HHS Public Access

Author manuscript

Bone. Author manuscript; available in PMC 2019 January 25.

Published in final edited form as:

Bone. 2018 August ; 113: 41–48. doi:10.1016/j.bone.2018.05.012.

Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions

Xiang-He Meng[a], **Xiang-Ding Chen**[a], **Jonathan Greenbaum**[b], **Qin Zeng**[a], **Sheng-Lan You**[a], **Hong-Mei Xiao**[c], **Li-Jun Tan**[a,*], and **Hong-Wen Deng**[a,b,c,**]

[a]Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan 410081, China

[b]Center of Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA

[c]Institute of Reproduction and Stem Cell Engineering, School of Basic Medical Science, Central South University, Changsha, Hunan 410013, China

Abstract

Purpose: Osteoporosis is a common global health problem characterized by low bone mineral density (BMD) and increased risk of fracture. Genome-wide association studies (GWAS) have identified > 100 genetic loci associated with BMD. However, the functional genes responsible for most associations remain largely unknown. We conducted an innovative summary statistic data-based Mendelian randomization (SMR) analysis to identify novel causal genes associated with BMD and explored their potential functional significance.

Methods: After quality control of the largest GWAS meta-analysis data of BMD and the largest expression quantitative trait loci (eQTL) meta-analysis data from peripheral blood samples, 5967 genes were tested using the SMR method. Another eQTL data was used to verify the results. Next we performed a fine-mapping association analysis to investigate the functional SNP in the identified loci. Weighted gene co-expression network analysis (WGCNA) was used to explore functional relationships for the identified novel genes with known putative osteoporosis genes. Further, we assessed functions of the identified genes through *in vitro* cellular study or previous functional studies.

Results: We identified two potentially causal genes (*ASB16-AS1* and *SYN2*) associated with BMD. *SYN2* was a novel osteoporosis candidate gene and *ASB16-AS1* locus was known to be associated with BMD but was not the nearest gene to the top GWAS SNP. Fine-mapping association analysis showed that rs184478 and rs795000 was predicted to be possible causal SNPs in *ASB16-AS1* and *SYN2*, respectively. *ASB16-AS1* co-expressed with several known putative osteoporosis risk genes. *In vitro* cellular study showed that over-expressed *ASB16-AS1* increased

*Correspondence to: L.-J. Tan, Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan, China. ljtan@hunnu.edu.cn. **Correspondence to: H.-W. Deng, Center of Bioinformatics and Genomics, Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 2001, New Orleans, LA 70112, USA. hdeng2@tulane.edu.

Competing interests

The authors have no financial interests to disclose.

the expression of osteoblastogenesis related genes (*BMP2* and *ALPL*), indicating its functional significance.

Conclusion: Our findings support that *ASB16-AS1* and *SYN2* may represent two novel functional genes underlying BMD variation. The findings provide a basis for further functional mechanistic studies.

Keywords

Osteoporosis; BMD; Summary data-based Mendelian randomization (SMR); Weighted gene co-expression network analysis (WGCNA)

1. Introduction

The widespread application of genome-wide association studies (GWAS) has contributed to a revolution in the research of human genetics and the genetic determinants that underlie complex disease. A recent study suggested that selecting genetically supported therapeutic targets implicated by GWAS could potentially double the success rate in the clinical development of new pharmaceutical treatments [1]. While GWAS have identified thousands of genetic variants associated with human complex traits and diseases, there are still several important limitations to these prevailing types of genetic association studies [2,3]. A longstanding challenge of GWAS lies in exploring the mechanisms by which these genetic loci affect the variation of complex traits and the pathophysiology of complex diseases.

Osteoporosis is a common skeletal disease characterized by reduced bone mineral density (BMD) and increased risk of low trauma fractures [4]. In the United States, it has been estimated that the prevalence rate of osteoporosis in older adults was about 10.3%, while low bone mass prevalence rate was 43.9% [5]. In China, the prevalence rate of osteoporosis in older adults was estimated to be 15.7%, and it will increase rapidly with the increasing age of the total population [6]. Despite the significant impact on human health, there is still a lack of highly effective osteoporosis treatments that are free of negative side effects [7]. Hence, identification of additional therapeutic molecular targets for effective and efficient prevention and treatment of osteoporosis are needed. Heritability for BMD is estimated to be > 50% [8,9] and more than one hundred loci have been found to be associated with BMD or osteoporosis by previous studies [10–16]. However, although many associated genetic variants have been identified by GWAS, the causal variants truly with biological effects have remained largely unknown. Furthermore, only ~10% of the total BMD heritability has been explained by the current GWAS findings [14]. Additional genes and biological mechanisms underlying osteoporosis could be identified from existing GWAS data by using novel biostatistic and bioinformatic methods [17].

A genetic variant that influences a particular gene-expression level is known as expression quantitative trait loci (eQTL). Two studies using eQTL analysis in primary bone cell cultures discovered numerous eQTLs associated with BMD [18,19]. Kwan et al. found an eQTL (rs136564) regulated the expression of a novel transcript of FAM118A, and this eQTL was also found to be associated with BMD by GWAS analysis [19]. There are many other studies that have attempted to assess whether a SNP detected by GWAS was also an eQTL [20];

however most of them have analyzed the GWAS data and eQTL data in two sequential and separate steps rather than in an integrative manner [21]. Zhu et al. recently proposed an integrative method called summary data–based Mendelian randomization (SMR) that integrates independent GWAS summary statistics data with eQTL data from whole blood tissue to identify potential functionally relevant genes at the loci identified in GWAS and to identify novel trait-associated genes for five complex traits [22], and further applied the method to the analyses of another 28 traits [23]. It has been shown that eQTL effect in blood can be a proxy for eQTL effect in most relevant tissues for various traits or diseases [22,24]. Mendelian randomization is an instrumental variable analysis approach that uses genetic variants as instrumental variables (for example, eQTLs) to test whether an exposure (for example, the expression level of a gene) has a causal effect on an outcome (for example, trait value or disease risk) [25,26]. In this study, the phenotypic trait is the outcome (Y), gene expression is the exposure (X), and the top cis-eQTL that is strongly associated with gene expression is used as the instrumental variable (Z). The previous study [22] suggested that there are three models consistent with a significant association from the SMR test using only a single genetic variant. These three models include causality (Z → X → Y), pleiotropy (Z → X and Z → Y) and linkage (Z1 → X, Z2 → Y, and Z1 and Z2 are two variants in linkage disequilibrium (LD) in the cis-eQTL region). The purpose of this study was to identify genes whose expression levels had causal effect (Z → X → Y) on BMD. In Mendelian randomization studies, multiple uncorrelated instrumental variables (IVs) [22] (for example, the uncorrelated trans-eQTLs and/or cis-eQTLs) or multiple correlated IVs [27] were needed to identify the causality. Due to multiple uncorrelated IVs were not available in the Westra eQTL study [24], we applied weighted generalized linear regression method using correlated IVs to distinguish causality from pleiotropy.

It is clear that complex bone traits, like BMD, are not only the results of cumulative effects of individual genetic factor, but also the effects from interactive biological networks [28]. The weighted gene co-expression network analysis (WGCNA) is a useful method to construct the co-expression network [29]. Co-expression networks are modular, and each module represents a group of co-expressed genes. These modules tend to contain genes involved in similar biological processes [7]. Hence, under WGCNA, we explored the functional relevance of the identified novel genes to other known putative osteoporosis genes in order to explore the potential functional mechanisms of the identified novel genes.

In this study, we will identify genes whose expression levels have causal effects on BMD by using the SMR method and the weighted generalized linear regression method, followed by assessing functions of the identified genes through *in vitro* cellular study or previous functional studies.

2. Materials and methods

2.1. Data used in this study

2.1.1. GWAS summary data—The GEnetic Factors for OSteoporosis (GEFOS) Consortium used meta-analysis of whole genome sequencing, whole exome sequencing and deep imputation of genotype data (in reference to the UK10K and 1000Genomes data) to identify low-frequency and rare variants associated with risk of osteoporosis in 53,236

Caucasians [16]. The detailed description about genotype and imputation of GWAS data can be found in the previous study [16]. Each SNP with a minor allele frequency (MAF) > 0.5% was tested for association with an additive effect on femoral neck (FN), lumbar spine (LS) and forearm (FA) BMD, adjusting for sex, age, age^2 and weight [16]. The summary statistic data are available online (http://www.gefos.org/).

2.1.2. eQTL summary data—Westra et al. performed the largest eQTL meta-analysis so far in non-transformed peripheral blood samples of 5311 European healthy individuals with replication in 2775 European individuals [24]. Another eQTL study [30] which were performed to investigate the genetic architecture of gene expression (GAGE) in peripheral blood in 2765 European individuals was used to verify the results. It is widely accepted that eQTL effect in blood tissue can be a proxy for eQTL effects in most relevant tissues for various traits or diseases [22,24]. Particularly, there are several types of cells such as peripheral blood monocytes (PBMs) and B and T lymphocytes that are related to bone metabolism [31]. For example, PBMs have been well established as a working cell model for studying gene expression patterns in relation to osteoporosis risk *in vivo* in humans [32]. PBMs may act as precursors of osteoclasts since they can differentiate into osteoclasts [33], and they express different cytokines which are important for osteoclast differentiation, activation, and apoptosis [32,34]. B lymphocytes, an important cell type of the immune system, express/secrete factors involved in osteoclastogenesis, such as receptor tumor necrosis factor superfamily member 11 and osteoprotegerin [35]. The gene expression data were quantile-normalized to the median distribution, and subsequently $\log_2$ transformed [24]. The eQTL summary data in SMR binary format can be downloaded from http://cnsgenomics.com/software/smr/download.html.

2.1.3. Gene expression data—We used previously published gene expression profile generated from PBMs [36] to generate gene co-expression networks to assess the potential interactions and thus potential mechanisms for the identified novel gene. PBMs were from 73 Caucasians females, which were stratified by hip BMD and menopausal status. In the high BMD group there were 42 subjects with 16 premenopausal and 26 postmenopausal women. In the low BMD group there were 31 subjects with 15 premenopausal and 16 postmenopausal women. Details of the samples' information can be found in the previous study [32]. We downloaded the raw data under the accession number GSE56814 from Gene Expression Omnibus (GEO) website.

2.2. Statistical analysis methods

2.2.1. Summary data–based Mendelian randomization (SMR) analysis—The SMR method was detailed in the previous paper [22]. In brief, SMR applies the principles of Mendelian randomization (MR) [37,38] to jointly analyze GWAS and eQTL summary statistics in order to test for association between gene expression and a trait due to a shared variant at a locus. In this study, the phenotypic trait is the outcome (Y), gene expression is the exposure (X), and the top cis-eQTL that is strongly associated with gene expression is used as the instrumental variable (Z). Equivalently, it is an analysis to test whether the effect of Z on Y is mediated by X (a model of Z → X → Y). The SMR software was downloaded from http://cnsgenomics.com/software/smr/.

Since the SMR analysis assumes that the instrument (top cis-eQTL) has a strong effect on the exposure (gene's expression level), only probes with at least one cis-eQTL at P_{eQTL} (a p value from the eQTL study indicating the significance of the eQTL associated with the gene expression) smaller than 5×10^{-8} in the cis-eQTL region were included in the eQTL summary data (hg19). We excluded cis-eQTL with MAF < 0.01 and cis-eQTL in the MHC region because of the complexity of LD patterns in this region [22]. After data processing, there were 5967 probes left. So the genome-wide significance level for SMR test was $P_{smr} < 8.4 \times 10^{-6}$ (0.05/5967, Bonferroni Correction).

Since the significant SMR results could also reflect linkage model that was of less biological interest, we used the heterogeneity in dependent instruments (HEIDI) test to distinguish pleiotropy (or causality) model from linkage model. The HEIDI test considers the pattern of associations using all the SNPs that are significantly associated with gene expression (eQTLs) in the cis-eQTL region (± 250 kb from the center of the gene probe). Under Hardy-Weinberg equilibrium and LD, b_{XY} estimated at the top associated cis-eQTL ($b_{XY(top)}$) will be equal to that estimated at any of the cis-SNPs in LD that is associated with gene expression. In the HEIDI test, we excluded the SNPs in strong LD with top cis-eQTL at $r^2 > 0.9$ which was calculated by using individual level data from the HapMap2 CEU. We also removed SNPs in the cis-eQTL region with a $P_{eQTL} > 1.6 \times 10^{-3}$ (equivalent to $\chi^2 < 10$) to avoid weak instrumental variables according to the original paper [22]. The null hypothesis of the HEIDI test was that there was a single causal variant which meant that there was no heterogeneity in the b_{XY} values. A p value threshold of $P_{HEIDI} > 0.05$ was conservative for a gene having "pleiotropy" effects (no heterogeneity) [22].

Next, causal estimate was conducted by applying weighted generalized linear regression method using correlated IVs [27] to distinguish causality model from pleiotropy model. First, we assume the estimate of association for IV k = 1, …, K with the gene's expression (X) is β_{kX} with standard error σ_{kX}. The estimate of association for IV k with BMD can be expressed as β_{kY} with standard error σ_{kY}. The correlation between IVs k_1 and k_2 can be defined as ρ_{k1k2}. Then we can perform a weighted generalized linear regression of the β_{kY} parameters on the β_{kX} parameters using the σ_{kY}^{-2} parameters as inverse-variance weights and taking into account the correlation between the IVs. If we define $\Omega_{k1k2} = \sigma_{k1Y}\sigma_{k2Y}\rho_{k1k2}$, b_{XY} can be estimated from a weighted generalized linear regression as $b_{XY} = (\beta_{kX}{}^T \Omega \beta_{kX})^{-1} \beta_{kX}{}^T \Omega^{-1} \beta_{kY}$ [39], with $\beta_{kX}{}^T$ being the transposed vector of β_{kX}. The standard error of the estimate is $\mathrm{se}(b_{XY}) = \sqrt{\left(\beta_{kX}^T \Omega^{-1} \beta_{kX}\right)^{-1}}$ [27]. After pruning for SNPs with linkage disequilibrium $r^2 > 0.8$, all variants with a $P_{eQTL} < 5 \times 10^{-8}$ in the cis-eQTL region of each probe available in both BMD and eQTL data were included in the analysis. The correlation between IVs was calculated in Plink software (version 1.0.7) by using HapMap 2 CEU as a reference panel. Cochran's Q statistic was performed to test heterogeneity using the *meta* R packages [27], a p value > 0.05 can be interpreted as no more heterogeneity between causal effects estimated using the variants individually than would be expected by chance.

2.2.2. The regional association plot—SNAP (for SNP Annotation and Proxy Search, https://archive.broadinstitute.org/mpg/snap/ldsearch.php) was used to characterize the

regional association plot for the interesting genes [40]. The LD between the putative causal SNP and other SNPs were calculated by using HapMap 2 CEU as a reference panel.

2.2.3. Fine-mapping—It is well documented that a SNP overlapping with a functional region is more likely to be a functional SNP [41]. To identify the causal SNPs in each identified gene, we used the annotation data from RegulomeDB [42], and performed a fine-mapping analysis called Probabilistic Identification of Causal SNPs (PICS) [43] to estimate the possibility of each SNP in the locus to be a causal SNP. By using information derived from the permutations, PICS score was calculated as the posterior probability of each SNP being the causal variant, given the observed pattern of association at the locus [43].

2.2.4. WGCNA—To identify functional connections for the identified novel genes with known genes, we performed WGCNA in gene expression profiles generated from PBMs. We used RMA (robust multiarray average) algorithm to correct for the background noise and normalize the expression data with the Expression Console software. The software can be downloaded from the Affymetrix website: http://www.affymetrix.com. The co-expression networks were generated using the WGCNA R package [44]. We extended gene boundaries by 20 kb upstream and downstream of the gene [45]. Any gene containing a SNP in the extended region with a p value < 0.01 for at least one of the two BMD trait (FN-BMD and LS-BMD) and 234 genes which were identified to be associated with BMD by previous studies [12,46,47] are here referred to as the nominally significant GWAS geneset (4796 genes). After excluding the non-expressed genes, we identified 3593 probes representing 3593 genes to construct the co-expression network. Finally, we exported the network of the interesting genes with Topology Overlap Matrix (TOM) value > 0.15. We visualized the networks by Cytoscape [48]. Due to that the TOM value could not exhibit the positive or negative correlation between co-expressed genes, we defined the signed expression similarity as the value of the correlation coefficient between the profiles of genes i and j: $s_{ij} = cor(x_i,x_j)$. Person correlation coefficient was calculated by using R software.

2.3. Gene overexpression experiment for the identified ASB16-AS1

To validate the functional implication of the *ASB16-AS1* gene in osteoblast, we performed gene overexpression experiment. Total cellular RNA was isolated from the human fetal osteoblastic 1.19 cell line (hFOB1.19, ATCC, Cat CRL-11372). Since the *ASB16-AS1* transcription variant 2 uses an alternate splice site in the terminal exon compared to variant 1, the transcription variant 2 is shorter than variant 1. The same primers were used to amplify both the variants, only variant 2 was successfully subcloned into the pCEP4 vector. The primers for amplification of *ASB16-AS1* were 5′-ggggtaccgtggcttcgcgactgcggaaggt-3′ (forward *Kpn*I) and 5′- cgggatccttttttttttttttttttttatttttttttggtgcatactgtttaattttct-3′ (reverse *BamH*I). The procedure of cell culture was described previously [49]. Cells were seeded at 6×10^5 cells/well in 6-well plates. After 24 h, cells were transfected with Lipofectamine 3000, expression vector pCEP4-ASB16-AS1 using pCEP4 as a control. Real-time Polymerase Chain Reaction (RT-PCR) was used to detect the relative mRNA level of osteoblastic genes (*BMP2* and *ALPL*) 48 h later after transfection. The primers for RT-PCR were shown in Table S1. Student's *t*-test was applied to determine statistical significance.

3. Results

3.1. SMR tests

SMR tests identified two genes (*ASB16-AS1* and *TMUB2*) that were significantly associated with FN-BMD, two genes (*SYN2* and *LRP3*) were associated with LS-BMD and no genes were identified to be associated with FA-BMD (Table 1). Of the four genes, only *ASB16-AS1* and *SYN2* passed HEIDI tests ($P_{HEIDI} > 0.05$), suggesting that there was no heterogeneity and gene expression of *ASB16-AS1* (or *SYN2*) and the BMD variation were affected by the same variant. Using the GAGE eQTL data, we re-conducted the SMR tests. *ASB16-AS1* was successfully verified to be associated with FN-BMD with $P_{SMR} = 1.98 \times 10^{-8}$ and $P_{HEIDI} = 0.13$. *SYN2* was successfully verified to be associated with LS-BMD with $P_{SMR} = 2.51 \times 10^{-5}$ and $P_{HEIDI} = 0.31$. In the cis-eQTL region of *SYN2* probe, there was no SNP with a GWAS p value $< 5 \times 10^{-8}$ (Fig. 1B). Therefore, *SYN2* was a new gene associated with BMD discovered by the SMR test. The details of the results were shown in Table 1. The genes *TMUB2* and *LRP3* failed to pass the HEIDI tests suggesting there were two genetic variants, one affecting the gene expression and the other affecting the BMD variation. Subsequently, our study only focused on the two genes (*ASB16-AS1* and *SYN2*) with potential functionally relevant variants in the downstream further analysis.

In the cis-eQTL region of *ASB16-AS1* probe, there were 73 SNPs with a $P_{eQTL} < 5 \times 10^{-8}$. After pruning for SNPs with LD $r^2 > 0.8$, a total of 16 SNPs in the cis-eQTL region of *ASB16-AS1* probe were included to detect whether the expression of *ASB16-AS1* had a causal effect on BMD. The estimated effect size of the expression of *ASB16-AS1* on BMD (b_{XY}) was −0.14 (95% CI: −0.18, −0.10). The p value from Cochran's Q statistic of heterogeneity in causal estimates for each genetic variant calculated individually was 0.45, suggesting that no heterogeneity existed. There were 160 eQTLs with a p value $< 5 \times 10^{-8}$, in the cis-eQTL region of *SYN2* probe. After pruning, 15 SNPs were included to give the causal estimate with a p value from Cochran's Q statistic 0.16. The estimated effect size of the expression of *SYN2* on BMD (b_{XY}) was −0.20 (95% CI: −0.27, −0.13). Association results of *ASB16-AS1* or *SYN2* with BMD were displayed graphically in Fig. 2.

3.2. Fine-mapping

With the functional annotation data from RegulomeDB, in the *ASB16-AS1* locus we found that rs184478, which was in perfect LD with rs227580 (Fig. 3), was an eQTL overlapping with many functional regions such as conserved motifs, transcription factor binding site, DNase I hypersensitive peak, Chip-Seq peaks and DNase I footprint region, while the top cis-eQTL rs227580 only overlapped with transcription factor binding site and DNase I hypersensitive peak. The PICS score of rs184478 was 0.987, while the PICS score of rs2275580 was only 0.055. When using rs184478 instead to perform the SMR test, we found the P_{smr} value ($P_{smr} = 2.39 \times 10^{-8}$) was smaller (thus more significant) than using the top cis-eQTL ($P_{smr} = 1.30 \times 10^{-7}$). In the *SYN2* locus, rs795000 achieved the highest PICS score. When using rs795000 instead to perform the SMR test, the result was more significant than using the top cis-eQTL rs310758 (Table 1). Therefore, rs184478 and rs795000 had the largest possibility to be causal SNPs for *ASB16-AS1* and *SYN2* genes, respectively.

3.3. WGCNA

The biological roles of *ASB16-AS1* in bone metabolism are not well known. WGCNA is a useful system genomics method to construct gene co-expression network and predict functional connections (thus potential functional mechanisms) between novel genes and known genes. After excluding non-expressed genes, we selected 3593 probes representing 3593 genes (~75% of the nominally significant GWAS geneset which contains 4796 genes). After hierarchical clustering and dynamic tree cutting, we chose the gene modules which contained *ASB16-AS1* gene. Using Cytoscape, we visualized the gene centered co-expression network, respectively (Fig. 4). The network only showed genes which were shown to be associated with BMD in the previous studies and connected to the centered gene with a TOM > 0.15. With this criterion, there were 45 known genes co-expressed with ASB16-AS1 (Fig. 4). *CTNNB1* positively co-expressed with *ASB16-AS1*, while *TNFSF11* (cor = −0.43) was negatively co-expressed with *ASB16-AS1*.

3.4. Over-expression experiment of ASB16-AS1

The RT-PCR results showed that the expression of *ASB16-AS1* transcript variant 1 was higher than the expression of *AB16-AS1* transcript variant 2 (Fig. 5A). However, using the same primers, we failed to get a clone of variant 1. The variant 2 was successfully subcloned into the pCEP4 vector. After transfection of pCEP4-ASB16-AS1, the expression of *ASB16-AS1* were highly enhanced (Fig. 5B). The expression of *BMP2* and *ALPL* was significantly increased in the *ASB16-AS1* over-expressed osteoblast compared with the control (Fig. 5C and D). The expression of *RUNX2* was increased but the alteration was not significant (Fig. 5E).

4. Discussion

In the present study, by using the summary data-based Mendelian randomization analysis, we identified *ASB16-AS1* associated with FN-BMD and *SYN2* associated with LS-BMD and successfully verified by using another eQTL study. Under the WGCNA analysis, we found that *ASB16-AS1* was co-expressed with dozens of known putative osteoporosis genes including *CTNNB1, TNFSF11* and *RUNX2*. Over-expressed *ASB16-AS1* would increase the expression of osteoblastic genes (*BMP2* and *ALPL*).

The causal estimate of b_{XY} for *ASB16-AS1* was −0.14 which meant that a genetic increase in $\log_2$ transformed expression of *ASB16-AS1* by one standard deviation decreased BMD by 0.14 standard deviation. Similarly, the causal estimate of b_{XY} for *SYN2* was −0.20 which meant that a genetic increase in $\log_2$ transformed expression of *SYN2* by one standard deviation decreased BMD by 0.20 standard deviation.

SMR was an efficient method to identify associations between gene expression and complex traits using summary data from GWAS and eQTL studies. One advantage of SMR was that it was useful to prioritize functionally relevant genes in a trait/disease associated locus. A major challenge of GWAS in interpreting which gene is functionally associated with the trait comes from the fact that the significant SNP represents a large region of LD. Regions of strong LD can be very large, and the significant SNPs have been found in perfect LD with

the causal SNP hundreds of kilobases away [41]. The SMR analysis (including a SMR test and a HEIDI test) was useful to prioritize the functional gene in the associated region. The top associated GWAS SNP rs228769 was located in *HDAC5* gene region, 60 kb downstream from *ASB16-AS1* (Fig. 3). The gene *HDAC5* may contribute to osteoporosis etiology by controlling sclerostin expression in osteocytes and regulating osteoblast differentiation [50]. Since there was no probe for the *HDAC5* gene in both eQTL studies, the present study cannot exclude the causal role of *HDAC5* gene for BMD variation. However, our study clearly demonstrated that *ASB16-AS1* was a putative novel functional gene in this known locus.

Complex traits, like BMD, may be not only the results of cumulatively additive effects of individual genetic factors, but also the results of gene interaction via biological pathways/networks [28]. Under WGCNA, we found *ASB16-AS1* co-expressed with more than forty genes which were proven to be associated with BMD by previous studies. The result thus suggested that *ASB16-AS1* may play an important role in bone biology. In the *ASB16-AS1* centered network, gene *CTNNB1* encodes β-catenin protein which can aggregate and move into the nucleus and then can active wnt/β-catenin signaling pathway. The target gene of β-catenin protein, *RUNX2*, is active to promote osteoblast proliferation and differentiation [51–53]. Interestingly, these genes which mainly play roles in osteoblasts, were found to be co-expressed with *ASB16-AS1* in PBMs which are precursors of osteoclasts [33]. The reason why these genes express in PBMs was worth of further exploration. *TNFSF11* was negatively correlated with *ASB16-AS1*. RANKL encoded by *TNFSF11* is a key factor for osteoclast differentiation and activation [54]. Taken together, we predicted that *ASB16-AS1* may play important roles in osteoblast and osteoclast proliferation and differentiation. *BMP2* has been shown to potently induce osteoblast differentiation [55]. *ALPL*, coding for the liver/bone/kidney isozyme of alkaline phosphatase, are known to be regulated during osteoblastic differentiation [56]. The expression of *BMP2* and *ALPL* were increased after transfection of pCEP4-ASB16-AS1, suggesting that *ASB16-AS1* may promote the differentiation of osteoblast.

SMR was demonstrated that it was useful to prioritize novel genes associated with BMD. There was no GWAS signal within 0.5 Mb of the probe in *SYN2* (Fig. 1B). However, by the SMR tests, our study identified significant association signal for *SYN2*. SMR tests reduced the multiple hypothesis burdens by testing tens of thousands of genes instead of millions of SNPs [17]. The synapsins are a family of 4 synaptic vesicle-associated proteins which are products from alternative splicing of two genes, *SYN1* and *SYN2*. The previous study which showed that the release of glutamate *via* synapsin can directly promote osteoblast differentiation [57], suggesting that *SYN2* played an important role in osteoblast differentiation.

Additionally, there are several other methods for detecting the association between genes and traits using the GWAS and eQTL data [58–60]. The COLOC method [58] is a very useful tool to detect co-localization of GWAS and eQTL signals at known GWAS risk loci with a Bayesian analysis approach; however it does not provide thresholds for the posterior probabilities to control for genome wide false positive rate. The PrediXican method [59] directly tests the molecular mechanism through which genetic variation affects a study trait;

however it requires individual-level genotype and gene expression data in the training data set, and individual-level genotype and phenotype data in the target data set. Therefore, this approach may currently have limited power and feasibility because of the limited availability and small sample sizes of such required data sets at present. Gusev et al. [60] proposed a method to overcome this problem by performing a polygenic prediction analysis using summary-level statistic data. However, unlike the SMR method, neither PrediXcan nor the Gusev's method distinguishes between pleiotropy and linkage.

The present study may have some limitations. First, in the Westra et al. eQTL data [24], there were only 5967 probes so that it had incomplete and sparse genomic coverage of relevant genes, which may lead to the fact that only four positive findings using the SMR test. Some potential genes may thus be missed. Second, in both eQTL data, there is no probe for *HDAC5*. So SMR test did not yield any signal of *HDAC5* but we cannot exclude the importance of the gene. Despite this, we found *ASB16-AS1* was another putative novel functional gene for BMD at the locus. Third, the Gene Tissue Expression (GTEx) project which was an NIH funded effort planned to generate RNA-seq expression profiles from > 40 tissues in a large genotyped human cohort [61], did not collect bone tissue and primary bone cells. We thus cannot get eQTL data from bone related cells/tissue from GTEx for our analyses at present. Fourth, SMR may lose power compared to standard GWAS when the true biological mechanism is independent of gene expression [17].

In conclusion, the present study identified that *ASB16-AS1* and *SYN2* may causally affect BMD through their gene expression. *ASB16-AS1* may play an important role in osteoblast proliferation and differentiation, at least by interacting with known genes in these processes.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bone.2018.05.012.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all the study subjects for volunteering to participate in the study. We thank Genetic Factors for Osteoporosis Consortium provided the BMD replication dataset to download for us. We also thank Zhu et al. for providing the SMR software to download and Westra et al. and Luke R. Lloyd-Jones et al. for providing the eQTL data to download.

Funding

This study was supported by Natural Science Foundation of China (NSFC; 81570807, 30900810, 31271344 and 31071097); Hunan Provincial Construct Program of the Key Discipline in Ecology (0713) and the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486). Hong-Wen Deng was partially supported by grants from the National Institutes of Health [AR069055, U19 AG055373, R01 MH104680, R01 AR059781 and P20 GM109036], the Edward G. Schlieder Endowment fund to Tulane University.

Abbreviations:

BMD	bone mineral density

GWAS	genome-wide association studies
eQTL	expression quantitative trait loci
SMR	summary data-based Mendelian randomization
HEIDI	heterogeneity in dependent instruments
LD	linkage disequilibrium
WGCNA	weighted gene co-expression network analysis
IVs	instrumental variables
PICS	Probabilistic Identification of Causal SNPs
TOM	Topology Overlap Matrix

References

[1]. Nelson MR, Tipney H, Painter JL, et al., The support of human genetic evidence for approved drug indications, Nat. Genet 47 (2015) 856–860. [PubMed: 26121088]

[2]. Hindorff LA, Sethupathy P, Junkins HA, et al., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, Proc. Natl. Acad. Sci. U. S. A 106 (2009) 9362–9367. [PubMed: 19474294]

[3]. Welter D, MacArthur J, Morales J, et al., The NHGRI GWAS catalog, a curated resource of SNP-trait associations, Nucleic Acids Res. 42 (2014) D1001–6. [PubMed: 24316577]

[4]. Ammann P, Rizzoli R, Bone strength and its determinants, Osteoporos. Int 14 (Suppl. 3) (2003) S13–8. [PubMed: 12730800]

[5]. Wright NC, Looker AC, Saag KG, et al., The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine, J. Bone Miner. Res 29 (2014) 2520–2526. [PubMed: 24771492]

[6]. Lin X, Xiong D, Peng YQ, et al., Epidemiology and management of osteoporosis in the People's Republic of China: current perspectives, Clin. Interv. Aging 10 (2015) 1017–1033. [PubMed: 26150706]

[7]. Sabik OL, Farber CR, Using GWAS to identify novel therapeutic targets for osteoporosis, Transl. Res 181 (2017) 15–26. [PubMed: 27837649]

[8]. Pocock NAEJ, Hopper JL, Yeates MG, Sambrook PN, Eberl S, Genetic determinants of bone mass in adults.pdf, J. Clin. Invest 80 (3) (1987) 706–710. [PubMed: 3624485]

[9]. Krall EA, Dawson-Hughes B, Heritable and life-style determinants of bone mineral density, J. Bone Miner. Res 8 (1) (1993) 1–9. [PubMed: 8427042]

[10]. Rivadeneira F, Styrkarsdottir U, Estrada K, et al., Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies, Nat. Genet 41 (2009) 1199–1206. [PubMed: 19801982]

[11]. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al., Multiple genetic loci for bone mineral density and fractures, N. Engl. J. Med 358 (2008) 2355–2365. [PubMed: 18445777]

[12]. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al., New sequence variants associated with bone mineral density, Nat. Genet 41 (2009) 15–17. [PubMed: 19079262]

[13]. Hsu YH, Zillikens MC, Wilson SG, 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. 6 (2010) e1000977. [PubMed: 20548944]

[14]. 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 44 (2012) 491–501. [PubMed: 22504420]

[15]. Zhang L, Choi HJ, Estrada K, et al., Multistage genome-wide association meta-analyses identified two new loci for bone mineral density, Hum. Mol. Genet 23 (2014) 1923–1933. [PubMed: 24249740]

[16]. Zheng HF, Forgetta V, Hsu YH, et al., Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture, Nature 526 (2015) 112–117. [PubMed: 26367794]

[17]. Pasaniuc B, Price AL, Dissecting the genetics of complex traits using summary association statistics, Nat. Rev. Genet 18 (2016) 117–127. [PubMed: 27840428]

[18]. Grundberg E, Kwan T, Ge B, et al., Population genomics in a disease targeted primary cell model, Genome Res. 19 (2009) 1942–1952. [PubMed: 19654370]

[19]. Kwan T, Grundberg E, Koka V, et al., Tissue effect on genetic control of transcript isoform variation, PLoS Genet. 5 (2009) e1000608. [PubMed: 19680542]

[20]. Farber CR, Systems genetics: a novel approach to dissect the genetic basis of osteoporosis, Curr. Osteoporos. Rep 10 (2012) 228–235. [PubMed: 22802146]

[21]. Farber CR, Lusis AJ, Integrating Global Gene Expression Analysis and Genetics, 60 (2008), pp. 571–601.

[22]. Zhu Z, Zhang F, Hu H, et al., Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets, Nat. Genet 48 (2016) 481–487. [PubMed: 27019110]

[23]. Pavlides JMW, Zhu Z, Gratten J, et al., Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits, Genome Med. 8 (2016).

[24]. Westra HJ, Peters MJ, Esko T, et al., Systematic identification of trans eQTLs as putative drivers of known disease associations, Nat. Genet 45 (2013) 1238–1243. [PubMed: 24013639]

[25]. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, et al., Methodological challenges in Mendelian randomization, Epidemiology 25 (2014) 427–435. [PubMed: 24681576]

[26]. Boef AG, Dekkers OM, le Cessie S, Mendelian randomization studies: a review of the approaches used and the quality of reporting, Int. J. Epidemiol 44 (2015) 496–511. [PubMed: 25953784]

[27]. Burgess S, Dudbridge F, Thompson SG, Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods, Stat. Med 35 (2016) 1880–1906. [PubMed: 26661904]

[28]. Schadt EE, Molecular networks as sensors and drivers of common human diseases, Nature 461 (2009) 218–223. [PubMed: 19741703]

[29]. Ghazalpour A, Doss S, Zhang B, et al., Integrating genetic and network analysis to characterize genes related to mouse weight, PLoS Genet. preprint (2005) e130.

[30]. Lloyd-Jones LR, Holloway A, McRae A, et al., The genetic architecture of gene expression in peripheral blood, Am. J. Hum. Genet 100 (2017) 228–237. [PubMed: 28065468]

[31]. Chalmers TC, Smith H, Jr., Blackburn B, et al., A method for assessing the quality of a randomized control trial, Control. Clin. Trials 2 (1981) 31–49. [PubMed: 7261638]

[32]. Liu YZ, Dvornyk V, Lu Y, et al., A novel pathophysiological mechanism for osteoporosis suggested by an in vivo gene expression study of circulating monocytes, J. Biol. Chem 280 (2005) 29011–29016. [PubMed: 15965235]

[33]. Geissmann F, Manz MG, Jung S, et al., Development of monocytes, macrophages, and dendritic cells, Science 327 (2010) 656–661. [PubMed: 20133564]

[34]. Deng FY, Lei SF, Zhang Y, et al., Peripheral blood monocyte-expressed ANXA2 gene is involved in pathogenesis of osteoporosis in humans, Mol. Cell. Proteomics 10 (2011) M111.011700.

[35]. Manabe N, Kawaguchi H, Chikuda H, et al., Connection between B lymphocyte and osteoclast differentiation pathways, J. Immunol 167 (2001) 2625–2631. [PubMed: 11509604]

[36]. Liu YZ, Zhou Y, Zhang L, et al., Attenuated monocyte apoptosis, a new mechanism for osteoporosis suggested by a transcriptome-wide expression study of monocytes, PLoS One 10 (2015) e0116792. [PubMed: 25659073]

[37]. Smith GD, Ebrahim S, Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol 32 (2003) 1–22. [PubMed: 12689998]

[38]. Davey Smith G, Hemani G, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, Hum. Mol. Genet 23 (2014) R89–98. [PubMed: 25064373]

[39]. Lynch MWB, Genetics and Analysis of Quantitative Traits, SinauerAssociates, 1998.

[40]. Johnson AD, Handsaker RE, Pulit SL, et al., SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap, Bioinformatics 24 (2008) 2938–2939. [PubMed: 18974171]

[41]. Schaub MA, Boyle AP, Kundaje A, et al., Linking disease associations with regulatory information in the human genome, Genome Res. 22 (2012) 1748–1759. [PubMed: 22955986]

[42]. Boyle AP, Hong EL, Hariharan M, et al., Annotation of functional variation in personal genomes using RegulomeDB, Genome Res. 22 (2012) 1790–1797. [PubMed: 22955989]

[43]. Farh KK-H, Marson A, Zhu J, et al., Genetic and epigenetic fine mapping of causal autoimmune disease variants, Nature 518 (2014) 337–343. [PubMed: 25363779]

[44]. Langfelder P, Horvath S, WGCNA: an R package for weighted correlation network analysis, BMC Bioinforma. 9 (2008) 559.

[45]. Jia P, Zheng S, Long J, et al., dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks, Bioinformatics 27 (2011) 95–102. [PubMed: 21045073]

[46]. Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, et al., Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies, Nat. Genet 41 (11) (2009) 1199–1206. [PubMed: 19801982]

[47]. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, et al., Multiple genetic loci for bone mineral density and fractures, N. Engl. J. Med 358 (22) (2008) 2355–2365. [PubMed: 18445777]

[48]. Doerks T, Copley RR, Schultz J, et al., Systematic identification of novel protein domain families associated with nuclear functions, Genome Res. 12 (2002) 47–56. [PubMed: 11779830]

[49]. Deng FY, Tan LJ, Shen H, et al., SNP rs6265 regulates protein phosphorylation and osteoblast differentiation and influences BMD in humans, J. Bone Miner. Res 28 (2013) 2498–2507. [PubMed: 23712400]

[50]. Wein MN, Spatz J, Nishimori S, et al., HDAC5 controls MEF2C-driven sclerostin expression in osteocytes, J. Bone Miner. Res 30 (2015) 400–411. [PubMed: 25271055]

[51]. Krishnan V, Bryant HU, Macdougald OA, Regulation of bone mass by Wnt signaling, J. Clin. Invest 116 (2006) 1202–1209. [PubMed: 16670761]

[52]. Baron R, Kneissel M, WNT signaling in bone homeostasis and disease: from human mutations to treatments, Nat. Med 19 (2013) 179–192. [PubMed: 23389618]

[53]. Chen G, Wang C, Wang J, et al., Antiosteoporotic effect of icariin in ovariectomized rats is mediated via the Wnt/beta-catenin pathway, Exp. Ther. Med 12 (2016) 279–287. [PubMed: 27347050]

[54]. Wittrant Y, Theoleyre S, Chipoy C, et al., RANKL/RANK/OPG: new therapeutic targets in bone tumours and associated osteolysis, Biochim. Biophys. Acta 1704 (2004) 49–57. [PubMed: 15363860]

[55]. Marie PJ, Debiais F, Hay E, Regulation of human cranial osteoblast phenotype by FGF-2, FGFR-2 and BMP-2 signaling, Histol. Histopathol 17 (2002) 877–885. [PubMed: 12168799]

[56]. Swallow DM, Povey S, Parkar M, et al., Mapping of the gene coding for the human liver/bone/kidney isozyme of alkaline phosphatase to chromosome 1, Ann. Hum. Genet 50 (1986) 229–235. [PubMed: 3446011]

[57]. Yang JE, Song MS, Shen Y, et al., The role of KV7.3 in regulating osteoblast maturation and mineralization, Int. J. Mol. Sci 17 (2016) 407. [PubMed: 26999128]

[58]. Giambartolomei C, Vukcevic D, Schadt EE, et al., Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, PLoS Genet. 10 (2014) e1004383. [PubMed: 24830394]

[59]. Gamazon ER, Wheeler HE, Shah KP, et al., A gene-based association method for mapping traits using reference transcriptome data, Nat. Genet 47 (2015) 1091–1098. [PubMed: 26258848]

[60]. Gusev A, Ko A, Shi H, et al., Integrative approaches for large-scale transcriptome-wide association studies, Nat. Genet 48 (2016) 245–252. [PubMed: 26854917]

[61]. Mele M, Ferreira PG, Reverter F, et al., Human genomics. The human transcriptome across tissues and individuals, Science 348 (2015) 660–665. [PubMed: 25954002]

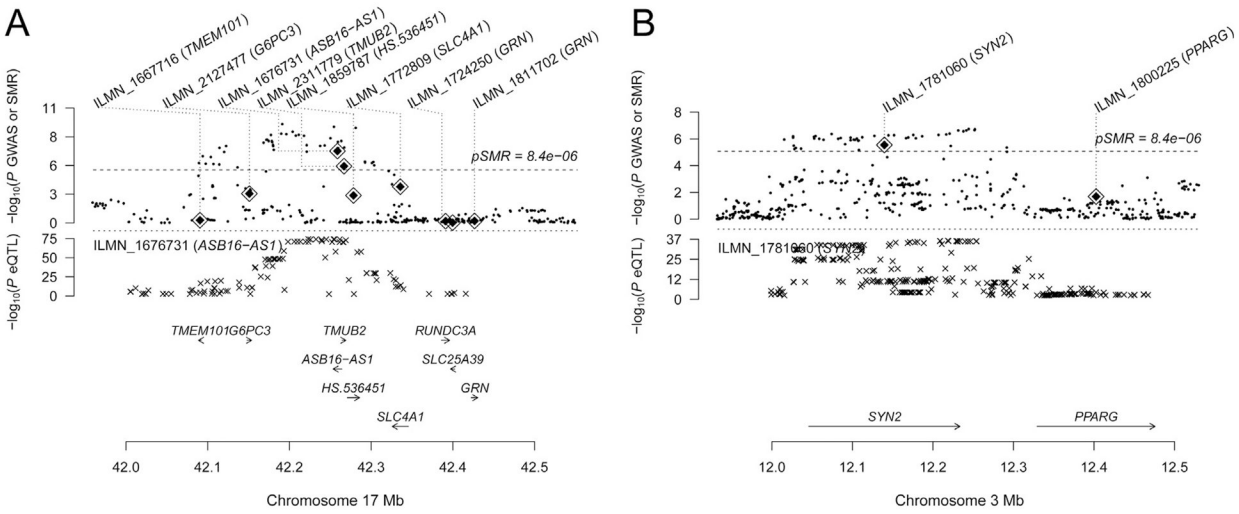

Fig. 1.

The SMR results at gene loci for BMD. (A) The SMR result at *ASB16-AS1* locus for FN-BMD. (B) The SMR result at *SYN2* locus for LS-BMD. In the top plot, black dots represent the p values for the SNPs from the latest GWAS meta-analysis for BMD (Y-axis), diamonds represent the p values for probes from the SMR test. In the bottom plot, the eQTL p values of the SNPs were from the eQTL study (Y-axis) for the ILMN_1676731 probe (or ILMN_1781060 probe) tagging *ASB16-AS1* (or *SYN2*).

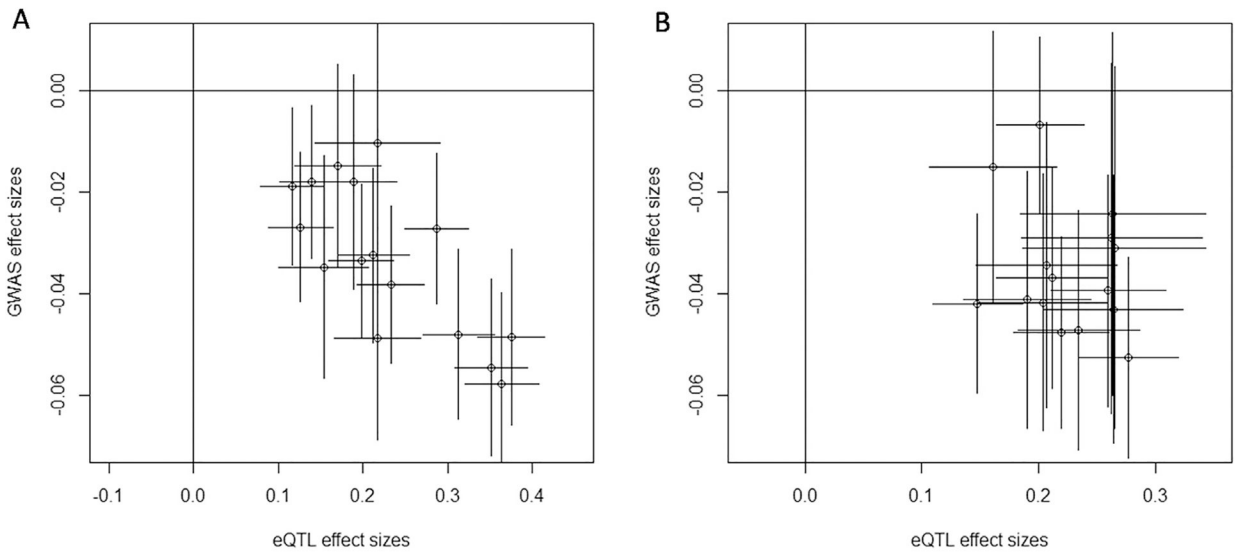

Fig. 2.

Estimated genetic associations and 95% confidence intervals with effect sizes in eQTL and GWAS studies for 16 genetic variants in the *ASB16-AS1* gene region (A) and 15 genetic variants in the *SYN2* gene region (B).

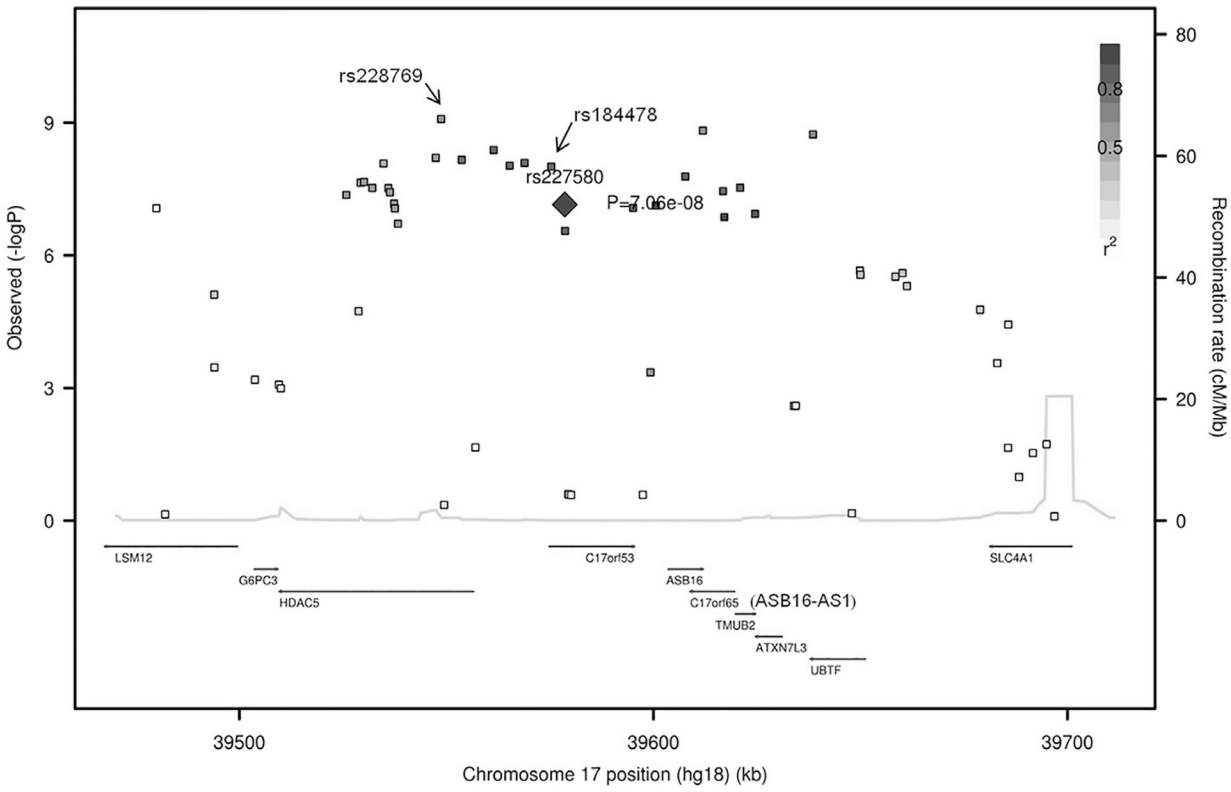

Fig. 3.

Regional association plot for *ASB16-AS1* on chromosome 17. SNPs which were in this region were selected with their p values from the GWAS data of FN-BMD. r^2 of pairwise LD is calculated between rs227580 and other SNPs. *C17ORF65* is also known as *ASB16-AS1.*

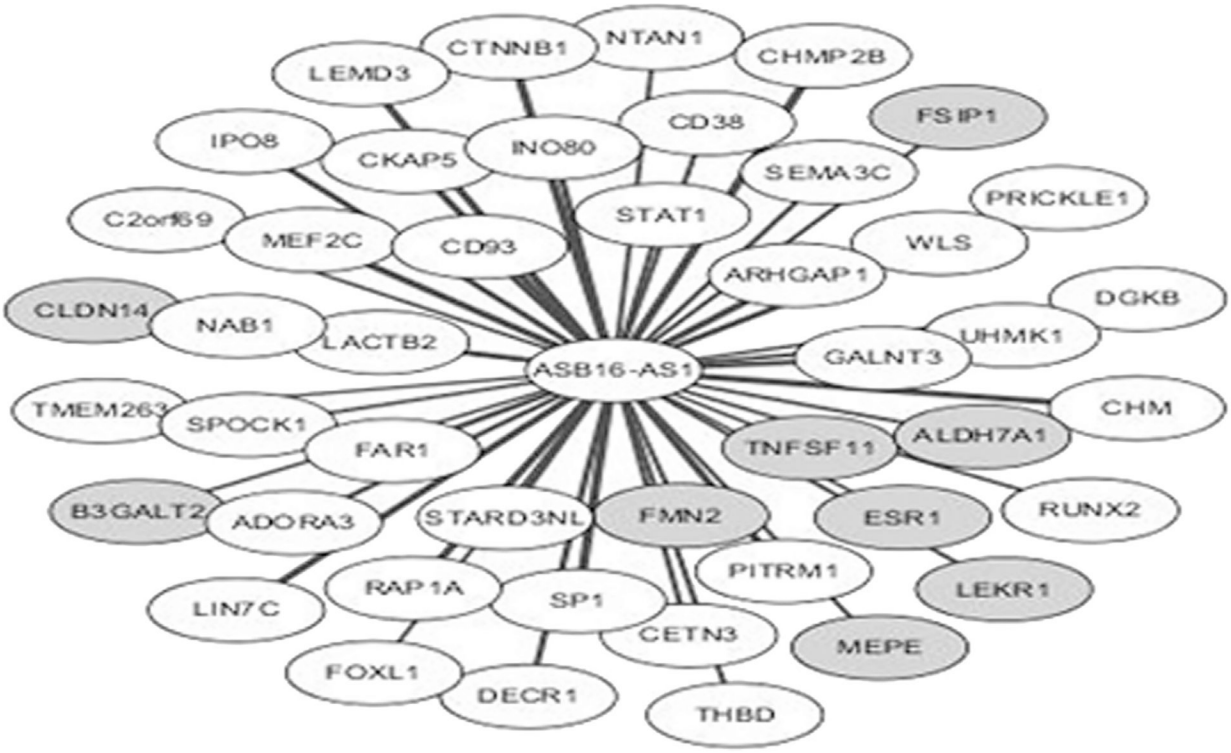

Fig. 4.

The gene co-expression relationships for *ASB16-AS1. ASB16-AS1* centered network provides a view of all edges and their corresponding nodes connected to *ASB16-AS1* with a TOM > 0.15. We only selected those nodes proved to be associated with BMD before. Genes are color coded based on their correlation with *ASB16-AS1*, white (cor > 0) and grey (cor < 0).

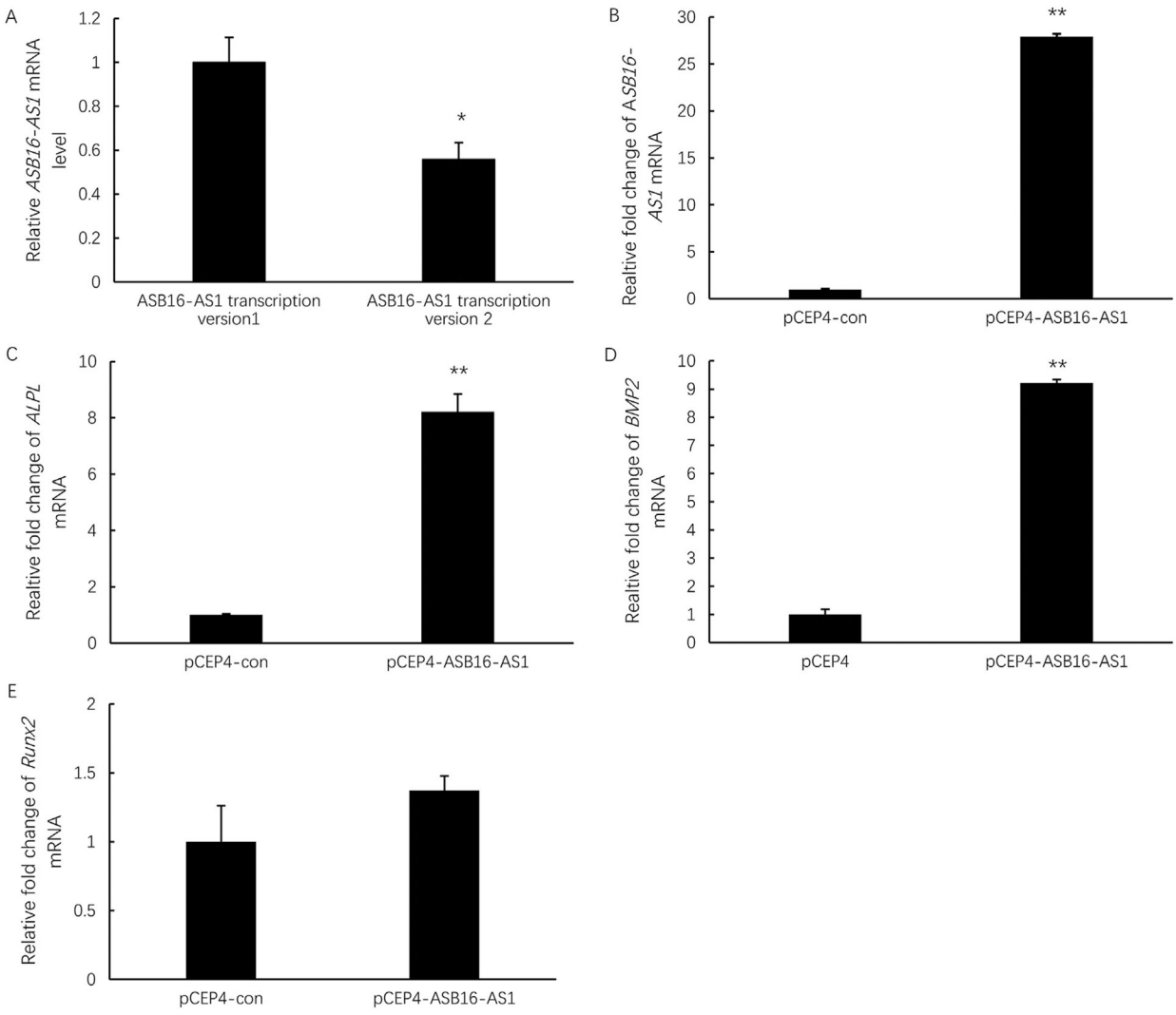

Fig. 5.

ASB16-AS1 can promote the expression of osteoblastic genes. (A) The relative expressions of two variants of *ASB16-AS1* were detected in hFOB1.19 cells. The relative expressions of *ASB16-AS1* (B), *ALPL* (C), *BMP2* (D) and *RUNX2* (E) were detected after transfection with the pCEP4-ASB16-AS1 using the pCEP4 as control. Bars represented S.D. *$p < 0.05$, **$p < 0.01$.

Table 1

Four genes identified by the SMR analyses.

Trait	Probe ID	Chr	Gene	cis-eQTL	P_{GWAS}	P_{eQTL}	b_{XY}	P_{SMR}	P_{HEIDI}	PICS
FN-BMD	ILMN_1676731	17	*ASB16-AS1*	rs227580	7.06×10^{-8}	4.66×10^{-75}	−0.13	1.30×10^{-7}	0.11	0.055
FN-BMD	ILMN_2311779	17	*TMUB2*	rs228754	4.20×10^{-6}	1.03×10^{-143}	−0.077	3.70×10^{-6}	0.007	NA
LS-BMD	ILMN_1781060	3	*SYN2*	rs310758	8.73×10^{-7}	8.18×10^{-37}	−0.18	2.83×10^{-6}	0.34	0.035
LS-BMD	ILMN_2127605	19	*LRP3*	rs11084710	6.99×10^{-6}	3.23×10^{-82}	0.11	7.55×10^{-6}	0.011	NA
FN-BMD	ILMN_1676731	17	*ASB16-AS1*	rs184478	9.34×10^{-9}	8.37×10^{-72}	−0.13	2.39×10^{-8}	0.11	0.99
LS-BMD	ILMN_1781060	3	*SYN2*	rs795000	3.58×10^{-7}	2.05×10^{-36}	−0.28	1.44×10^{-6}	0.34	1

Note: The first four lines were calculated with top cis-eQTLs which were the top associated cis-eQTL of the probes in the Westra et al. eQTL data. The last two lines were calculated with the possible causal SNPs in the region.

b_{XY} was the gene expression effect on BMD.

PICS was Probabilistic Identification of Causal SNPs.