporosis due to reduced generation of testosterone and estrogens by adrenal androgens, less physical activity, and decreased exposure to the sun [42]. Meanwhile, our study found that *AAAS*, *nucleoporin 155 (NUP155)*, and *Cvclin B1 (CCNB1)* were significantly enriched in the GO term of mitotic nuclear envelope disassembly and directly interacted with the hub genes of the biological network (Fig.2C). The nuclear envelope has a fundamental impact on cellular organization and complexity, and mutations in genes encoding nuclear envelope proteins may induce skeletal dysplasia and osteoporosis [43]. In addition, the SNP rs6486303 within *C11orf58* was reported as an important SNP associated with wrist BMD ($P=3.34E-05$) in 1000 unrelated Caucasian subjects [44], but it is not significantly associated with BMDs (FN- and LS-BMD) in GEFOS2 or with BMDs (FN-, LS-, and FA-BMD) in GEFOS-seq. To date, the functional role of *C11orf58* in bone metabolism is not yet well understood.

Compared to the SNP-based GWAS, the gene-based GWAS analysis identified a higher overall correlation of P values of matched variants and an increased correlation of P values of significant associated variants, efficiently improving the stability and consistency for testing the association between the genotype and phenotype. This improvement may be due to the natural advantages of gene-based strategy, e.g., overcoming the potential heterogeneities from alleles and/or cohorts and reducing the burden of multiple testing for traditional GWASs. Additionally, our previous simulation investigation indicated that overlapping samples between two meta-analyses can lead to an increase of replicability of significantly associated variants compared to that between two independent meta-analyses [6]. In the present study, the significant associated variants were identified with higher replicabilities than the estimation from our previous simulation study, which may be partially due to the more overlapping samples between the two GEFOS studies than that investigated in our earlier study. While, by referring the correlation of the false positive

association arising from overlapping samples, the results that a moderately high correlation of association of genes/SNPs with BMDs indicated that, the identified potential BMD-associated genes/SNPs have high between-study stability and consistency between the GEFOS2 and GEFOS-seq studies. Taken together, the results of correlation analyses supported that, in the two GEFOS studies, the gene-based analysis can identify additional potential causal variants with the higher reliability of discoveries than the prevailing SNP-based analysis.

The purpose of our study is to identify additional candidate genes involved in the mechanism of bone metabolism by integrating multiple-level biological evidences. The major contribution of the study is to provide clues (unbiased hypothesis generating) to select candidate genes for bone diseases for downstream *in vivo* and/or *in vitro* studies. Comparatively, it's not of our primary study scope to generate knock-out mice with functional evidence and phenotypic consequences in regulating bone metabolism. Actually, some of the potential casual genes, which have been used as the study knock-out targets in mice models to reveal how disruption of their functions could influence the balance of bone metabolism, were verified for their significance in humans as BMD-associated genes in our gene-based analysis.

Conclusion

Traditional GWAS studies provided a relative low efficiency for discovering candidate variants of BMD, and these identified variants explain a small proportion of osteoporosis heritability. As a supplementary strategy to SNP-based GWAS, the gene-based analysis is shown that it helped identify some novel BMD-associated genes. In addition to its empirically increased statistical power, gene-based analysis also provides a higher testing stability for identification of BMD genes. The present study represents a systemic endeavor integrating the multilevel biological evidence and bioinformatics analyses to reveal BMD-associated genes and their roles in skeletal metabolism, which may shed light on the genetic and pathophysiological mechanisms of osteoporosis. Considering the common situation that majority of public resources only provided data as summary statistics, this study is a practical example to extract more genetic information without heavy burden of computation and time costs with the raw genotype information, which could be easily used to investigate summary statistics of other phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This study was partially supported by and/or benefited from grants from National Institutes of Health [AR069055, U19 AG055373, R01 MH104680, R01AR059781 and P20GM109036], and Edward G. Schlieder Endowment to Tulane University.

Abbreviations

SNP Single-nucleotide polymorphism

GWASs	genome-wide association studies
BMD	bone mineral density
GEFOS	Genetic Factors for Osteoporosis studies
LD	linkage disequilibrium
GO	gene ontology
PPI	protein-protein interaction
AOGC	Australasian Osteoporosis Genetics Consortium
FHS	Framingham Heart Study
RS-I	Rotterdam Study-I
ERF	Erasmus Rucphen Family Study
TWINSUK	TwinsUK
FN	femoral neck
LS	lumbar spine
FA	forearm
GATES	Extended Simes procedure method
KGG	Knowledge-based mining system for Genome-wide Genetic studies
QQ	Quantile-Quantile
HYST	hybrid set-based test
FDR	false discovery rate
GEO	Gene Expression Omnibus
WHO	World Health Organization
DEGs	differentially expressed genes
NHGRI	National Human Genome Research Institute
RANKL	Receptor activator of NFkB ligand
RANK	Receptor activator of NFkB
UBTF	Upstream binding transcription factor
rRNA	ribosomal RNA
AAAS	Aladin WD repeat nucleoporin
NUP155	nucleoporin 155

CCNB1

Cyclin B1

Reference

1. Kanis JA, et al., Assessment of fracture risk. *Osteoporos Int*, 2005 16(6): p. 581–9. [PubMed: 15616758]
2. Richards JB, Zheng HF, and Spector TD, Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Genet*, 2012 13(8): p. 576–88. [PubMed: 22805710]
3. Liu YJ, et al., Genome-wide Association Studies for Osteoporosis: A 2013 Update. *J Bone Metab*, 2014 21(2): p. 99–116. [PubMed: 25006567]
4. Deng HW, et al., Relevance of the genes for bone mass variation to susceptibility to osteoporotic fractures and its implications to gene search for complex human diseases. *Genet Epidemiol*, 2002 22(1): p. 12–25. [PubMed: 11754470]
5. Evangelou E and Ioannidis JP, Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet*, 2013 14(6): p. 379–89. [PubMed: 23657481]
6. Pei YF, et al., On individual genome-wide association studies and their meta-analysis. *Hum Genet*, 2014 133(3): p. 265–79. [PubMed: 24114349]
7. Ioannidis JP, et al., Replication validity of genetic association studies. *Nat Genet*, 2001 29(3): p. 306–9. [PubMed: 11600885]
8. McCarthy MI, et al., Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 2008 9(5): p. 356–69. [PubMed: 18398418]
9. Rivadeneira F, et al., Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet*, 2009 41(11): p. 1199–206. [PubMed: 19801982]
10. Estrada K, 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(5): p. 491–501. [PubMed: 22504420]
11. Zheng HF, et al., Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*, 2015 526(7571): p. 112–7. [PubMed: 26367794]
12. Tak YG and Farnham PJ, Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*, 2015 8: p. 57. [PubMed: 26719772]
13. Hazelett DJ, et al., Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet*, 2014 10(1): p. e1004102. [PubMed: 24497837]
14. Spain SL and Barrett JC, Strategies for fine-mapping complex traits. *Hum Mol Genet*, 2015 24(R1): p. R111–9. [PubMed: 26157023]
15. Hagg S, et al., Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. *Hum Mol Genet*, 2015 24(23): p. 6849–60. [PubMed: 26376864]
16. Wood AR, et al., Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Hum Mol Genet*, 2011 20(20): p. 4082–92. [PubMed: 21798870]
17. Li MX, et al., GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet*, 2011 88(3): p. 283–93. [PubMed: 21397060]
18. Wilk MB and Gnanadesikan R, Probability plotting methods for the analysis of data. *Biometrika*, 1968 55(1): p. 1–17. [PubMed: 5661047]
19. Li MX, Kwan JS, and Sham PC, HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am J Hum Genet*, 2012 91(3): p. 478–88. [PubMed: 22958900]
20. Szklarczyk D, et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 2015 43(Database issue): p. D447–52. [PubMed: 25352553]
21. Cochran WG, The Combination of Estimates from Different Experiments. *Biometrics*, 1954 10(1): p. 101–129.
22. Higgins JP, et al., Measuring inconsistency in meta-analyses. *BMJ*, 2003 327(7414): p. 557–60. [PubMed: 12958120]

23. Bindea G, et al., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 2009 25(8): p. 1091–3. [PubMed: 19237447]
24. Bindea G, Galon J, and Mlecnik B, CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*, 2013 29(5): p. 661–3. [PubMed: 23325622]
25. Ramasamy A, et al., Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med*, 2008 5(9): p. e184. [PubMed: 18767902]
26. Chen XD, 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(2): p. 339–55. [PubMed: 19594299]
27. Lei SF, et al., An in vivo genome wide gene expression study of circulating monocytes suggested GBP1, STAT1 and CXCL10 as novel risk genes for the differentiation of peak bone mass. *Bone*, 2009 44(5): p. 1010–4. [PubMed: 19223260]
28. Liu YZ, et al., A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes. *J Biol Chem*, 2005 280(32): p. 29011–6. [PubMed: 15965235]
29. Reppe S, et al., Eight genes are highly associated with BMD variation in postmenopausal Caucasian women. *Bone*, 2010 46(3): p. 604–12. [PubMed: 19922823]
30. Xiao P, et al., In vivo genome-wide expression study on human circulating B cells suggests a novel ESR1 and MAPK3 network for postmenopausal osteoporosis. *J Bone Miner Res*, 2008 23(5): p. 644–54. [PubMed: 18433299]
31. Prevention and management of osteoporosis. *World Health Organ Tech Rep Ser*, 2003 921: p. 1–164, back cover.
32. Zhou PR, et al., SOST polymorphisms and response to alendronate treatment in postmenopausal Chinese women with osteoporosis. *Pharmacogenomics*, 2015 16(10): p. 1077–88. [PubMed: 26250343]
33. Dallas SL, Prideaux M, and Bonewald LF, The osteocyte: an endocrine cell... and more. *Endocr Rev*, 2013 34(5): p. 658–90. [PubMed: 23612223]
34. Horowitz MC, Fretz JA, and Lorenzo JA, How B cells influence bone biology in health and disease. *Bone*, 2010 47(3): p. 472–9. [PubMed: 20601290]
35. He H, et al., Network-Based Meta-Analyses of Associations of Multiple Gene Expression Profiles with Bone Mineral Density Variations in Women. *PLoS One*, 2016 11(1): p. e0147475. [PubMed: 26808152]
36. Welter D, et al., The NHGRIG WAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*, 2014 42(Database issue): p. D1001–6. [PubMed: 24316577]
37. Koller DL, et al., Contribution of the LRP5 gene to normal variation in peak BMD in women. *J Bone Miner Res*, 2005 20(1): p. 75–80. [PubMed: 15619672]
38. Holmen SL, et al., Decreased BMD and limb deformities in mice carrying mutations in both *Lrp5* and *Lrp6*. *J Bone Miner Res*, 2004 19(12): p. 2033–40. [PubMed: 15537447]
39. Styrkarsdottir U, et al., Sequence variants in the *PTCH1* gene associate with spine bone mineral density and osteoporotic fractures. *Nat Commun*, 2016 7: p. 10129. [PubMed: 26733130]
40. Sanij E, et al., A novel role for the Pol I transcription factor UBTF in maintaining genome stability through the regulation of highly transcribed Pol II genes. *Genome Res*, 2015 25(2): p. 201–12. [PubMed: 25452314]
41. Ali SA, et al., A RUNX2-HDAC1 co-repressor complex regulates rRNA gene expression by modulating UBF acetylation. *J Cell Sci*, 2012 125(Pt 11): p. 2732–9. [PubMed: 22393235]
42. Dumic M, et al., Low bone mineral density for age/osteoporosis in triple A syndrome-an overlooked symptom of unexplained etiology. *Osteoporos Int*, 2016 27(2): p. 521–6. [PubMed: 26243364]
43. Dauer WT and Worman HJ, The nuclear envelope as a signaling node in development and disease. *Dev Cell*, 2009 17(5): p. 626–38. [PubMed: 19922868]
44. Tan L, et al., A genome-wide association analysis implicates *SOX6* as a candidate gene for wrist bone mass. *Sci China Life Sci*, 2010 53(9): p. 1065–72. [PubMed: 21104366]

Fig. 1: Distribution of BMD-associated Genes

(A). In the GEFOS2 study, the gene-based analysis replicated 20 BMD-related genes reaching genome-wide significance level. (B). Compared with original studies, the gene-based analysis replicated 17 BMD-related genes reaching genome-wide significance level from GEFOS-seq study. (C). In the present gene-based analysis, 24 and 15 genes were uniquely identified in GEFOS2 and GEFOS-seq studies, and detection of 47 genes was overlapped. (D). In the original paper for SNP-based analysis, 49 and 9 genes were exclusively detected in GEFOS2 and GEFOS-seq studies respectively, and 13 genes were common. (E). In the gene-based analysis for GEFOS2 study, 23 genes and 21 genes were site-specifically identified in FN-BMD and LS-BMD, and 27 genes were confirmed in both

BMD traits. (F). In the gene-based analysis for GEFOS-seq study, 3 genes were associated with FA-BMD, 32 genes with FN-BMD and 43 genes with LS-BMD, respectively. Two common BMD-related genes were identified for FA-and FN-BMD, 2 genes for FA-and LS-BMD and 14 genes for FN-and LS-BMD.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

identified in this discovery study were marked in red (*AAAS*, *NUP155* and *CCNBI*), and they directly interacted with hub genes (*NEK6*, *NEK9*, and *PLKI*) of the network.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig.3: Correlation Analysis of the P Values of Same Variants

In the correlation analyses for FN-BMD (A) and LS-BMD (B), compared to the SNP-based GWAS, the gene-based GWAS analysis identified a higher overall correlation of P values of matched variants and an increased correlation of P values of significant associated variants across different significance thresholds.

Gene Enrichment Analysis

Table 1.

Phenotype	ID	GO Term	Ontology Source	GEFOS2		GEFOS-seq		Associated Genes
				P value	Q value	P value	Q value	
FN-BMD	GO:0000428	Wnt Signaling Pathway	WikiPathways	1.00E-03	1.10E-03	7.10E-09	1.20E-07	AXIN1, LRP5, WNT16, CCND1 ² , CTNNB1 ² , MYC ² , WNT5A ²
	GO:0003050	Proteasome	KEGG	1.30E-05	1.60E-05	4.80E-04	5.70E-04	PSMA5 ¹ , PSMD12, PSMD13, SHEM1
	GO:0002018	RANKL/RANK (Receptor activator of NFKB (ligand) Signaling Pathway	WikiPathways	3.30E-05	3.30E-05	9.40E-04	9.40E-04	RELA ¹ , TNFRSF11A, TNFRSF11B, TRAF1 ¹ , TNFSF11
	GO:0004310	Wnt signaling pathway	KEGG	2.30E-08	1.00E-07	1.60E-09	3.20E-08	AXIN1, LRP5, SFRP4, SOST, WNT1, WNT16, WNT3, WNT4 ¹ , MMP2 ² , SMAD3 ²
	GO:0005217	Basal cell carcinoma	KEGG	2.60E-08	8.70E-08	4.00E-05	6.20E-05	AXIN1, PTCH1 ¹ , WNT1, WNT16, WNT3, WNT4 ¹
LS-BMD	GO:0005224	Breast cancer	KEGG	2.70E-08	7.10E-08	1.90E-09	1.90E-08	AXIN1, ESRI, JAG1, LRP5, WNT1, WNT16, WNT4 ¹ , WNT3, SPI ² , TNFSF11 ²
	GO:0003664	Regulation of Wnt/B-catenin Signaling by Small Molecule Compounds	WikiPathways	2.70E-05	2.90E-05	3.20E-05	5.30E-05	AXIN1, SFRP4, WNT1
	GO:0000399	Wnt Signaling Pathway and Pluripotency	WikiPathways	1.40E-06	2.10E-06	1.90E-06	6.50E-06	AXIN1, LRP5, WNT1, MMP2 ² , WNT16, WNT3, WNT4 ¹
	GO:0000428	Wnt Signaling Pathway	WikiPathways	3.50E-09	2.30E-08	1.70E-07	1.10E-06	AXTN1, LRP5, SFRP4, WNT1, WNT16, WNT3, WNT4 ¹
	GO:0009798	axis specification	GO_BiologicalProcess	2.90E-08	6.40E-08	1.00E-06	5.20E-06	AXTN1, LRP5, PTCH1 ¹ , SOST, WLS, WNT1, WNT3
	GO:0009950	dorsal/ventral axis specification	GO_BiologicalProcess	9.30E-07	1.70E-06	1.10E-06	4.50E-06	AXIN1, LRP5, SOST, WNT3

The Ontology Source includes Biological Process, KEGG pathway, and Wiki pathway. The *P* value indicates the raw *P* values of genes, and *Q* value means the *P* values adjusted with Benjamini-Hochberg multiple test correction. For the associated genes, lack of subscript indicates the common genes identified in both GEFOS studies; the genes with subscripts of 1 and 2 indicate they were specifically identified in GEFOS2 and GEFOS-seq studies, respectively.

Table 2.

Candidate Genes with Gene Expression Validation

Phenotype	Categories	Symbol	GEFOS2	GEFOS-seq	Meta6 mRNA	
			<i>P</i> value ¹	<i>P</i> value ¹	<i>P</i> value ²	<i>Q</i> value
FN-BMD	GEFOS2	<i>ESR1</i>	2.26E-12	1.70E-04	1.48E-03	1.93E-02
		<i>MEPE</i>	3.14E-08	1.80E-04	2.70E-03	2.60E-02
		<i>SHFM1</i>	1.09E-07	4.20E-06	3.06E-03	2.80E-02
		<i>RPS6KA5</i>	7.08E-07	8.60E-04	3.45E-03	2.98E-02
	Common	<i>HDAC5</i>	1.58E-08	1.18E-08	3.68E-03	3.11E-02
		<i>UBTF</i>	1.21E-09	1.33E-07	1.73E-03	2.09E-02
	GEFOS-seq	<i>C11orf58</i>	3.40E-03	1.18E-06	5.78E-05	4.19E-03
	LS-BMD	GEFOS2	<i>MEPE</i>	5.37E-09	8.15E-06	2.70E-03
<i>SHFM1</i>			3.75E-07	1.58E-05	3.06E-03	2.80E-02
<i>ESR1</i>			2.08E-12	1.43E-07	1.48E-03	1.93E-02
Common		<i>AAAS</i>	1.81E-10	2.83E-07	5.94E-03	4.05E-02
		<i>RPS6KA5</i>	5.05E-09	8.42E-07	3.45E-03	2.98E-02
GEFOS-seq		<i>C11orf58</i>	1.70E-03	4.96E-07	5.78E-05	4.19E-03

The *P* value¹ indicates the *P* values of genes for BMD in GEFOS projects, and *P* value² shows the *P* values of genes identified by differential gene expression study. The *Q* values are the *P* values that were adjusted using FDR to correct for multiple hypothesis testing. The *P* value¹ with genome-wide significance (*P* value¹ < 2E-06), and the *P* value² and *Q* value with empirical significance (*P* < 0.05) are shown in bold.

Table 3.

Correlation Analysis for Variants

Variants	Phenotype	Threshold	Both Significant SNPs			GEFOS2*			GEFOS2-seq*		
			number	r	P value	number	ratio	r	P value	number	ratio
SNPs	FN-BMD	5.00E-02	29700	0.66	< 2.2E-16	121834	24%	0.64	< 2.2E-16	135726	22%
		1.00E-02	6972	0.76	< 2.2E-16	30768	23%	0.78	< 2.2E-16	33180	21%
		1.00E-03	2162	0.82	< 2.2E-16	6967	31%	0.84	< 2.2E-16	6500	33%
		1.00E-04	1192	0.83	< 2.2E-16	2892	41%	0.85	< 2.2E-16	2267	53%
		1.00E-05	714	0.81	< 2.2E-16	1477	48%	0.88	< 2.2E-16	1061	67%
		1.00E-06	500	0.81	< 2.2E-16	841	59%	0.86	< 2.2E-16	699	72%
		1.00E-07	359	0.79	< 2.2E-16	552	65%	0.87	< 2.2E-16	498	72%
		5.00E-08	326	0.79	< 2.2E-16	493	66%	0.87	< 2.2E-16	471	69%
		5.00E-02	30842	0.67	< 2.2E-16	121784	25%	0.66	< 2.2E-16	133642	23%
		1.00E-02	7680	0.78	< 2.2E-16	29808	26%	0.74	< 2.2E-16	32806	23%
Genes	FN-BMD	1.00E-03	2730	0.82	< 2.2E-16	6701	41%	0.73	< 2.2E-16	6225	44%
		1.00E-04	1195	0.79	< 2.2E-16	3024	40%	0.75	< 2.2E-16	1897	63%
		1.00E-05	810	0.78	< 2.2E-16	1927	42%	0.7	< 2.2E-16	1114	73%
		1.00E-06	596	0.78	< 2.2E-16	1516	39%	0.63	< 2.2E-16	713	84%
		1.00E-07	440	0.59	< 2.2E-16	671	66%	0.62	< 2.2E-16	518	85%
		5.00E-08	417	0.57	< 2.2E-16	630	66%	0.61	< 2.2E-16	486	86%
		5.00E-02	378	0.78	< 2.2E-16	1914	20%	0.78	< 2.2E-16	1611	23%
		1.00E-02	147	0.83	< 2.2E-16	637	23%	0.83	< 2.2E-16	504	29%
		1.00E-03	81	0.83	< 2.2E-16	204	40%	0.82	< 2.2E-16	138	59%
		1.00E-04	49	0.84	< 2.2E-16	100	49%	0.84	< 2.2E-16	69	71%
Genes	FN-BMD	1.00E-05	32	0.84	< 2.2E-16	60	53%	0.84	6.63E-13	44	73%
		2.00E-06	23	0.85	8.88E-14	46	50%	0.83	1.04E-08	30	77%
		5.00E-02	446	0.79	< 2.2E-16	1797	25%	0.75	< 2.2E-16	1731	26%
		1.00E-02	166	0.84	< 2.2E-16	609	27%	0.79	< 2.2E-16	498	33%
		1.00E-03	78	0.87	< 2.2E-16	207	38%	0.8	< 2.2E-16	161	48%
		1.00E-04	46	0.84	< 2.2E-16	108	43%	0.83	< 2.2E-16	72	64%
		1.00E-05	37	0.8	5.28E-14	59	63%	0.79	1.85E-11	49	76%
	LS-BMD										

Variables	Phenotype	Threshold	Both Significant SNPs			GEFOS2*			GEFOS-seq*		
			number	r	P value	number	r	P value	number	r	ratio
		2.00E-06	31	0.76	5.42E-09	42	0.74	1.99E-07	37		84%

“Threshold” indicates the significant *P* value thresholds for the association of variants with BMD. “r” shows the correlation of same variants across the two GEFOS studies, and “*P* value” indicates the *P* values of the correlation tests. “Number” indicates the number of matched variants between the two GEFOS studies, and “ratio” denotes the ratios of variants with significance in both GEFOS studies in the significant variants from one of the two GEFOS studies. “GEFOS2*” indicates the group of matched variants with significant *P* values in GEFOS2, and “GEFOS-seq*” indicates the group of matched variants with significant *P* values in GEFOS-seq.