BJR

■ **RESEARCH**

Integrative analysis of GWAS, eQTLs and meQTLs data suggests that multiple gene sets are associated with bone mineral density

W. Wang,*
S. Huang,*
W. Hou,*
Y. Liu,
Q. Fan,
A. He,
Y. Wen,
J. Hao,
X. Guo,
F. Zhang

Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, Shaanxi, China

■ W. Wang,* MS, Student, School of Public Health,
■ Y. Liu, MS, Student, Department of Breast Surgery, First Affiliated Hospital, Health Science Center, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
■ A. He, BS, School of Public Health,
■ Y. Wen, PhD, Postdoctoral Researcher,
■ J. Hao, MS, Research Assistant, School of Public Health,
■ X. Guo, MD, Professor, School of Public Health,
■ F. Zhang, PhD, Professor, School of Public Health, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Health Science Center, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
■ W. Hou,* MD, Attending Doctor, Department of Joint Surgery, Osteonecrosis and Joint Reconstruction Ward, HongHui Hospital, Health Science Center, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
■ S. Huang,* MS, Attending Doctor, Department of Radiotherapy, Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
■ Q. Fan, BS, Student, School of Public Health, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Health Science Center, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

Correspondence should be sent to F. Zhang; email: fzhxjtu@mail.xjtu.edu.cn

doi: 10.1302/2046-3758.610. BJR-2017-0113.R1

Bone Joint Res 2017;6:572–576.
Received: 16 April 2017; Accepted: 5 July 2017

Objectives

Several genome-wide association studies (GWAS) of bone mineral density (BMD) have successfully identified multiple susceptibility genes, yet isolated susceptibility genes are often difficult to interpret biologically. The aim of this study was to unravel the genetic background of BMD at pathway level, by integrating BMD GWAS data with genome-wide expression quantitative trait loci (eQTLs) and methylation quantitative trait loci (meQTLs) data

Method

We employed the GWAS datasets of BMD from the Genetic Factors for Osteoporosis Consortium (GEFOS), analysing patients' BMD. The areas studied included 32 735 femoral necks, 28 498 lumbar spines, and 8143 forearms. Genome-wide eQTLs (containing 923 021 eQTLs) and meQTLs (containing 683 152 unique methylation sites with local meQTLs) data sets were collected from recently published studies. Gene scores were first calculated by summary data-based Mendelian randomisation (SMR) software and meQTL-aligned GWAS results. Gene set enrichment analysis (GSEA) was then applied to identify BMD-associated gene sets with a predefined significance level of 0.05.

Results

We identified multiple gene sets associated with BMD in one or more regions, including relevant known biological gene sets such as the Reactome Circadian Clock (GSEA p-value = 1.0×10^{-4} for LS and 2.7×10^{-2} for femoral necks BMD in eQTLs-based GSEA) and insulin-like growth factor receptor binding (GSEA p-value = 5.0×10^{-4} for femoral necks and 2.6×10^{-2} for lumbar spines BMD in meQTLs-based GSEA).

Conclusion

Our results provided novel clues for subsequent functional analysis of bone metabolism, and illustrated the benefit of integrating eQTLs and meQTLs data into pathway association analysis for genetic studies of complex human diseases.

Cite this article: *Bone Joint Res* 2017;6:572–576.

Article focus

■ The genetic predisposition of osteoporosis is poorly understood. Integrating the information from the genome-wide association studies (GWAS) of bone mineral density (BMD) with expression of quantitative trait loci (eQTLs) and methylation quantitative trait loci (meQTLs) data, however, may provide a novel insight into the mechanism of osteoporosis.

Key messages

■ Integration of GWAS and eQTLs (meQTLs) data identified multiple pathways associated with BMD.
■ Many pathways were associated with BMD in more than one region, including the Reactome Circadian Clock insulin-like growth factor receptor binding.

Strengths and limitations

- Strength: Our new analysis framework can provide more biological interpretable results by integrating eQTL and meQTL data.
- Strength: The method can be applied to widely existed, publicly available GWAS summary data.
- Limitation: Further eQTL and meQTL data are warranted to provide more tissue-specific exploration of GWAS results.

Introduction

Bone mineral density (BMD) is the most commonly used indicator for assessing the risk of a fracture from osteoporosis. It has been estimated that genetic factors contribute to over half of the risk of low BMD, suggesting that BMD is a highly inherited phenotype.[1,2] Genome-wide association studies (GWAS) are capable of simultaneously assessing the correlation between target phenotypes and millions of genetic loci. GWAS of osteoporosis and BMD have successfully identified multiple susceptibility genes.[1,2] However, isolated susceptibility genes are often difficult to interpret biologically and have largely been neglected.

GWAS have limited power to detect the causal loci with moderate or weak genetic effects, due to their genome-wide threshold of strict statistical significance.[3] Since individual genes can participate in multiple cellular processes, identifying several disease-associated genes is insufficient when used for revealing the pathogenesis of complex human diseases. A better solution, therefore, is to test the associations between target diseases and multiple functionally related loci, or biological pathways, simultaneously. Inspired by the gene set enrichment analysis (GSEA) of microarray data, GWAS-based pathway association analysis has been proposed and applied in several studies of GWAS.[3-6] By integrating GWAS results with known gene set annotation databases, GWAS-based pathway association analysis has been shown to perform better in clarifying GWAS results.[7] However, traditional pathway association analysis does not usually take into account the genetic effect of eQTLs and meQTLs.

A large proportion of genetic variants affect the phenotypes by causal regulatory effect rather than by directly influencing the protein structure.[8] For instance, it has been demonstrated that eQTLs and meQTLs play important roles in the pathogenesis of complex human diseases.[8,9] A genome-wide meQTLs dataset has been used for next-generation sequencing to assay blood DNA methylation at approximately 4.5 million loci, and for testing their associations with about 4.5 million single-nucleotide polymorphisms (SNPs).[9] Recently, Zhu et al[8] proposed using summary data-based Mendelian randomisation (SMR) analysis to study this problem. SMR can integrate GWAS and genome-wide eQTLs datasets to identify causal genes, the expression of which is associated with target diseases.[8] SMR was applied to five real GWAS datasets and identified multiple novel susceptibility genes for complex human diseases, demonstrating the power for integrating eQTLs datasets with GWAS in genetic analysis. Integrating the genome-wide eQTLs and meQTLs dataset in GWAS analysis could therefore provide novel clues for clarifying the genetic mechanism of BMD. However, Zhu et al[8] mainly focused on single gene mapping, while GWAS-based pathway association analysis, evaluating eQTLs and meQTLs, has more potential to discover BMD associated gene sets.

In this study, we integrated the Genetic Factors for Osteoporosis Consortium (GEFOS) GWAS of BMD with the genome-wide eQTLs and meQTLs data sets, in order to scan for possible BMD-associated gene sets. We first computed the gene score by SMR software and meQTLs, which were aligned with the GWAS results. The GSEA algorithm was then applied to the gene score to scan for BMD-associated gene sets. This novel method involved identifying significant pathways with the aid of prior data sets so that the results reflected real biological circumstances.

Materials and Methods

Gene score calculation. GWAS of BMD were integrated with eQTLs and meQTLs data before conducting pathway association analysis. A detailed description of the GWAS, eQTLs, and meQTLs data is presented as supplementary information. To integrate eQTLs with GWAS, SMR (an eQTLs-based GWAS analytical approach) was adopted to perform single-gene expression association analysis for BMD.[8] SMR used disease related GWAS and eQTLs data to evaluate the effect of gene expression on phenotypes. The effect size of the most significant SNP in the SMR test was denoted as the gene score in eQTL-based gene score calculation.

For meQTLs-based gene score calculation, meQTLs were eliminated without gene annotation from a total of 386 767 meQTLs. The BMD GWAS SNPs were then aligned with the significant meQTLs (meQTLs $p < 4.05 \times 10^{-5}$, corresponding to false discovery rate (FDR) < 0.01), in order to identify those that overlapped. The largest GWAS statistic of the meQTLs was denoted within each gene as its gene scores for subsequent analysis.

Gene set enrichment analysis. The GWAS-based GSEA algorithm was adopted for this study.[3,10] The latest gene set annotation database, the Molecular Signatures Database (MSigDB, Version 5.1, Broad Institute, Cambridge, Massachusetts), containing a total of 13 311 annotated gene sets, was downloaded from a publicly available database (http://software.broadinstitute.org/gsea/msigdb/index.jsp).[11] A total of 5000 permutations were conducted to calculate the p-value of each gene set.

Results

We found multiple eQTLs-based pathways for BMD, some of which showed significant associations in more

expression quantitative trait loci - based results   **methylation quantitative trait loci - based results**
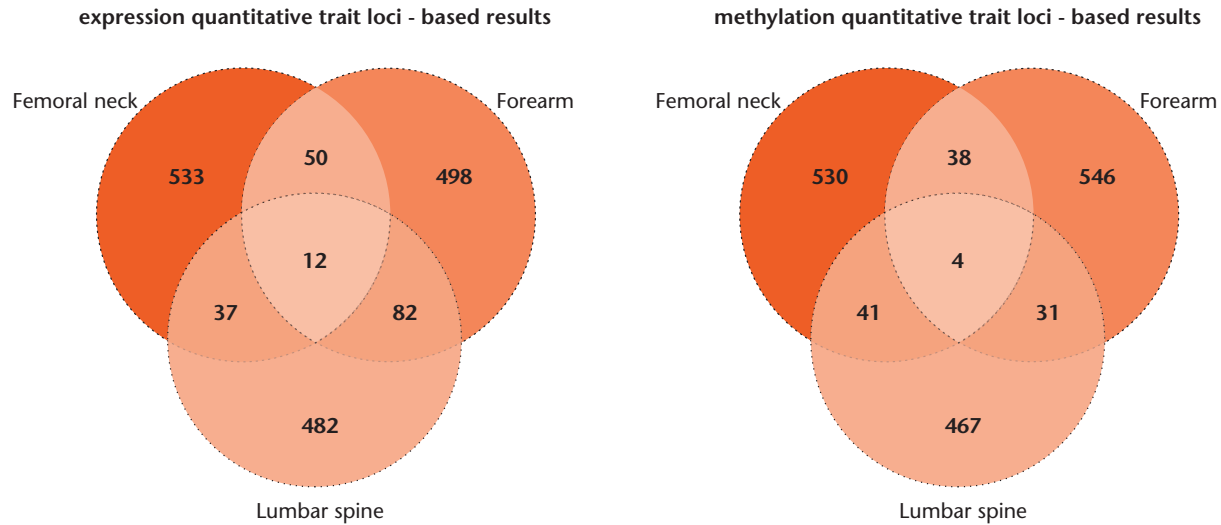

Fig 1

Summary of the expression of eQTL-based and meQTL-based pathway association analysis. Different pathways were associated with BMD in different regions (FN, femoral neck; LS, lumbar spine; FA, forearm).

Table I. Gene sets that achieved significance in all three regions.

Pathway name	p-value*		
	Forearm	**Femoral neck**	**Lumbar spine**
Expression quantitative trait loci (eQTL)-based results			
GSE17721 PAM3CSK4 *vs* CPG 12H BMDM up	0.028	0.038	< 0.001
GSE5589 IL6 KO *vs* IL10 KO LPS and IL10 Stim macrophage 45 minutes up	0.036	0.043	0.002
Module 195	0.042	0.011	0.017
Module 480	0.014	0.005	0.014
NABA ECM glycoproteins	0.049	0.041	0.028
KEGG pathways in cancer	0.049	0.038	0.008
GSE14769 unstim *vs* 60 minutes LPS BMDM up	0.020	0.008	0.021
YTTCCNNNGGAMR unknown	0.037	0.002	0.032
Darwiche papilloma risk high up	0.040	0.001	0.038
GUO hex targets DN	0.047	0.045	0.024
Module 427	0.047	0.001	0.003
Module 279	0.039	< 0.001	0.003
Methylation quantitative trait loci (meQTL)-based results			
Chang cycling genes	0.038	0.010	0.009
GNF2 KISS1	0.043	0.049	0.013
GNF2 MMP11	0.032	0.043	0.013
chr6q14	0.026	0.006	0.030

*Kolmogorov-Smirnov running sum statistics was used and p-values were decided based on permutation[3]

than one organ. Within each region, several pathways achieved significance (Fig. 1). In more than one region, a total of 181 gene sets were found to be significantly associated with BMD. Analogously, for meQTLs-based results, a total of 114 gene sets were identified that also achieved significance in more than one region (Table I, Supplementary Tables i and ii).

The results included multiple pathways with a known biological relevance to BMD, such as Reactome Circadian Clock (p < 0.001 for lumbar spine and 0.027 for femoral neck, respectively, in eQTLs-based GSEA) and insulin-like

growth factor (IGF) receptor binding (p < 0.001 for femoral neck BMD and 0.026 for lumbar spine BMD, respectively, in meQTLs-based GSEA).

Discussion

Multiple GWAS studies have been conducted using considerable sample sizes,[1,2,12-14] yet the overall results have only partly explained the genetic variance for BMD. This implies that the commonly used GWAS analytical strategy lacks the power to interpret the large quantity of information available within the GWAS datasets.[15] To

study the genetic background of bone metabolism further and to make full use of the existing large amounts of BMD GWAS data, we have conducted a re-investigation of previous results by integrating genome- and transcriptome- /epitome- level data, and have identified several gene sets associated with BMD.

Our eQTLs-based pathway association analysis also identified an interesting gene set, namely the Reactome Circadian Clock, which we found to be significantly associated with BMD in the regions of the femoral neck and lumbar spine. The role of the circadian clock in bone physiology and pathology has been characterised in many studies. Histological studies have identified the existence of circadian rhythm in both bone formation and resorption.[16,17] Biochemical parameters of bone remodelling have been shown to have a clear circadian pattern.[18,19] Moreover, functional studies have revealed that genes involved in mineral deposition also behave in a circadian fashion.[20] An animal experiment reported that a deficiency of the circadian clock protein BMAL1 in mice resulted in a low bone mass phenotype.[21] Epidemiological studies have found that individuals undertaking shift work developed circadian disruption, resulting in lower BMD and an increased risk of osteoporosis.[22] These findings have all demonstrated a relationship between the circadian clock and BMD. Our study suggests that the circadian clock may affect the mineral metabolism of human bones by using the eQTLs effect.

Another typical gene set was the IGF receptor binding. This small gene set contains several other genes that interact selectively with the IGF receptors, including IGF-1 and IGF-2, which are essential for maintaining bone mass. These potent anabolic peptides promote formation in bone modelling.[23] The synthesis of IGF-1 by osteoblasts is under the control of growth hormones and a deficiency of the growth hormone (GH)/IGF axis contributes to the development of low bone mass.[24] Our meQTLs-based gene set enrichment analysis reported significant associations between this gene set and the femoral neck and lumbar spine BMD. We also identified significant associations between Reactome GH receptor signalling and BMD in both the femoral neck and the lumbar spine, suggesting that genetically controlled methylation of the GH/IGF axis may play a critical role in bone metabolism.

In our study, however, none of the previously mentioned gene sets reached significance in the forearm, which may be because of the relatively small sample size of this area in our study, or may be due to the tissue specificity in that region. Despite these factors, there were still pathways that achieved significance in all three regions, few of which have been reported before. For example, our eQTLs-based analysis found YTTCCNNNGGAMR, consisting of 52 genes located in the promoter regions, which has not previously been associated with BMD in any of the three regions that we have studied. To the best of our knowledge, the common motif within the gene sets does not match any known transcription factor. These findings could help in subsequent functional studies of BMD.

The main purpose of this study was to demonstrate that SNPs identified by GWAS exert their effects on phenotypes mainly or partly by influencing the gene expression (GE) and by methylation.[8,9] These regulatory genetic variants, located in non-coding areas or regions, have often been neglected in biological and functional analysis. Our integration analysis took advantage of both pathway strategies and eQTLs and meQTLs results, which should allow for better anticipation of such variants. Our analysed framework does not require individual-level genotype data and can be applied to other open access large-scale GWAS summary data sets. Our study employed classic GSEA to conduct enrichment analysis, as described by Maciejewski.[25] GSEA, which is a preferable tool for gene set association analysis, produces biologically interpretable p-values and generally performs well.[25] Additionally, the GSEA statistic was employed in our study as it is one of the most popular choices for pathway association analysis,[5,6,26-28] although other proper enrichment algorithms can be used.[29]

There are limitations to our study. First, both the eQTLs and meQTLs data are based on blood samples, and as such may lose power for eQTLs/meQTLs analysis, especially when considering tissue-specific effects. Further high-quality eQTL and meQTL mapping studies will be required for more accurate analysis. Second, a standard SMR analysis incorporates a heterogeneity in dependent instruments (HEIDI) test to distinguish linkage from causal/pleiotropy association. However, such processes were evaluated to decrease the overall number of enrolled genes. As this may decrease the power of subsequent pathway analysis, these processes were not used in our study. The meQTL data were only employed as a filter, which may have less power compared with SMR strategy. SMR-based meQTL analysis, as well as other forms of *a priori* information, may also suit the framework of Mendelian randomisation and hence warrant further investigation.

In conclusion, we have identified a number of candidate gene sets for BMD through integrating the GWAS, eQTLs and meQTLs summary data sets. Our results have provided novel clues for subsequent functional analysis and our study has illustrated the integration of eQTLs and meQTLs data into pathway association analysis for genetic studies of complex human diseases.

Supplementary information

ë Further information regarding methods and tables showing eQTLs-based and meQTLs-based gene set enrichment analysis results is available alongside this article online at www.bjr.boneandjoint.org.uk

References

1. **Estrada K, Styrkarsdottir U, Evangelou E, et al.** Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* 2012;44:491-501.

2. **Zheng HF, Forgetta V, Hsu YH, et al.** Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* 2015;526:112-117.

3. **Wang K, Li M, Bucan M.** Pathway-based approaches for analysis of genome wide association studies. *Am J Hum Genet* 2007;81:1278-1283.

4. **Zhang L, Guo YF, Liu YZ, et al.** Pathway-based genome-wide association analysis identified the importance of regulation-of-autophagy pathway for ultradistal radius BMD. *J Bone Miner Res* 2010;25:1572-1580.

5. **Bouzigon E, Nadif R, Thompson EE, et al.** A common variant in RAB27A gene is associated with fractional exhaled nitric oxide levels in adults. *Clin Exp Allergy* 2015;45:797-806.

6. **Ambalavanan N, Cotten CM, Page GP, et al.** Integrated genomic analyses in bronchopulmonary dysplasia. *J Pediatr* 2015;166:531-537.e13.

7. **O'Dushlaine C, Kenny E, Heron EA, et al.** The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* 2009;25:2762-3.

8. **Zhu Z, Zhang F, Hu H, et al.** Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;48:481-487.

9. **McClay JL, Shabalin AA, Dozmorov MG, et al.** High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol* 2015;16:291.

10. **Wen Y, Wang W, Guo X, Zhang F.** PAPA: a flexible tool for identifying pleiotropic pathways using genome-wide association study summaries. *Bioinformatics* 2016;32:946-948.

11. **Subramanian A, Tamayo P, Mootha VK, 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.

12. **Kemp JP, Morris JA, Medina-Gomez C, et al.** Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet* 2017. (Epub ahead of print)

13. **Duncan EL, Danoy P, Kemp JP, et al.** Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet* 2011;7:e1001372.

14. **Kung AW, Xiao SM, Cherny S, et al.** Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am J Hum Genet* 2010;86:229-239.

15. **Yang TL, Guo Y, Zhang JG, et al.** Genome-wide Survey of Runs of Homozygosity Identifies Recessive Loci for Bone Mineral Density in Caucasian and Chinese Populations. *J Bone Miner Res* 2015;30:2119-2126.

16. **Simmons DJ.** Circadian mitotic rhythm in epiphyseal cartilage. *Nature* 1964;202:906-907.

17. **Russell JE, Simmons DJ, Huber B, Roos BA.** Meal timing as a Zeitgeber for skeletal deoxyribonucleic acid and collagen synthesis rhythms. *Endocrinology* 1983;113:2035-2042.

18. **Greenspan SL, Dresner-Pollak R, Parker RA, London D, Ferguson L.** Diurnal variation of bone mineral turnover in elderly men and women. *Calcif Tissue Int* 1997;60:419-423.

19. **Shao P, Ohtsuka-Isoya M, Shinoda H.** Circadian rhythms in serum bone markers and their relation to the effect of etidronate in rats. *Chronobiol Int* 2003;20:325-336.

20. **McElderry JD, Zhao G, Khmaladze A, et al.** Tracking circadian rhythms of bone mineral deposition in murine calvarial organ cultures. *J Bone Miner Res* 2013;28:1846-1854.

21. **Samsa WE, Vasanji A, Midura RJ, Kondratov RV.** Deficiency of circadian clock protein BMAL1 in mice results in a low bone mass phenotype. *Bone* 2016;84:194-203.

22. **Feskanich D, Hankinson SE, Schernhammer ES.** Nightshift work and fracture risk: the Nurses' Health Study. *Osteoporos Int* 2009;20:537-542.

23. **Celiker R, Arslan S.** Comparison of serum insulin-like growth factor-1 and growth hormone levels in osteoporotic and non-osteoporotic postmenopausal women. *Rheumatol Int* 2000;19:205-208.

24. **Lindsey RC, Mohan S.** Skeletal effects of growth hormone and insulin-like growth factor-I therapy. *Mol Cell Endocrinol* 2016;432:44-55.

25. **Maciejewski H.** Gene set analysis methods: statistical models and methodological differences. *Brief Bioinform* 2014;15:504-518.

26. **Brossard M, Fang S, Vaysse A, et al.** Integrated pathway and epistasis analysis reveals interactive effect of genetic variants at TERF1 and AFAP1L2 loci on melanoma risk. *Int J Cancer* 2015;137:1901-1909.

27. **Hou S, Du L, Lei B, et al.** Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p31.2 and 10q21.3. *Nat Genet* 2014;46:1007-1011.

28. **Li WD, Jiao H, Wang K, et al.** Pathway-Based Genome-wide Association Studies Reveal That the Rac1 Pathway Is Associated with Plasma Adiponectin Levels. *Sci Rep* 2015;5:13422.

29. **Huang DW, Sherman BT, Lempicki RA.** Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1-13.

Author Contribution

■ W. Wang: Conceiving and designing the study, Analysing and interpreting the data, Writing the manuscript
■ S. Huang: Designing the study, Reviewing and editing the manuscript
■ W. Hou: Writing the paper, Reviewing and editing the manuscript
■ Y. Liu: Reviewing and editing the manuscript
■ Q. Fan: Collecting the data, Literature research
■ A. He: Collecting the data, Literature research
■ Y. Wen: Collecting the data, Literature research
■ J. Hao: Collecting the data
■ X. Guo: Final version approval, Reviewing and editing the manuscript
■ F. Zhang: Final version approval, Reviewing and editing the manuscript

■ *W. Wang, S. Huang, and W. Hou contributed equally to this paper.