



# Twelve New Genomic Loci Associated With Bone Mineral Density

Lu Liu<sup>1,2,3†</sup>, Min Zhao<sup>1,2†</sup>, Zong-Gang Xie<sup>4†</sup>, Ju Liu<sup>3</sup>, Hui-Ping Peng<sup>3</sup>, Yu-Fang Pei<sup>5\*‡</sup>, Hong-Peng Sun<sup>2,6\*‡</sup> and Lei Zhang<sup>1,2\*‡</sup>

<sup>1</sup> Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Jiangsu, China, <sup>2</sup> Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Jiangsu, China, <sup>3</sup> Kunshan Hospital of Traditional Chinese Medicine, Jiangsu, China, <sup>4</sup> The Second Affiliated Hospital of Soochow University, Jiangsu, China, <sup>5</sup> Department of Epidemiology and Health Statistics, School of Public Health, Medical College of Soochow University, Jiangsu, China, <sup>6</sup> Department of Children Health and Social Medicine, School of Public Health, Medical College of Soochow University, Jiangsu, China

## OPEN ACCESS

### Edited by:

Krasimira Tsaneva-Atanasova,  
University of Exeter, United Kingdom

### Reviewed by:

Elena Lopez-Isac,  
University of Manchester,  
United Kingdom  
David Karasik,  
Bar-Ilan University, Israel

### \*Correspondence:

Yu-Fang Pei  
ypei@suda.edu.cn  
Hong-Peng Sun  
hpsun@suda.edu.cn  
Lei Zhang  
lzhang6@suda.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

<sup>‡</sup>These authors have jointly supervised  
this study

### Specialty section:

This article was submitted to  
Systems Endocrinology,  
a section of the journal  
Frontiers in Endocrinology

**Received:** 16 January 2020

**Accepted:** 02 April 2020

**Published:** 22 April 2020

### Citation:

Liu L, Zhao M, Xie Z-G, Liu J,  
Peng H-P, Pei Y-F, Sun H-P and  
Zhang L (2020) Twelve New Genomic  
Loci Associated With Bone Mineral  
Density. *Front. Endocrinol.* 11:243.  
doi: 10.3389/fendo.2020.00243

Aiming to identify more genomic loci associated with bone mineral density (BMD), we conducted a joint association analysis of 2 genome-wide association study (GWAS) by the integrative association method multi-trait analysis of GWAS (MTAG). The first one is the single GWAS of estimated heel BMD (eBMD) in the UK biobank (UKB) cohort ( $N = 426,824$ ), and the second one is the GWAS meta-analysis of total body BMD (TB-BMD) in 66,628 participants from 30 studies. Approximate conditional association analysis was performed in the identified novel loci to identify secondary association signal. Statistical fine-mapping was conducted to prioritize plausible credible risk variants (CRVs). Candidate genes were prioritized based on the analyses of cis-expression quantitative trait locus (cis-eQTL) and cis-protein QTL (cis-pQTL) information as well as the functional category of the SNP. By integrating the information carried in over 490,000 participants, this largest joint analysis of BMD GWAS identified 12 novel genomic loci at the genome-wide significance level (GWS,  $p = 5.0 \times 10^{-8}$ ), nine of which were for eBMD and four were for TB-BMD, explaining an additional 0.11 and 0.23% heritability for the two traits, respectively. These loci include 1p33 (lead SNP rs10493130,  $p_{\text{eBMD}} = 3.19 \times 10^{-8}$ ), 5q13.2 (rs4703589,  $p_{\text{eBMD}} = 4.78 \times 10^{-8}$ ), 5q31.3 (rs9324887,  $p_{\text{TB-BMD}} = 1.36 \times 10^{-9}$ ), 6p21.32 (rs6905837,  $p_{\text{eBMD}} = 3.32 \times 10^{-8}$ ), 6q14.1 (rs10806234,  $p_{\text{eBMD}} = 2.63 \times 10^{-8}$ ), 7q21.11 (rs10806234,  $p_{\text{TB-BMD}} = 3.37 \times 10^{-8}$ ), 8q24.12 (rs11995866,  $p_{\text{eBMD}} = 6.72 \times 10^{-9}$ ), 12p13.31 (rs1639122,  $p_{\text{eBMD}} = 4.43 \times 10^{-8}$ ), 12p12.1 (rs58489179,  $p_{\text{eBMD}} = 4.74 \times 10^{-8}$ ), 12q24.23 (rs75499226,  $p_{\text{eBMD}} = 1.44 \times 10^{-8}$ ), 19q13.31 (rs7255083,  $p_{\text{TB-BMD}} = 2.18 \times 10^{-8}$ ) and 22q11.23 (rs13056137,  $p_{\text{TB-BMD}} = 2.54 \times 10^{-8}$ ). All lead SNPs in these 12 loci are nominally significant in both original studies as well as consistent in effect direction between them, providing solid evidence of replication. Approximate conditional analysis identified one secondary signal in 5q13.2 (rs11738874,  $p_{\text{conditional}} = 5.06 \times 10^{-9}$ ). Statistical fine-mapping analysis prioritized 269 CRVs. A total

of 65 candidate genes were prioritized, including those with known biological function to bone development (such as *FGF1*, *COL11A2* and *DEPTOR*). Our findings provide novel insights into a better understanding of the genetic mechanism underlying bone development as well as candidate genes for future functional investigation.

**Keywords:** bone mineral density, genome-wide association study, osteoporosis, joint analysis, MTAG

## INTRODUCTION

Osteoporosis is a common aging-related disease characterized by low bone mass and micro-architectural deterioration of bone tissue with a consequent increase in bone fragility and susceptibility to fracture (1). These later complications are associated with significant individual morbidity and related healthcare costs. Fifteen per cent of white people over 50 years old suffer osteoporotic fracture in their remaining lifetime, and the projected costs expended on this disease will exceed \$25 billion in the United States alone by year 2025 (2). Therefore, a better understanding of the mechanisms underlying osteoporosis may help to develop medications for osteoporosis prevention and treatment.

The diagnosis of osteoporosis is made from the measurement of bone mineral density (BMD), which is a highly heritable trait with heritability ranging from 50 to 80% (3). Previous genome-wide associations studies (GWAS) and their meta-analyses have identified over 500 genomic loci for BMD, accounting for up to 20% phenotypic variation (4–17). Nonetheless, compared with the total 36% of GWAS-attributable heritability (14), there is still a large portion of “missing” heritability to be discovered by enlarged GWAS or more efficient analysis.

In the present study, aiming to identify more genomic loci that are responsible for BMD variation, we conduct a joint analysis of 2 GWAS analyses. The first one is the largest single GWAS study of estimated heel BMD (eBMD) in 426,824 UK biobank (UKB) participants (17), and the second one is the GWAS meta-analysis of total body BMD (TB-BMD) in 66,628 participants (15), which is so far the second largest study for BMD. This analysis integrates the information from an expanded size of over 490,000 participants, and therefore has the potential maximal statistical power of gene mapping to date. We further prioritize plausible functional variants and candidate genes for future experimental validation.

## MATERIALS AND METHODS

We performed a joint analysis of summary statistics from two large-scale BMD GWAS studies. No new ethnic approval was required.

### Study Samples

Two studies were incorporated into this joint association analysis. The first one is the single GWAS of eBMD in the UKB cohort ( $N = 426,824$ ). In brief, the UKB sample is a large prospective cohort study of ~500,000 participants from across the United Kingdom, aged between 40 and 69 at recruitment.

Heel BMD was estimated based on quantitative ultrasound speed of sound (SOS) and broadband ultrasound attenuation (BUA). Genome-wide genotypes were available for all participants at 784,256 genotyped autosome markers, and were imputed into UK10K haplotype, 1000 Genomes project phase 3 and Haplotype Reference Consortium (HRC) reference panels by IMPUTE2 (18). After quality controls, 426,824 qualified participants were used in GWAS analysis (17).

The second one is the GWAS meta-analysis of TB-BMD in 66,628 participants from 30 studies (15). In brief, TB-BMD was measured by dual energy X-ray absorptiometry (DXA). All participants had genome-wide genotype data and were imputed into the 1,000 Genomes project phase 1 or the combined 1,000 Genomes project phase 3 and the UK10K reference panel. Almost all popular imputation methods were used across studies (**Supplementary Table 1**). In their analyses, association was performed in each individual study, and the summary statistics across the 30 studies were meta-analyzed by a fixed-effects model.

A total of 1,553 participants from the UKB cohort were included in the TB-BMD study, accounting for 2.3 and 0.4% of TB-BMD and eBMD sample sizes, respectively. Association summary statistics for both studies were publicly available at the genetic factors for osteoporosis consortium (GEFOS) website (<http://www.gefos.org>), and were downloaded for analysis.

### SNP Inclusion Criteria

The SNP inclusion criteria were same as described previously (19). In brief, in each study, non-SNP variants and ambiguous SNPs (i.e., multiple SNPs corresponding to one single identifier) were excluded. In addition, SNPs presenting significant meta-analysis genetic heterogeneity ( $I^2 > 50\%$  or  $Q$  p-value  $< 0.1$ ) in the TB-BMD study were excluded. After quality control (QC), 8,680,009 SNPs are available in both studies. After removing SNPs that are not concordant between the two studies (i.e., A/G vs. T/G polymorphisms), a total of 7,369,691 SNPs in common to both studies were used in subsequent analyses.

### Joint Association Analysis

The recently developed multi-trait analysis of GWAS (MTAG) method was used for joint association analysis, which accounts for sample overlap and trait heterogeneity between studies (20). In brief, MTAG estimates per SNP effect size for each trait by incorporating information contained in other correlated traits, and therefore has potential to improve statistical power of association test. MTAG takes summary statistics from multiple studies as input. The output of MTAG contains re-estimated effect size and  $p$ -value for each SNP in each trait.

The linkage disequilibrium score regression (LDSC) method was applied to the MTAG results to estimate the amount of

genomic inflation due to confounding factors such as population stratification and cryptic relatedness (21). LDSC takes GWAS summary statistics as input and partitions overall inflated association statistic into one part attributable to polygenic architecture and another part due to population stratification and cryptic relatedness. Reference LD scores for the European population were downloaded from the software website (<https://github.com/bulik/ldsc>). The relative contribution of confounding factors was measured by attenuation ratio (AR), which is defined as  $(\text{intercept}-1)/(\text{mean } \chi^2 - 1)$ , where intercept and mean  $\chi^2$  are estimates of confounding and the overall association inflation, respectively (21).

Genome-wide significance (GWS) level was set to be  $5.0 \times 10^{-8}$ . An independent locus was defined as one 1-MB region, which consists of two 500 kb regions from the lead SNP to both directions. A novel locus was declared if it was neither reported in previous GWAS studies nor significant at the GWS level in eBMD or TB-BMD study.

Individual variant effect size was estimated with the formula  $2f(1-f)\beta^2$ , where  $f$  is allele frequency and  $\beta$  is regression coefficient, which was estimated by MTAG.

### Approximate Conditional Analysis

To identify additional signals in regions of association, approximate conditional association analysis was performed in each region using the genome-wide complex trait analysis (GCTA) tool (22). A reference sample of 100,000 unrelated participants from the UKB cohort was generated for estimating LD pattern. Specifically, a total of 369,968 unrelated participants were inferred with kinship-based inference for GWAS (KING) software (23), from whom the 100,000 participants of the reference sample were randomly drawn.

A recursive conditional association analysis was performed. In each iteration, an approximate conditional analysis conditioning on the current list of lead variants was performed. A secondary significant variant was defined at the conditional GWS level (conditional  $p < 5 \times 10^{-8}$ ). The variant with the smallest  $p$ -value among such identified ones was added into the list of lead variants. Iterations of the conditional analysis were run until no significant signal can be identified.

### Statistical Fine-Mapping of Credible Risk Variants

Statistical fine-mapping of credible risk variants (CRVs) was performed with FINEMAP (24). FINEMAP uses GWAS summary statistics and applies a shotgun stochastic search algorithm to efficiently exploring a set of most important causal configurations in the associated region. It relies on a matched reference panel for LD estimation. Again, the above reference panel of 100,000 unrelated UKB participants was used for LD estimation with LDstore (25). Software parameters were set by default. The outputs of FINEMAP include the posterior probability of each SNP being causal. For each locus, we sorted the posterior probabilities in an descending order, and constructed the set of CRVs by including those SNPs whose posterior probabilities were within one order of magnitude of the largest posterior probability.

### Candidate Gene Prioritization

We prioritized candidate genes in the identified novel loci by annotating the CRVs for their cis-expression quantitative trait locus (cis-eQTL) and cis-protein QTL (cis-pQTL) effects, and for their distances to genes.

Cis-eQTL effect was assessed from two datasets. The first one is the 44 tissues from the GTEx project (v6) (26). Pre-compiled cis-eQTL results were downloaded from the GTEx web portal ([www.gtexportal.org/](http://www.gtexportal.org/)). The distance between the SNP and the transcription starting site (TSS) of the target gene was assumed to be <500 kb. Assuming a maximal number of 5,000 independent variants over such a 1-MB region, significant cis-eQTL was declared at  $p < 1.0 \times 10^{-5}$  (0.05/5,000). The second one is the lymphoblastoid cell lines of 373 European individuals from the 1,000 genomes project (27). Pre-compiled cis-eQTL results were downloaded from the study website (<https://www.ebi.ac.uk/Tools/geuvadis-das/>). Significant cis-eQTL was declared under the same criteria.

Cis-pQTL information was accessed from Sun et al. (28). In that largest proteome study to date, the authors measured plasma protein levels of 3,301 healthy individuals using the SOMAscan platform (SomaLogic, Inc., Boulder, Colorado, USA) comprising 4,034 distinct aptamers (SOMAmers) covering 3,623 proteins. GWAS summary statistics for 3,284 proteins were downloaded from the study's website. Cis-pQTL was searched within 500 kb distance from a target gene. Analogously to the cis-eQTL analysis, significant cis-pQTL was declared at  $p < 1.0 \times 10^{-5}$ .

SNPs were annotated by variant effect predictor (VEP) for their functional category (29). SNP-gene distance was annotated by prioritization of candidate causal genes at molecular QTLs (ProGeM) software (30). Genes nearest, second nearest and third nearest to each lead SNP were listed.

## RESULTS

### Main Association Results

There are 13,753,401 and 18,259,434 SNPs in the eBMD and TB-BMD studies. After QC, 7,369,691 SNPs present in both studies were qualified for analysis. Genetic correlation coefficient was reported to be 0.57 (19), implying shared BMD heritability between the two studies. In the MTAG analysis of the eBMD study, the intercept and mean chi-square are 1.09 and 3.30, respectively, suggesting that 96.1% of the inflation in the mean chi-squared statistic is from polygenic architecture rather than from population stratification. Similarly, the intercept and mean chi-square in the MTAG analysis of the TB-BMD study are 0.78 and 1.84, respectively, suggesting that most of the inflation is from polygenic architecture.

The original UKB eBMD study summary statistics contain 76,400 SNPs significant at the GWS level, encompassing 659 distinct loci. In the MTAG analysis, all the 659 lead SNPs are consistent in effect direction with original ones. Among them, 556 (84.4%) remain significant at the GWS level. Though the other 103 lead SNPs become non-significant at the GWS level, they are all nominally significant ( $p < 0.05$ ) with  $p$ -value ranging from  $5.02 \times 10^{-8}$  to  $8.88 \times 10^{-5}$ . Of the 556 GWS significant

TABLE 1 | Main results of the identified novel loci.

SNP	CHR	BP	EA/OA	EAF	eBMD (N = 426,824)					TB-BMD						
					Original			MTAG		Original			MTAG			
					BETA	SE	P	BETA	SE	P	BETA	SE	P	BETA	SE	P
rs10498130	1	48,341,005	G/A	0.68	0.002	4.20E-7	-0.01	0.003	<b>3.19E-8</b>	-0.02	0.006	5.85E-3	66,546	-0.02	0.004	4.97E-6
rs4703559	5	72,097,351	T/C	0.47	0.002	2.20E-6	0.01	0.002	<b>4.78E-8</b>	0.02	0.006	9.67E-4	66,604	0.02	0.004	1.32E-6
rs9324887	5	142,047,831	A/G	0.72	0.002	2.00E-5	-0.02	0.003	<b>2.20E-8</b>	-0.03	0.006	1.57E-7	66,480	-0.03	0.004	<b>1.36E-9</b>
rs6905837	6	32,626,205	C/T	0.95	0.004	2.00E-7	-0.03	0.006	<b>3.32E-8</b>	-0.05	0.019	4.21E-3	37,410	-0.05	0.010	2.89E-6
rs10806234	6	82,685,299	A/G	0.40	0.002	4.70E-6	-0.01	0.002	<b>2.63E-8</b>	-0.02	0.006	3.78E-5	66,048	-0.02	0.004	6.51E-8
rs1019203	7	84,784,053	C/T	0.64	0.002	3.60E-5	0.01	0.003	8.58E-8	0.03	0.006	5.21E-6	66,553	0.02	0.004	<b>3.37E-8</b>
rs11995866	8	121,058,098	G/A	0.61	0.002	6.80E-7	-0.01	0.002	<b>6.72E-9</b>	-0.01	0.006	0.04	66,569	-0.02	0.004	1.94E-5
rs1639122	12	6,711,147	C/A	0.56	0.002	6.30E-7	-0.01	0.002	<b>4.43E-8</b>	-0.02	0.006	7.85E-3	66,038	-0.02	0.004	7.27E-6
rs58489179	12	25,525,053	G/A	0.91	0.003	1.10E-7	0.02	0.004	<b>4.74E-8</b>	0.04	0.012	2.32E-3	63,629	0.03	0.007	2.59E-6
rs75499226	12	120,515,773	C/A	0.91	0.003	7.60E-7	0.02	0.004	<b>1.44E-8</b>	0.03	0.009	4.85E-3	66,572	0.03	0.006	2.96E-6
rs7255083	19	44,337,803	T/C	0.57	0.002	7.30E-6	-0.01	0.002	1.77E-7	-0.03	0.006	2.08E-6	66,595	-0.02	0.004	<b>2.18E-8</b>
rs13056137	22	23,407,261	C/A	0.74	0.002	4.60E-6	0.01	0.003	7.01E-8	0.03	0.006	5.23E-6	66,419	0.02	0.004	<b>2.54E-8</b>

CHR, chromosome; BP, base-pair, based on GRCh37 genome assembly; EA, effect allele; OA, other allele; EAF, effect allele frequency that was derived from the UKB study summary results; BETA, regression coefficient; SE, standard error of BETA; N, sample size; P, p-value. P-values less than the genome-wide significance (GWS,  $5.0 \times 10^{-8}$ ) were marked in bold.

SNPs, p-value at 193 (34.7%) SNPs gets smaller while those at 363 SNPs gets higher.

The original GEFOS TB-BMD study summary statistics contain 3,842 SNPs significant at the GWS level, encompassing 68 distinct loci. In the MTAG analysis, all the 68 lead SNPs are consistent in effect direction with original ones. Among them, 59 (86.8%) remain significant at the GWS level, while the other nine become non-significant at the GWS level. Only one SNP rs1037011 becomes non-significant even at the loosely nominal significance level ( $p = 0.07$ ). Interestingly, this SNP is GWS significant in both original studies ( $p_{eBMD} = 3.40 \times 10^{-9}$ ,  $p_{TB-BMD} = 1.54 \times 10^{-12}$ ), but the effect allele T has opposite direction ( $\beta_{eBMD} = 0.01$ ,  $\beta_{TB-BMD} = -0.04$ ). The possible reason for the opposite directions could be strand alignment error or true opposite genetic effects, pending further investigation. Of the 59 GWS significant SNPs, p-value at up to 53 SNPs gets smaller while that at the remaining 6 SNPs gets higher.

To search for additional loci, we evaluated MTAG results of all SNPs excluding those within 500 kb of a locus that was either GWS significant in either original study or was reported previously. This identified 225 SNPs in the eBMD study and 15 SNPs in the TB-BMD study. To increase the confidence of association results, only SNPs of p-value < 0.05 in both original studies were kept, resulting in 79 SNPs in the eBMD study and again 15 SNPs in the TB-BMD, encompassing nine and four distinct loci, respectively. Only one locus overlapped between the two studies, resulting in a total number of 12 distinct novel loci. All the 12 lead SNPs are consistent in effect direction in both original studies. Manhattan plot is displayed in Figure 1 and main results are listed in Table 1. Regional plots at all the novel loci are displayed in Supplementary Figure 1. The 12 lead SNPs collectively explain 0.11 and 0.23% phenotypic variance for eBMD and TB-BMD, respectively.

In Supplementary Table 2, we listed genomic distance and LD value ( $r^2$ ) between the lead SNP at each of the identified novel loci and the nearest GWS SNP at the nearest known locus. The results show that all pairs of SNPs are in complete linkage equilibrium, confirming the novelty of the identified loci.

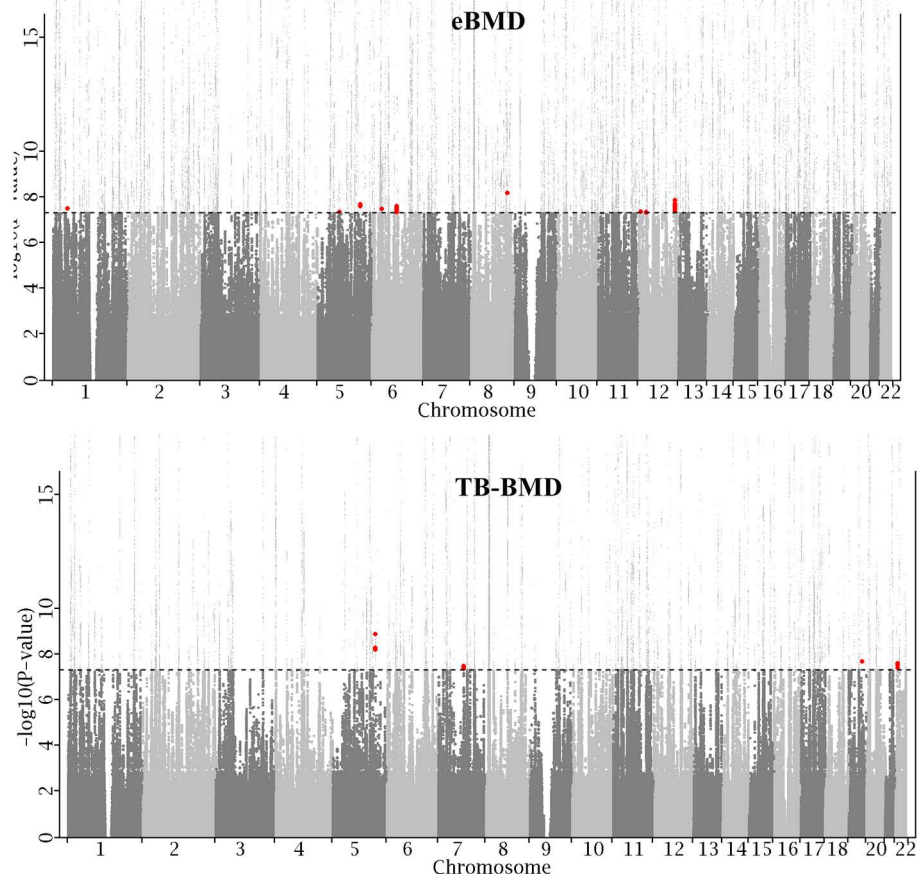
### Approximate Conditional Analysis

Using GCTA, we performed an approximate conditional analysis. In the eBMD analysis, one secondary signal was identified at rs11738874 ( $p_{MTAG} = 5.38 \times 10^{-8}$ ,  $p_{conditional} = 5.06 \times 10^{-9}$ ). This SNP is significant in both original studies ( $p_{eBMD} = 7.60 \times 10^{-6}$ ,  $p_{TB-BMD} = 7.30 \times 10^{-3}$ ). It is in linkage equilibrium with the lead SNP rs4703589 (LD  $r^2 = 0.0$ ), indicating an independent association signal.

No secondary associations were found in the TB-BMD analysis.

### Credible Risk Variants

With FINEMAP, we performed a statistical fine-mapping analysis. In the eBMD analysis, 214 CRVs were prioritized at the 12 novel loci, with an average of 18 variants per locus. The locus with most number of variants is 12q24.23, in which up to 38 variants were prioritized. There is one locus at 19q13.31 with only one causal variant being prioritized, which is the lead



**FIGURE 1** | Manhattan plot. X-axis represents genomic position on chromosome and Y-axis represents  $-\log_{10}(P\text{-value})$ . For ease of presentation, Y-axis is truncated at 15. Dotted line represents genome-wide significance level (GWS,  $5.0 \times 10^{-8}$ ). All known loci, including GWS loci in either eBMD (top) or TB-BMD (bottom) study, and previously reported loci that were retrieved from the EBI GWAS catalog, are plotted in light gray. Newly identified loci are plotted in red.

SNP rs7255083. Its posterior probability is up to 0.87, being 15-fold larger than that of the second variant rs2356401 (posterior probability 0.06) (**Supplementary Table 3**).

In the TB-BMD analysis, 64 variants are prioritized at 12 loci, with an average of 5 variants per locus. The locus with most number of variants is again 12q24.23, in which 7 variants are prioritized. No locus is found to have only one causal variant (**Supplementary Table 3**).

The two sets of CRVs derived from both analyses were merged into one single set of 269 CRVs, where 9 CRVs overlap between the two analyses.

### Newly Identified Loci/Genes

We next prioritized candidate genes based on the annotations of cis-eQTL and cis-pQTL effects and functional categories in the above set of CRVs.

In the 44 tissues of the GTEx (v6) dataset, 116 CRVs from 10 regions are cis-associated with the expression of nearby genes at a variety of tissues. The two regions with no evidence of cis-eQTL effect are 6q14.1 and 7q21.11 (**Supplementary Table 4**).

In the lymphoblastoid cell lines, 12 variants from four regions are associated with nearby gene expressions, where seven variants overlapped with the GTEx variants. There are three associated genes that are not found in the list of GTEx prioritized genes (**Supplementary Table 4**).

In the pQTL dataset, multiple SNPs at 12p13.31 are cis-associated with *TAPBPL* (TAP Binding Protein Like) protein level. The most significant association is observed at rs1639122 ( $p = 3.31 \times 10^{-43}$ ), which is a mis-sense mutation in a nearby gene *CHD4* (Chromodomain Helicase DNA Binding Protein 4). Another SNP rs11609726 at this region is also associated with one additional protein *C1RL* (Complement C1r Subcomponent Like,  $p = 4.36 \times 10^{-8}$ ) (**Supplementary Table 5**).

Combining the various annotations, a total of 65 candidate genes were prioritized at the 12 novel loci (**Table 2**).

### DISCUSSION

By conducting a joint association study of 2 large-scale GWAS analyses, the present study integrated the information contained

**TABLE 2** | Prioritized candidate genes at the identified novel loci.

Locus	Lead SNP	CHR	BP	Gene
1p33	rs10493130	1	48,341,005	SLC5A9(E), SKINTL(E), SPATA6(E), TRABD2B(N)
5q13.2	rs4703589	5	72,097,351	MAP1B(E), FCHO2(E,S), TNPO1(N), ZNF366(T)
5q31.3	rs9324887	5	142,047,831	FGF1(E,N), ARHGAP26(S), SPRY4(T)
6p21.32	rs6905837	6	32,626,205	CYP21A1P(E), HCG23(E), TNXA(E), PRRT1(E), PSMB9(E), TAP1(E), COL11A2(E), RPL32P1(E), HLA-DQB1(N), HLA-DQA1(S), HLA-DRB1(T)
6q14.1	rs10806234	6	82,685,299	IBTK(N), FAM46A(S), TPBG(T)
7q21.11	rs1019203	7	84,784,053	SEMA3D(N)
8q24.12	rs11995866	8	121,058,098	DSCC1(E), DEPTOR(E,N), COL14A1(S), DSCC1(T)
12p13.31	rs1639122	12	6,711,147	TAPBPL(E,P), MRPL51(E), LRRC23(E), CHD4(E,N), C1RL(P), LPAR5(S), NOP2(T)
12p12.1	rs58489179	12	25,525,053	LRMP(E), CASC1(E), IFLTD1(N), KRAS(S)
12q24.23	rs75499226	12	120,515,773	PRKAB1(E), TMEM233(E), CCDC64(E,N), COQ5(E), POP5(E), GATC(E), RAB35(S), GCN1L1(T)
19q13.31	rs7255083	19	44,337,803	ZNF575(E), XRCC1(E), PHLDB3(E), LYPD3(E), PINLYP(E), ZNF283(E,N), ZNF155(E), ZNF223(E), ZNF404(E,T), ZNF45(E), LYPD5(S)
22q11.23	rs13056137	22	23,407,261	BCR(E), RAB36(E,T), GNAZ(E,S), FBXW4P1(E), RTDR1(N)

Genes were prioritized as following: gene nearest to the lead SNP (N); gene second nearest to the lead SNP (S); gene third nearest to the lead SNP (T); gene with mRNA level in association with one or more CRVs (cis-eQTL, E); gene with protein level in association with one or more CRVs (cis-pQTL, P).

in over 490,000 participants, making it the largest BMD GWAS analysis to date. As a result, 12 new genomic loci were identified, demonstrating the enhanced statistical power.

Because the two integrated traits are not same, naive meta-analysis is not applicable. Instead, we used the recently developed method MTAG for integration analysis. MTAG is applicable to analyzing genetically correlated traits. It uses two sources of information to integrate association signals. The first one is that the true effect of target SNP is correlated across traits, and the second one is that the estimation error of the SNP' effects is correlated across traits (20). When two GWAS summary statistics datasets have overlapped individuals, for example, shared control individuals, MTAG is indeed capable of handling this relatedness. It accomplishes this by applying the bivariate LD score regression to estimating the correlation in GWAS estimation error due to sample overlap. In our analysis, 1,533 participants overlapped between the eBMD and the TB-BMD studies, accounting for 0.36% of the total UKB participants. This tiny portion of overlap is not expected to have a major effect on the results, let alone that MTAG analysis took the overlap into account.

We were able to identify only one bone-related dataset, which is the lymphoblastoid cell lines. Only 12 CRVs from this dataset were found to exert cis-eQTL activity. The use of the 44 GTEx tissues was not motivated by their specific relationship to BMD, but was by a maximal coverage of available cis-eQTL SNPs. Because while some cis-eQTL activities are tissue-specific, some others are common across tissues. For example, Of the 12 cis-eQTL variants identified from the lymphoblastoid cell lines, seven overlapped with the GTEx cis-eQTL variants, implying that a considerable portion of cis-eQTL sites were common across tissues.

The prioritized candidate genes include those being linked to bone biology in previous literature. At 5q31.3, for example, the lead SNP rs9324887 is located in the intron of *FGF1* (Fibroblast Growth Factor-1) gene and is associated with its

expression. *FGF1* is a member of the *FGF* signaling pathway that participates in the regulation of bone development (31). Local and systemic *FGF1* increases new bone formation and bone density. It also appears to restore bone micro-architecture and prevent bone loss associated with estrogen-withdrawal (32). At 6p21.32, *COL11A2* (Collagen Type XI Alpha 2 Chain) is one of the prioritized genes by cis-eQTL analysis. It is a member of the collagen family of extracellular proteins. It is a critical positive factor in the regulation of extra cellular matrix (ECM), which mineralizes to bone (33, 34). At 8q24.12, two genes *DSCC1* (DNA Replication And Sister Chromatid Cohesion 1) and *DEPTOR* (DEP Domain Containing MTOR Interacting Protein) were prioritized. It was only recently that *DEPTOR* was found to play a novel function in osteogenic differentiation by inhibiting *MEG3*-mediated activation of *BMP4* signaling, suggesting its involvement in osteoporosis (35).

In addition to those established candidate genes to bone biology, we also prioritized multiple candidate genes with no known function to bone metabolism. Among them, *SEMA3D* (Semaphorin 3D) at 7q21.11 belongs to a member of the semaphorin III family of secreted signaling proteins that are involved in axon guidance during neuronal development (36). Despite the lack of evidence for its involvement in bone development, one of its paralogs, *SEMA3A* (Semaphorin 3A), is found to regulate bone remodeling indirectly by modulating sensory nerve development instead of by acting on osteoblasts (37). In addition, *SEMA3A* and *SEMA3E* (Semaphorin 3E), are also reported to be associated with hypogonadotropic hypogonadism (38, 39). Hypogonadotropic hypogonadism is known to regulate bone density (40). These lines of evidence imply *SEMA3D* may have a regulatory role on bone development.

In conclusion, by conducting a joint analysis of two large-scale genome-wide association study and meta-analysis, we have

identified 12 novel loci associated with BMD. Our findings provide candidate genes for future functional investigations and for a better understanding of the genetic mechanism underlying bone development.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.gefos.org/?q=content/gefos-lifecourse-tb-bmd-gwas-results>, <http://www.gefos.org/?q=content/data-release-2018>.

## AUTHOR CONTRIBUTIONS

LZ was in charge of conceptualization and design. LL, MZ, and Z-GX contributed to data collection, formal analysis, and manuscript drafting. JL and H-PP contributed to the literature review and manuscript editing. Y-FP, H-PS, and LZ are responsible for supervision

and funding. All authors approved the final version to be published.

## ACKNOWLEDGMENTS

We are grateful to the GEFOS consortium and the eBMD study for releasing the large-scale GWAS summary statistics that were used in this study. LZ and Y-FP are partially supported by the national natural science foundation of China (31571291, 31771417 and 31501026) and a project of the priority academic program development (PAPD) of Jiangsu higher education institutions. The numerical calculations in this paper have been done on the supercomputing system of the National Supercomputing Center in Changsha.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2020.00243/full#supplementary-material>

## REFERENCES

- Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet*. (2002) 359:1761–7. doi: 10.1016/S0140-6736(02)08657-9
- Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005–2025. *J Bone Miner Res*. (2007) 22:465–75. doi: 10.1359/jbmr.061113
- Peacock M, Turner CH, Econs MJ, Foroud T. Genetics of osteoporosis. *Endocr Rev*. (2002) 23:303–26. doi: 10.1210/edrv.23.3.0464
- Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet*. (2008) 371:1505–12. doi: 10.1016/S0140-6736(08)60599-1
- Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, et al. Multiple genetic loci for bone mineral density and fractures. *N. Engl. J. Med.* (2008) 358:2355–65. doi: 10.1056/NEJMoa0801197
- Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet*. (2009) 41:1199–206. doi: 10.1016/j.bone.2009.03.085
- Xiong DH, Liu XG, Guo YF, Tan LJ, Wang L, Sha BY, et al. Genome-wide association and follow-up replication studies identified ADAMTS18 and TGFBR3 as bone mass candidate genes in different ethnic groups. *Am J Hum Genet*. (2009) 84:388–98. doi: 10.1016/j.ajhg.2009.01.025
- Duncan EL, Danoy P, Kemp JP, Leo PJ, McCloskey E, Nicholson GC, 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. doi: 10.1371/journal.pgen.1001372
- Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, 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. doi: 10.1038/ng.2249
- Zhang L, Choi HJ, Estrada K, Leo PJ, Li J, Pei YF, et al. Multistage genome-wide association meta-analyses identified two new loci for bone mineral density. *Hum Mol Genet*. (2014) 23:1923–33. doi: 10.1093/hmg/ddt575
- Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*. (2015) 526:112–7. doi: 10.1038/nature14878
- Pei YF, Hu WZ, Hai R, Wang XY, Ran S, Lin Y, et al. Genome-wide association meta-analyses identified 1q43 and 2q32.2 for hip Ward's triangle areal bone mineral density. *Bone*. (2016) 91:1–10. doi: 10.1016/j.bone.2016.07.004
- Pei YF, Xie ZG, Wang XY, Hu WZ, Li LB, Ran S, et al. Association of 3q13.32 variants with hip trochanter and intertrochanter bone mineral density identified by a genome-wide association study. *Osteoporos Int*. (2016) 27:3343–54. doi: 10.1007/s00198-016-3663-y
- Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youlten SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet*. (2017) 49:1468–75. doi: 10.1038/ng.3949
- Medina-Gomez C, Kemp JP, Trajanoska K, Luan J, Chesni A, Ahluwalia TS, et al. Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. *Am J Hum Genet*. (2018) 102:88–102. doi: 10.1016/j.ajhg.2017.12.005
- Pei YF, Hu WZ, Yan MW, Li CW, Liu L, Yang XL, et al. Joint study of two genome-wide association meta-analyses identified 20p12.1 and 20q13.33 for bone mineral density. *Bone*. (2018) 110:378–85. doi: 10.1016/j.bone.2018.02.027
- Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet*. (2019) 51:258–66. doi: 10.1038/s41588-018-0302-x
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. (2009) 5:e1000529. doi: 10.1371/journal.pgen.1000529
- Pei YF, Liu L, Liu TL, Yang XL, Zhang H, Wei XT, et al. Joint association analysis identified 18 new loci for bone mineral density. *J Bone Miner Res*. (2019) 34:1086–94. doi: 10.1002/jbmr.3681
- Turley P, Walters RK, Maghziyan O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. (2018) 50:229–37. doi: 10.1038/s41588-017-0009-4
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. (2015) 47:291–5. doi: 10.1038/ng.3211
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. (2011) 88:76–82. doi: 10.1016/j.ajhg.2010.11.011
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. (2010) 26:2867–73. doi: 10.1093/bioinformatics/btq559

24. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*. (2016) 32:1493–501. doi: 10.1093/bioinformatics/btw018
25. Benner C, Havulinna AS, Jarvelin MR, Salomaa V, Ripatti S, Pirinen M. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am J Hum Genet*. (2017) 101:539–51. doi: 10.1016/j.ajhg.2017.08.012
26. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. (2013) 45:580–5. doi: 10.1038/ng.2653
27. Lappalainen T, Sammeth M, Friedlander MR, Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. (2013) 501:506–11. doi: 10.1038/nature12531
28. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature*. (2018) 558:73–9. doi: 10.1038/s41586-018-0175-2
29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol*. (2016) 17:122. doi: 10.1186/s13059-016-0974-4
30. Stacey D, Fauman EB, Ziemek D, Sun BB, Harshfield EL, Wood AM, et al. ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res*. (2019) 47:e3. doi: 10.1093/nar/gky837
31. Sahn M, Ambrosetti DC, Mansukhani A, Gertner R, Levy D, Basilico C. FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes Dev*. (1999) 13:1361–6. doi: 10.1101/gad.13.11.1361
32. Dunstan CR, Boyce R, Boyce BF, Garrett IR, Izbicka E, Burgess WH, et al. Systemic administration of acidic fibroblast growth factor (FGF-1) prevents bone loss and increases new bone formation in ovariectomized rats. *J Bone Miner Res*. (1999) 14:953–9. doi: 10.1359/jbmr.1999.14.6.953
33. Okazaki K, Sandell LJ. Extracellular matrix gene regulation. *Clin Orthop Relat Res*. (2004) S123–8. doi: 10.1097/01.blo.0000144478.51284.f3
34. Goto T, Matsui Y, Fernandes RJ, Hanson DA, Kubo T, Yukata K, et al. Sp1 family of transcription factors regulates the human alpha2 (XI) collagen gene (COL11A2) in Saos-2 osteoblastic cells. *J Bone Miner Res*. (2006) 21:661–73. doi: 10.1359/jbmr.020605
35. Chen S, Jia L, Zhang S, Zheng Y, Zhou Y. DEPTOR regulates osteogenic differentiation via inhibiting MEG3-mediated activation of BMP4 signaling and is involved in osteoporosis. *Stem Cell Res Ther*. (2018) 9:185. doi: 10.1186/s13287-018-0935-9
36. Liu Y, Halloran MC. Central and peripheral axon branches from one neuron are guided differentially by Semaphorin3D and transient axonal glycoprotein-1. *J Neurosci*. (2005) 25:10556–63. doi: 10.1523/JNEUROSCI.2710-05.2005
37. Fukuda T, Takeda S, Xu R, Ochi H, Sunamura S, Sato T, et al. Sema3A regulates bone-mass accrual through sensory innervations. *Nature*. (2013) 497:490–3. doi: 10.1038/nature12115
38. Cariboni A, Andre V, Chauvet S, Cassatella D, Davidson K, Caramello A, et al. Dysfunctional SEMA3E signaling underlies gonadotropin-releasing hormone neuron deficiency in Kallmann syndrome. *J Clin Invest*. (2015) 125:2413–28. doi: 10.1172/JCI78448
39. Cariboni A, Davidson K, Rakic S, Maggi R, Parnavelas JG, Ruhrberg C. Defective gonadotropin-releasing hormone neuron migration in mice lacking SEMA3A signalling through NRP1 and NRP2: implications for the aetiology of hypogonadotropic hypogonadism. *Hum Mol Genet*. (2011) 20:336–44. doi: 10.1093/hmg/ddq468
40. Finkelstein JS, Klibanski A, Neer RM, Doppelt SH, Rosenthal DI, Segre GV Jr, et al. Increases in bone density during treatment of men with idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. (1989) 69:776–83. doi: 10.1210/jcem-69-4-776

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Zhao, Xie, Liu, Peng, Pei, Sun and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.