Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.ajhg.2017.12.005

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
American Journal of Human Genetics

Publisher Rights Statement:
This is the author’s peer-reviewed manuscript as accepted for publication.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Jul. 2019
Life-course Genome-Wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-specific Effects


1) Department of Internal Medicine, Erasmus MC, 3000 CA Rotterdam, The Netherlands; 2) The Generation R Study Group, Erasmus MC, 3000 CA Rotterdam, The Netherlands; 3) Department of Epidemiology, Erasmus MC, 3000 CA Rotterdam, The Netherlands; 4) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4102, Australia; 5) MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK; 6) MRC Epidemiology Unit, University of Cambridge, Cambridge CB2 0QQ, UK; 7) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 8) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen 2820, Denmark; 9) Steno Diabetes Center Copenhagen 2820, Denmark; 10) Department of Clinical Epidemiology, Leiden University Medical Centre, 2333 ZA Leiden, The Netherlands; 11) Department of Public Health and Primary Care, Leiden University Medical Centre, 2333 ZA Leiden, The Netherlands; 12) Postgraduate Program in Epidemiology, Federal University of Pelotas, 96020220 Pelotas, Brazil; 13) California Pacific Medical Center Research Institute, San Francisco, CA 94107, USA; 14) Institute of Biomedicine, Physiology, University of Eastern Finland, Kuopio 70211, Finland; 15) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montreal H3A 1A2, Canada; 16) Department of Medicine, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal H3T 1E2, Canada; 17) Hangzhou Normal University, 311121 Zhejiang, China; 18) Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands 6009, Australia; 19) School of Medicine and Pharmacology, University of Western Australia, Crawley 6009, Australia; 20) Department of Biostatistics, Boston University School of Public Health, Boston, MA 02131, USA; 21) Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, 41390 Gothenburg, Sweden; 22) deCODE Genetics/Amgen, Reykjavik, IC-101, Iceland; 23) Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI 48201, USA; 24) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA; 25) Data Tecnica International, Glen Echo, MD 20812, USA; 26) School of Women’s and Infants’ Health, University of Western Australia, Crawley 6009, Australia; 27) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH16 4UX, Scotland; 28) Centre for Cardiovascular Sciences, Queen’s Medical Research Institute of Edinburgh, University of Edinburgh, EH16 4UX, Scotland.
Institute, University of Edinburgh, Edinburgh EH16 4TJ, Scotland; 29) Hebrew SeniorLife, Institute for Aging Research, Roslinlade, MA 02131, USA; 30) Faculty of Medicine in the Galilee, Bar-Ilan University, Safed 1311502, Israel; 31) Department of Internal Medicine, Section of Geriatrics, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; 32) Division of GI, Hepatology, and Nutrition, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 33) Wake Forest School of Medicine Winston-Salem, NC 27101, USA; 34) Lund University, Department of Clinical Sciences Malmö, Clinical and Molecular Osteoporosis Research Unit, 20502 Malmö, Sweden; 35) Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland; 36) Department of Medicine, University of California, Davis; Sacramento, CA 95817, USA; 37) Department of Twin Research & Genetic Epidemiology, King’s College London, London SE1 7EH, UK; 38) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, MD 20892, USA; 39) Skåne University Hospital, Department of Orthopedics, 21428 Malmö, Sweden; 40) Division of Endocrinology, Diabetes, and Metabolism, Department of Internal Medicine, The Ohio State University, Columbus, OH 43210, USA; 41) Department of Endocrinology and Metabolism, Landslpati, The National University Hospital of Iceland, Reykjavik, IC-101, Iceland; 42) Faculty of Medicine, University of Iceland, Reykjavik, IC-101, Iceland; 43) Department of Clinical Epidemiology and Department of Internal Medicine, VU Medical Center, 1081 HV Amsterdam, The Netherlands; 44) Department of Pediatrics, Erasmus MC, 3000 CA Rotterdam, The Netherlands; 45) Department of Public Health & Preventive Medicine, Oregon Health & Science University, Oregon, Portland, OR 97239, USA; 46) Kuopio Research Institute of Exercise Medicine, Kuopio, 70100, Finland; 47) Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, 70210, Finland; 48) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 49) Geriatric Medicine, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, University of Gothenburg, 40530 Gothenburg, Sweden; 50) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland; 51) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98101, USA; 52) Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA; 53) Department of Medicine Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215, USA; 54) Broad Institute of MIT and Harvard, Boston, MA 02115, USA; 55) Center for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 41345 Gothenburg, Sweden; 56) Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, 45110, Greece; 57) Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN 55905, USA; 58) Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; 59) Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, London W12 0NN, UK; 60) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London W2 1PG, UK. 61) Center for Musculoskeletal Research, University of Rochester, Rochester, NY 14642, USA; 62) School of Clinical Sciences, University of Bristol, Bristol BS10 5NB, UK.

63) These authors contributed equally

64) These authors contributed equally

Corresponding author:

Dr. Fernando Rivadeneira MD PhD
Associate Professor Musculoskeletal Genomics
Genetic Laboratory - Department of Internal Medicine
Erasmus University Medical Center Rotterdam
The Netherlands
Email: f.rivadeneira@erasmusmc.nl
Abstract

Bone mineral density (BMD) assessed by DXA is used to evaluate bone health. In children, total body (TB) measurements are commonly used; in older individuals, BMD at the lumbar spine (LS) and femoral neck (FN) is used to diagnose osteoporosis. To date, genetic variants in more than 60 loci have been identified as associated with BMD. To investigate the genetic determinants of TB-BMD variation along the life course and test for age-specific effects, we performed a meta-analysis of 30 genome-wide association studies (GWAS) of TB-BMD including 66,628 individuals overall and divided across five age-strata each spanning 15 years. We identified variants associated with TB-BMD at 80 loci, of which 36 have not been previously identified; overall they explain approximately 10% of the TB-BMD variance when combining all age groups and influence the risk of fracture. Pathway and enrichment analysis of the association signals showed clustering within gene-sets implicated in the regulation of cell growth and SMAD proteins; overexpressed in the musculoskeletal system; and enrichment in enhancer and promoter regions. These findings reveal TB-BMD as a relevant trait for genetic studies of osteoporosis, enabling the identification of variants and pathways influencing different bone compartments. Only variants in ESR1 and close proximity to RANKL showed a clear effect dependency on age. This most likely indicate that the majority of genetic variants identified influence BMD early in life and their effect can be captured throughout the life course.

Introduction

Osteoporosis is a disease characterized by low bone mass and microarchitectural deterioration of bone tissue leading to increased risk of fracture. It is diagnosed through the measurement
of bone mineral density (BMD) utilizing dual-energy X-ray absorptiometry (DXA), which is the single best predictor of fracture\(^1\).

Bone is a dynamic tissue constantly undergoing resorption and formation. Bone mass increases steadily during childhood and markedly during adolescent growth\(^2\). Peak bone mass is attained at approximately the third decade of life. Thereafter, until about 50 years of age, BMD remains fairly stable, by virtue of the coupling between bone formation and resorption (e.g., bone remodeling). Subsequently, bone resorption exceeds the rate of bone formation, resulting in a decrease in BMD, particularly in women after the onset of menopause\(^3\).

The International Society for Clinical Densitometry recommends performing DXA measurements at the lumbar spine, femoral neck and total hip to diagnose osteoporosis in postmenopausal women and men who are 50 years or older\(^4\). Consequently, studies of BMD determinants are frequently based on measurements at these skeletal sites. By contrast, for the assessment of bone health in children and adolescents, total body (excluding head) and lumbar spine are the preferred sites to minimize measurement artifacts resulting from changing areas in growing bones\(^4\). Nevertheless, in elderly individuals degenerative changes in the spine can give elevated BMD readings\(^5\). Moreover, total body DXA scans have been obtained in a number of adult research cohorts, primarily to assess body composition. Therefore, the total body BMD (TB-BMD) measurement is the most appropriate method for an unbiased assessment of BMD variation in the same skeletal site from childhood to old age.

To date, nearly 80 independent genetic variants have been shown to be robustly associated with variability in bone parameters\(^6\)\(^-\)\(^18\). Most of these markers have been identified in studies
comprising tens of thousands of adult and elderly individuals with DXA-derived BMD measurements, although a few of them have been associated with BMD specifically in studies of pediatric cohorts. Furthermore, several of the associated variants display significant site-specific effects, possibly reflecting differences in bone composition across skeletal sites (e.g., cortical bone vs. trabecular bone) or differential response to mechanical loading. Moreover, genetic studies on measures from peripheral quantitative computed tomography (pQCT) and bone quantitative ultrasound, which provide additional information regarding bone size, geometry and (micro) architecture identified genetic variants that may have specific effects on bone properties that are poorly captured by conventional DXA measurements.

Given the complex physiological processes underlying age-related changes in BMD across the life course, it is possible that genetic studies in more refined age groups will reveal variants in unreported loci as well as age-specific genetic effects. Thus, the purpose of this study was to identify gene variants associated with TB-BMD across the life span and investigate possible differences of genetic effects across age periods.
Methods

**TB-BMD GWAS meta-analyses**

**Study Populations**

**Subjects**

This study comprised 30 epidemiological studies comprising ~66,628 individuals from populations across America, Europe, and Australia, with a variety of designs (Supplemental Data; Table S1) and participant characteristics (Table S2). In summary, most participants came from population-based cohorts of European ancestry (86%), two cohorts comprising African-American individuals (2%) and other four studies holding a fraction of individuals from admixed background (14%). All research aims and the specific measurements have been approved by the correspondent Medical Ethical Committee of each participating study. Written informed consent was provided by all subjects or their parents in the case of children.

**BMD measurement**

Total body BMD (g/cm²) was measured by DXA following standard manufacturer protocols. As recommended by the International Society for Clinical Densitometry total body less head (TBLH) was the measurement used in pediatric cohorts (e.g., 0-15 years). Detailed information on the assessments performed by each study can be found in Table S1.

**GWAS data and imputation**

All individuals included in this study had genome-wide array data. Quality control of genotypes is summarized in Table S1. To enable meta-analysis, each study performed genotype
imputation using the cosmopolitan (all ethnicities combined) 1000 genomes phase 1 version 3 (March 2012) reference panel, yielding ~30,000,000 SNPs for analysis. Three studies used the combined 1000 genomes and the UK10K reference panels as presented in Table S1.

**Association Analysis**

TB(LH)-BMD was corrected for age, weight, height and genomic principal components (derived from GWAS data), as well as any additional study-specific covariates (e.g. recruiting center), in a linear regression model. For studies with non-related individuals, residuals were computed separately by sex, whereas for family-based studies sex was included as a covariate in the model. Finally, residuals were inverse normal transformed. The analyses were performed in each study for the overall population as well as in subgroups of individuals by age-strata, defined by bins of 15 years (i.e., 0-15 years, 15-30 years, 30-45 years, 45-60 years, and 60 or more years). SNP association was tested for autosomal variants, in which the additive effect of each SNP on the normalized BMD-residuals was estimated via linear regression.

**Quality control of TB-BMD association summary statistics**

A centralized quality-control procedure implemented in EasyQC\(^1\) was applied to all study-specific files of association results to identify cohort-specific issues. We excluded variants if they had missing information (e.g., missing association P-value, beta estimate, alleles, allele frequency), or nonsensical values (e.g., absolute beta estimates or standard errors >10, association P-values >1 or <0; or imputation quality < 0; infinite beta estimates or standard errors); minor allele frequency (MAF) less than 0.5%; imputation quality scores <0.4 (Impute2) or <0.3 (Minimac). Moreover, variants were flagged if they had large allele frequency
deviations from reference populations (>0.6 for admixed studies and >0.3 for ancestry-homogeneous studies).

GWAS meta-analyses

In the first instance, no exclusion criteria based on ancestry were applied for the meta-analysis (N=66,628). In addition, meta-analyses were carried out across age strata (minimum sample size per bin N=200 for each study) comprising: 1) 0-15 years (N=11,807), 15-30 years (N=4,180), 30-45 years (N=10,062), 45-60 years (N=18,805), and 60 or more years (N=22,504). Further, summary data from cohorts of European ancestry only were meta-analyzed and used in subsequent analyses. We discarded variants present in less than three studies. Approximately 23,700,000 markers (including SNPs and INDELS) were assessed for association. We applied the conventional genome-wide significance level (GWS, P<5x10^{-8}) for SNP discovery.

Assessment of Age-dependent effects

We selected SNPs which were suggestively (12,567 SNPs, P<5x10^{-6}) associated with BMD in the overall meta-analysis, present in at least 2 studies per age-bin and with MAF differences across these meta-analyses lower than 0.5. We clumped this dataset with an r^2 ≥ 0.8, using as reference the most strongly associated SNPs with BMD and, pruning remaining SNPs within 0.7 Mb of each other. Age-dependent effects were assessed using a meta-regression approach for 1,464 SNPs obtained after this selection procedure. We ran a linear regression of the SNP effect estimates onto an intercept and the median age of each subgroup (e.g., each study stratified in age-bins). As proposed previously^{20}, standard errors of the effect estimates of each subgroup were multiplied by the square root of the genomic inflation factor when it was greater than 1.
We performed the meta-regression using the Metafor package\textsuperscript{21}, and any statistical evidence of linear association was corrected for multiple testing (Bonferroni correction; $0.05/1,464=3.4\times10^{-5}$). The difference between beta-estimates in children vs. elderly meta-analyses (Pdiff) was tested using Easy-strata\textsuperscript{22}.

**Approximate conditional meta-analyses**

Conditional analyses were undertaken based on the meta-analysis of the studies of European ancestry only (N=56,284). Only variants in the loci that reached GWS in this meta-analysis were assessed. The Rotterdam Study I (n=6,291) was used as reference for precise calculation of the linkage disequilibrium (LD) between the analyzed markers. We used an iterative strategy as implemented in GCTA\textsuperscript{23} to determine: 1) independence of association signals within loci discovered in our study, by means of stepwise model selection procedure per chromosome (--massoc-slct routine); and 2) the novelty of the association signals discovered by our meta-analysis with regard to variants reported in previous well-powered GWAS of different bone traits (**Table S3**). To this end, we performed the association analysis conditional on 78 variants present in our data and associated with different bone-traits (--massoc-cond routine). These 78 SNPs were selected from different GWAS publications\textsuperscript{6-10,12-14}, assuring their independence to avoid collinearity issues.
Shared Genetic architecture of TB-BMD fracture and other traits

LD score regression analyses

We used the LD score regression package to estimate the heritability of TB-BMD and rule out that our results were a product of bias (e.g., residual population stratification or cryptic relatedness). LD score regression uses GWAS summary statistics and assesses the SNP-heritability based on the expected relationship between linkage disequilibrium (LD) of neighboring SNPs and strength of association under a polygenic model. As this methodology relies on the LD structure throughout the genome, we restricted this analysis to summary statistics from the meta-analysis of cohorts comprising only individuals from European ancestry. We used the publicly available, pre-computed LD structure data files specific to European populations of the HAPMAP 3 reference panel. An extension of this method allows estimating the genetic correlation between two traits. This can be performed in the LDhub pipeline, a web utility which gathers data from many different GWAS meta-analysis. From the 199 traits, currently available in the website, we have restricted our analysis to those traits whose heritability z-scores were larger than 4 and were analyzed only in European ancestry individuals (following the recommendations in the LD score software website (Web Resources)). Additionally, we incorporated data from a recent GWAS meta-analysis of any-type of fracture in individuals from European ancestry (N= 264,267; 37,778 cases) (K.T, unpublished data). In total, we assessed the genetic correlation between TB-BMD and 74 traits.
Mendelian randomization analysis

We undertook a two-sample Mendelian randomization approach\textsuperscript{27} to estimate the causal effect of TB-BMD on any-type of fracture in the Europeans samples. In short, we constructed a score based on the independent genetic variants from the TB-BMD meta-analysis (European set and excluding secondary signals), whenever the selected variant was not present in the fracture meta-analysis, the second variant with the lowest p-value in the locus (P<5x10^{-8}) and $r^2 > 0.8$ was used as proxy. Thereafter, estimates derived from the TB-BMD summary statistics were pooled using methods similar to inverse-variance weighted fixed meta-analysis using the meta R-package (Web Resources).

Search for biological and functional knowledge of the identified association regions

For all those SNPs outside a 500Kb window from previously known bone associated SNPs we did a literature search in PubMed and Web of Science to evaluate if nearby genes (within 500Kb) were known to play a role in bone metabolism. Also, we determined if the annotated genes underlie any human Mendelian disorder with a skeletal manifestation, had knockout mouse models with a skeletal phenotype or were annotated to pathways critical to bone metabolism. Genomic annotation for all SNPs was made based on UCSC hg19.

DEPICT analyses

We used DEPICT\textsuperscript{28}, a recently developed tool to prioritize genes at the associated regions, define possible pathways by enrichment testing, and identify tissue and cell types in which genes from loci associated with TB-BMD. The methodology first selects all lead SNPs below a certain threshold with respect to a target P-value. We tested both the complete set of GWS
SNPs and the subset of those mapping only to loci not previously reported. Enriched gene-set were group based on the degree of gene overlap into ‘meta gene-sets’ as proposed earlier and their correlation visualized using Cytoscape 3.4 (Web Resources).

**Functional annotation to microRNA binding sites**

We used the PolymiRTS, miRdSNP, and microSNiPer databases to obtain a list of variants located in predicted microRNA binding sites on the 3’UTRs of genes, as described in detail elsewhere. In summary, index SNPs (most associated variant) of the GWS loci were submitted to SNAP (Web Resources) to retrieve their high LD proxy SNPs (with $r^2 > 0.8$, limit distance 500 kb, and CEU panel) in the 1000 genomes project. The resulting list of SNPs was annotated to the list of microRNA binding site variants obtained from the above mentioned publicly available databases.

**Functional enrichment analysis of trait-associated variants**

GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) was used to characterize the putative functional contribution of TB-BMD associated variants mapping to non-coding regions. GARFIELD employs a non-parametric analysis to calculate fold enrichment values for regulatory marks, at given significance thresholds and then tests them for significance via permutation testing while accounting for LD, MAF and local gene density. We used data regarding DNase I hypersensitive sites, transcription factor binding sites, histone modifications and chromatin states (ENCODE and Roadmap Epigenomics) from 424 cell types and tissues to capture and characterize possible cell-type-specific patterns of enrichment, as provided in the GARFIELD software (Web
Resources). Fold enrichment statistics were tested at the four different significance thresholds (i.e., $1 \times 10^{-8}$, $1 \times 10^{-7}$, $1 \times 10^{-6}$ and $1 \times 10^{-5}$). Multiple-testing correction was performed on the effective number of annotations used, using the default P-value threshold of $1 \times 10^{-4}$.

Knockout animal models and gene expression in bone cells

Animal models survey

We surveyed databases from The International Mouse Phenotyping Consortium\textsuperscript{34} together with The International Knockout Mouse Consortium\textsuperscript{35} to identify knockout models of candidate genes resulting in skeletal phenotypes. Furthermore we mined data from The Origins of Bone and Cartilage Disease (OBCD) project\textsuperscript{36}, specialized in murine skeletal phenotypes including Digital X-ray microradiography on femurs and tail vertebrae, Micro-CT analysis, femur three-point bend test load–displacement curves and tail vertebrae compression testing from knockout mice and wild-type controls at 16 weeks of age.

Gene expression in murine bone cells

Gene expression profiles of candidate genes were examined in primary mouse osteoblasts undergoing differentiation and bone marrow derived osteoclasts. To study murine osteoblasts, pre-osteoblast-like cells were obtained from neonatal calvaria collected from C57BL/6J. Next Generation RNA sequencing using an Illumina HiSeq 2000 was used to evaluate the transcriptome every two days from day 2 to 18 days post osteoblast differentiation\textsuperscript{7}. Expression of genes in murine osteoclasts was determined using publicly available data obtained using Next-Gen RNA-sequencing applied to bone marrow derived osteoclasts obtained from 6-8 week old C57BL/6 mice\textsuperscript{37}.
Gene expression in human bone cells

Gene expression profiles of candidate genes were examined in human bone marrow derived mesenchymal stem cells differentiated into osteoblast. Total RNA (n=3) was isolated at day 0 (MSCs) and day 4 of osteoblast differentiation\textsuperscript{38}. Also, RNA was isolated during osteoclast differentiation. Peripheral blood mononuclear cells derived from buffy coats (Sanquin, Amsterdam, the Netherlands) were seeded in 96-well plates (5x10\textsuperscript{5} cells per well) as previously described\textsuperscript{39}. Total RNA (n=3) was isolated using Trizol at day 0 (PBMCs) and at day 7 of osteoclast differentiation. Illumina HumanHT-12 v3 BeadChip human whole-genome expression arrays were used for expression profiling. The quality of isolated RNA was assessed on a 2100 Bioanalyzer (Agilent Technologies). Data were analyzed as described in detail previously\textsuperscript{38}. Genes were designated as being expressed when at least one probe coding for the gene was significantly present in at least 2 of the 3 biological replicates.

Results

TB-BMD GWAS meta-analyses

Analyses including all age-strata

Our meta-analysis of TB-BMD GWAS summary statistics (N=66,628) identified variants in 76 independent loci associated with TB-BMD at a genome-wide significant (GWS, P\textless{}5x10\textsuperscript{-8}) level (Figure 1, Table S4). Overall, there was no evidence of a strong inflation (genomic inflation factor (\(\lambda\)) of 1.08, Figure S1). Yet, inflation was observed in the range of common variants (0.2>MAF<0.5, \(\lambda\)=1.19) due to polygenicity (LD score regression intercept = 1.007). In our results, one of the signals mapping to LDLRAD3 was driven entirely by individuals of African
background (MAF=0.043 in YRI panel) since the two associated variants are monomorphic in all other populations. The low allele frequency of this variant in our study (MAF= 0.025) and our limited statistical power (N=6,748) in non-European samples warrants independent replication efforts to exclude the possibility of a false-positive association.

In addition, a meta-analysis comprising 56,284 individuals of European ancestry (~84% of the study population) identified variants in two additional GWS loci (Figures S1-S2, Table S5). Association signals mapping to these loci were close to the GWS threshold in the overall meta-analysis (P=1x10^-7) and showed no evidence of heterogeneity (P_{het}>0.1). One of them, in 12q24.21 (MED13L), has not been previously associated with bone parameters (Table 1, Figure S3), while the other in 21q22.13 (CLDN14), is not fully independent from the previously reported hip-BMD association signal^{13} (Table S5).

Of the 78 identified loci, variants in 35 (45%) were not located within 500 kb of known association signals nor in regions of extended LD with them (Table 1, Figure S4). Index SNPs at these 35 loci were, in general, common non-coding variants. Twenty-two of these, are located in close proximity to genes likely to influence bone metabolism as shown by previous functional studies (Table 1, Figure S3), including CSF1 ([MIM 120420] important for osteoclast differentiation^{40}) and SMAD3 ([MIM 603109] a critical component of the TGF-beta signaling pathway^{41}). Across these 35 signals, 31 of the index SNPs were nominally associated (P<0.05) with either lumbar spine or femoral neck BMD in the same direction as in the previously published GEFOS GWAS meta-analysis^{7} (Table 1). This comparison was not possible for the rs113964474 variant, because it was not available in the GEFOS study. Moreover, we found directionally-concordant effect estimates (P < 0.05) for 73 of the 78 index SNPs of known bone
association signals (Table S3). The markers which failed to replicate in our study were either previously associated with lumbar spine BMD but not femoral neck BMD (rs3905706 [MPP7, 1p12.1] and rs1878526 [INSIG2, 2q14.2]), associated specifically with the hip trochanter and intertrochanteric subregions (rs1949542 [RP11-384F7.1, 3q13.32]), or associated with BMD only in women (rs7017914 [XKR9, 8q13.3]) or only in children (rs754388 [RIN3, 14q32.12]).

Age-dependent effects

Meta-analyses across age strata resulted in the identification of variants mapping to 2 additional loci that were not detected in the overall meta-analysis (Figure S5; Table S6). In children (age group 0-15 years), the previously known 14q32.12 locus, harboring RIN3 (rs72699866, P=1x10⁻⁸); and in the middle-aged (age group 45-60 years), a signal in the 19q12 locus mapping in the vicinity of TSHZ3 (rs6510186, P=3.1x10⁻⁸) were identified. The rs72699866 variant leading the RIN3 signal in the youngest age stratum showed no evidence of association (P=0.16) and high heterogeneity (P_{het}=6.6x10⁻⁵) in the overall meta-analysis. In fact, the effect of rs72699866 decreased significantly with age (P_{trend}=1.69x10⁻⁹) (Figure S6) and showed a significant difference between the two extreme groups, i.e. children vs elderly (β_{0-15}=0.099 [0.066, 0.134]; β_{>60}=-0.035 [-0.060, -0.010]; P_{diff}=4.32x10⁻¹⁰). In contrast, the rs6510186 variant [19q12] showed nominal evidence of association and heterogeneity in the overall meta-analysis (P=0.02; P_{het}=0.03). Nevertheless, no clear pattern of age-dependency was identified (P=0.2) for this SNP (Figure S6).

We also applied meta-regression analysis and found that variants mapping to 42 different loci showed nominally significant age dependent effect (P<0.05) (Table S7, Figure S7). In summary,
27 (64%) of the loci showed stronger effects in the older age groups. Of these, variants in the 6q25.1 (ESR1) and 13q14.11 (RANKL) loci remained significant after multiple-testing correction (P<3.4x10^{-5}) (Figure 2); while variants in 6p21.1 (RUNX2, rs148460475), 15q21.2 (CYP19A1, rs2414098), 17q21.31 (MEOX1, rs74835612) and 11p15.1 (SOX6, rs11822790) were only suggestive at P<1x10^{-3}.

Conditional association analyses

The step-wise conditional approach included studies comprising only individuals of European ancestry, as the method used relies on appropriate representability of the LD reference. Of the 76 GWS loci identified in the overall analysis, variants in 57 (19 previously unreported) loci were also GWS in the European-only analysis (Figure S2), likely a consequence of the lower power in this subgroup. We identified 81 SNPs independently associated with TB-BMD mapping to 58 different loci (one European-specific), 18 of which depicted multiple distinct signals attaining GWS (Table S8). These independent variants together explained 10.2% of TB-BMD variance. This proportion is slightly higher than the 7.4% TB-BMD variance explained by the 78 known variants associated with bone traits. Moreover, we identified independent signals in 13 of the 78 known bone loci after conditional analyses. (Figure S2; Table S8).

Shared Genetic architecture of TB-BMD, fracture and other traits

SNP-heritability of TB-BMD in the European samples was estimated to be 0.259 (SE 0.017). TB-BMD was highly genetically correlated with BMD measured at other skeletal sites (p>0.9). Among the non-BMD traits, all-type of fracture showed the highest correlation [p=-0.61 (P=1.6x10^{-27})]. The MR approach indicated a strong causal relation where per 1 standard
deviation decrease in genetically determined TB-BMD there is 56% increase in the risk of fracture (Odds ratio 1.56 [1.50-1.62]). Other anthropometric, metabolic and disease traits showed significant (yet weak) correlation with TB-BMD (Table S9, Figure 3). In contrast, other established risk factors for osteoporosis such as menopause or age of menarche showed no significant genetic correlation with TB-BMD.

**Biological and functional knowledge of the genes in BMD-associated loci**

Loci not previously reported and their potential role in bone metabolism are summarized in Table 1. Several loci harbor genes implicated directly in bone metabolism (SLC8A1 [MIM 182305], PLCL1 [MIM 600597], ADAMTS5 [MIM 605007]), affecting osteoblast or osteoclast differentiation and activity (CSF1 [MIM 120420], DUSP5 [MIM 603069], SMAD3 [MIM 603109], SMAD9 [MIM 603295], CD44 [MIM 107269]), participating in Wnt signaling (FZD7 [MIM 603410], TCF7L1 [MIM 604652]), or regulating processes such as manganese or calcium absorption (GCKR [MIM 600842], DGKD [MIM 601826], SLC30A10 [MIM 611146]) among others; while genes in at least 14 loci exert a potential novel role in bone biology. Rodent knockout models of several genes in the implicated loci, show an altered skeletal phenotype (e.g., osteopetrosis [Csf1], increased bone resorption [Aqp1, Cyp19a1, Cd44], impaired skeletogenesis [Apc, Runx2, Smad3], deformities in the axial skeleton [Btg1, Atpaf2]). Whereas an effect on bone can be inferred for genes in other associated loci, for example, CYP19A1 [MIM 107910] in 15q21.2 is an estrogen synthesis gene, being estrogen a key compound for bone maturation and maintenance, and ZKSCAN5 [MIM 611272] in 7q22.1 is associated with circulating dehydroepiandosterone sulphate (DHEAS) levels. DHEAS levels are positively correlated with BMD in adults and post-menopausal women. Across these loci,
not previously reported as associated with BMD variation, we identified six exonic variants
associated with TB-BMD, three of which were nonsynonymous variants all cataloged as benign
both by SIFT and polyphen2. We also identified 53 GWS coding variants in known loci, of which
33 are non-synonymous (Table S10). Only a low-frequency variant in LRP5 [MIM 603506],
rs4988321/A (11:68174189, MAF=0.04), has a clinical annotation, constituting a homozygous G-
to-A transition variant identified in a person with osteoporosis-pseudoglioma syndrome (OPPG
[MIM 259770])

**DEPICT analyses**

Based on the overall meta-analysis, 53 genes were prioritized (FDR<0.05), 15 of them mapping
to loci not previously described (Table S11). Cells and tissues from the musculoskeletal system
presented the largest enrichment of gene expression within the associated loci (Figure 4).
These genes were overrepresented in 182 pathways clustered in 25 ‘meta gene-sets’ (Table
S12). The large majority of the clusters are involved in musculoskeletal development and bone
homeostasis (Figure 4). The most significant of these implicated the regulation of cell growth,
and the TGFB signaling pathway and its mediating SMAD proteins.

Restricting the DEPICT analysis to the subset of not previously reported associated regions
resulted in significant enrichment of genes expressed in the musculoskeletal and immunological
systems (Figure S8). Genes mapping to these loci were overrepresented in the SMAD binding
pathway and TGFBR2 PPI (protein-protein interaction) subnetwork (FDR<0.05).

**Functional annotation to microRNA binding sites**
We then assessed if the index SNPs of the 80 GWS loci detected in the main and subsequent GWAS (or their proxies in strong LD; \( r^2 > 0.8 \)) were located in predicted microRNA binding sites within the genes’ 3’UTRs and thus, were expected to disrupt the regulation of gene expression (Table S13). The index SNP within the 3’UTR of ZKSCAN5 (mapping to a locus not previously identified), rs34670419 (MAF=0.04), is predicted to create a binding site for miR-382-3p, a microRNA which is expressed in osteocytes and has been recently shown to be involved in osteogenic differentiation\(^66\). In addition, eight proxy SNPs (mapping to PSMD13, ABCF2, GALNT3, PKDCC, REEP5, PPP6R3, AAGAB and TOM1L2) are predicted to influence the binding of microRNAs to transcripts of their host gene.

**Functional enrichment analysis of trait-associated variants**

As typically found in GWAS, the great majority of identified associations emerged from non-coding common variants and hold no direct annotation to molecular mechanisms. To assess if there is relative enrichment of regulatory genomic marks underlying the associated variants in a cell-specific context, we used GARFIELD\(^33\). We found relative ubiquitous enrichment for TB-BMD variants (Empirical \( P < 2.4 \times 10^{-4} \)) in DNase I hypersensitive sites across the different cell types (Figure S9). Further, we found higher levels of fold-enrichment for enhancers (median 3.6, range [2.7, 4.4]) and promoters (median 3.2, range [2.9, 3.5]) than for transcribed regions (median 1.8, range [1.5, 2.2]).

**Gene expression in bone cells and knockout animal models**

From the 53 genes prioritized by DEPICT only 49 had a mouse orthologue (Table S14). From these genes, only Mepe (osteocyte-specific) and Foxl1 were not expressed in murine osteoblast
or osteoclast. Moreover, 61% of the prioritized genes were expressed in human cells in vitro during osteoblast or osteoclast differentiation (Table S14). AQP1 was the only prioritized gene mapping to a locus not previously reported showing no expression in the human bone cells differentiation experiments.

Knockout models were widely available in at least one of the different databases assessed. Nevertheless in-depth bone phenotyping performed under the OBCD project was only available for four knockout models (Table S15). Two of these, DUSP5 and CD300LG showed no significant bone phenotype. The TCF7L1 knockout model only showed lower cortical diameter in the femur without other clear bone phenotype. Nevertheless, TCF7L1 was shown to be expressed during osteoblastogenesis. Conversely, homozygous knockout for CREB3L1 showed a clear bone phenotype consisting of low BMC both at the vertebrae and femur together with a strong trabecular and cortical phenotype affecting bone strength (Figure S10). CREB3L1 maps to 11p11.2, a previously identified BMD locus harboring ARHGAP1 and LRP4 as candidates to underlie the GWAS signal in a region of extended LD.

**Discussion**

This meta-analysis of TB-BMD comprising up to 66,000 individuals identified variants in 36 loci not previously reported and replicated at GWS level several association signals identified by GWAS of diverse bone phenotypes. Bioinformatics analyses suggest enrichment of these 36 loci for genes expressed in the musculoskeletal system, and solidly represented in the SMAD binding pathway and the TGFBR2 PPI subnetwork. We also demonstrate that for variants in few loci the size of the effect is age dependent; variants in two loci (RIN3 and TSHZ3) were...
identified only by the age-stratified analyses despite less power (smaller sample size); while for
variants in two other loci (ESR1 and RANKL) there was significant evidence of age heterogeneity
derived from a meta-regression of the genetic effects with age. Our results strengthen the
evidence that genetic variants influence BMD from a young age and support the value of peak
bone mass as an important determinant of bone health later in life.

Traditionally, DXA-BMD measurements performed at sites of high fracture risk (i.e., femoral
neck, lumbar spine and forearm) have been used in genetic epidemiological investigations of
bone health in adults. Instead, we have used BMD measurements derived from total body
scans. Not only do we show a high overlap of association signals with previous GWAS of
different bone traits, including DXA, pQCT and ultrasound measurements, but we have also
identified unreported loci. Five known associations failed to replicate in our studies, even
though we cannot discard these associations constitute false-positives, these results might also
indicate that variants whose effect is highly specific to skeletal sites, skeletal properties, sex or
age groups cannot be detected in our TB-BMD meta-analysis. It is plausible that more variants
of this type exist and will be discovered as site-specific BMD meta-analyses are performed in
increasingly powered settings. Furthermore, the genetic correlation of TB-BMD with BMD
measured at other sites was close to one. Whilst, we found that a decrease of one standard
deviation in the genetically determined TB-BMD resulted in at least 50% higher odds of
suffering a fracture. Significant genetic correlations with other traits (i.e., BMI, IGF1 and
ulcerative colitis) reflect the systemic context of skeletal biology and merit further study by
future efforts to elucidate the underlying mechanisms.
Genes in the associated loci were highly expressed in the musculoskeletal system and overrepresented in gene-sets related to bone development. The prioritized gene CREB3L1 [MIM 616215] in 11p11.2 observed a clear bone phenotype in our mouse knockout model, which corroborates the findings of previous work showing substantial rescue of CREB3L1 deficiency with bisphosphonates and its critical role for bone formation. This locus characterized by extended LD, also harbors LRP4 [MIM 604270] whose knockout model presents with increased trabecular and cortical bone mass. This is in line with our conditional analysis identifying multiple independent signals in the region making it likely that both genes are influencing bone biology. Altogether, we demonstrated that TB-BMD offers a powerful alternative to identify genetic variants associated with bone metabolism.

Variants mapping to 14q32 harboring RIN3 [MIM 610223] were only associated at a GWS level in children (i.e., <15 years), and were only nominally significant in the elderly group (i.e., >60 years). This age-related heterogeneity may explain why this locus has not been detected in BMD meta-analyses in adults, although being identified in relation to pediatric BMD and Paget’s disease (PDB [602080]) GWAS. In addition, another signal mapping to 19q12 harboring TSHZ3 [MIM 614119] was significant in adults aged 45-60 years but not in other age groups analyzed or in previous studies, alluding to a false-positive association, thus replication of this finding is necessary.

Our analyses revealed variants in the 6q25.1 (ESR1) and 13q14.11 (RANKL) loci demonstrating the most compelling evidence for age-modulation effects. The 6q25.1 locus harboring ESR1 [MIM 133430], an important genetic factor in normal BMD variability, was not associated with BMD in children below 15 years of age, where the largest cohorts (i.e., Avon Longitudinal Study...
of Parents and Children (ALSPAC) and the Generation R Study) comprise predominantly pre-pubertal children. As levels of estradiol before puberty are low\(^70\), a negligible effect of \(ESR1\) variants on BMD is expected. Likewise, in mouse models the expression of \(RANKL\) [MIM 602642] in bone is markedly increased with advancing age from young to adult and related to bone loss\(^71\). Accordingly, variants influencing \(RANKL\) expression show a larger effect later in life.

In general, a substantial heterogeneity of the genetic effects in the overall meta-analysis was explained by age, nevertheless, the inclusion of larger sample sizes (avoiding age exclusion criteria and incrementing statistical power) leveled off the loss of power due to the heterogeneity of the genetic effects.

In brief, variants with evidence of age-specific effects were exceptional in our study. These results might reflect a lack of statistical power as only SNPs showing suggestive evidence (\(P<5\times10^{-6}\)) of association with TB-BMD in the overall meta-analysis were tested for age-specific effects. This selection criteria aimed to include SNPs whose heterogeneity might have hampered their statistical significance in the overall meta-analysis, and at the same time maximize the power to discover variants with real age-dependet effects. Alternatively, these results indicate that most of the genetic variants identified so far, by us and others, influence BMD from early ages onwards, and their effect persist throughout the life course. However, variants in 27 of the 42 loci (64\%) showing nominal evidence for age dependent effects had larger effects in the older groups. Nonetheless, this requires careful interpretation given the uneven sample sizes between the age groups and the criteria to select markers for the meta-regression based on significance in the overall meta-analysis. Collectively, this argues in favor of
enlarging studies focused on younger populations – where the statistical power is still restricted – to discover additional genetic variants influencing BMD.

Our study has some limitations. A key disadvantage of our design is that we group the data based on age spans rather than life stages. Crucial information for this assessment, such as puberty onset in children and adolescents or menopausal status in the adults, was not available across the majority of the cohorts. Other strategies like using shorter age spans will result in even less statistical power of the discovery setting. Similarly, despite the large sample size of our study, we identified very few variants in the low-frequency spectrum (MAF <5%) indicating that comprehensive surveys of rare variation influencing BMD still require even larger sample sizes, on top of better resources for imputation of the rarer variants, possibly needing population-specific references. Such strategies will be key to explain a larger fraction of the genetic variability of BMD phenotypes, as illustrated for other traits such as height or BMI.

Moreover, the identified SNPs are in their vast majority, non-coding variants, raising the possibility that the causal genes are different from the candidate genes we have prioritized based on the current biological knowledge and bioinformatic prediction tools. Additional functional studies are required to determine the potential role of the genes in the identified loci.

In conclusion, we performed a genome-wide survey for association with DXA derived TB-BMD, combining data from five age groups including children and older individuals. In contrast to previous large-scale meta-analyses, we used DXA derived TB-BMD rather than measurements on specific skeletal sites prone to fracture to identify genetic factors influencing BMD variation. We demonstrate that TB-BMD is a valid phenotype for this purpose, as we replicated more than
90% of the previously reported signals. Most importantly, we identify variants in 36 loci associated with TB-BMD not previously reported by previous GWAS of bone phenotypes. Our results show steadiness in the magnitude of the genetic effects on BMD for most of the BMD-associated variants. While the contrasting skeletal physiology across different age periods is well established (i.e. endochondral ossification, linear growth, modelling, remodeling, etc.), peak bone mass acquisition remains the major determinant of variability at any age. These findings strongly support the importance of the bone accrual process in the definition of BMD status and fracture susceptibility throughout the life course.

Accession Numbers

GWAS Summary data for the main and age-strata meta-analyses together with the corresponding regional plots of GWS signals have been deposited in the GEFOS website (Web Resources). Gene expression data presented in this paper can be retrieved from the Gene Expression Omnibus (GEO) as follows: Murine osteoclasts (GSM1873361) and osteoblasts (GSE54461); human osteoblast differentiation (GSE54461).

Supplemental Data

Supplemental data include a full list of acknowledgements, cohort short descriptions, 15 tables and 10 figures.

Acknowledgements

The authors would like to thank the many colleagues who contributed to collection and phenotypic characterization of the clinical samples, as well as genotyping and analysis of the
GWAS data. Part of this work was conducted using the UK Biobank resource.

Conflicts of interests

Psaty serves on the DSMB of a clinical trial for the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.

Web Resources

LDhub, http://ldsc.broadinstitute.org/
Meta R-package, https://github.com/guido-s/meta)
OBCD, http://www.boneandcartilage.com/
References


and pubertal boys and girls by a novel ultrasensitive gas chromatography-tandem mass spectrometry method. J Clin Endocrinol Metab 95, 82-92.


Figure Titles and Legends

**Figure 1.** Manhattan plot of association statistics (-log10(P-values)) for TB-BMD overall meta-analysis. Each dot represents a SNP and the x-axis indicates its chromosomal position (built 37 NCBI). Red dots represent SNPs at GWS loci that are not within ±500Kb of leading SNPs in previous GWAS with different bone traits. Dashed horizontal red and yellow lines mark the GWS threshold (P<5x10^{-8}) and suggestive threshold (P<1x10^{-6}), respectively. Novel loci in the only-CEU analysis are not shown.

**Figure 2.** Age dependence of the genetic variant effect in the meta-regression. The panels display leading SNPs from two loci exhibiting significant evidence for age influences. Heterogeneity P-values (P_{het}) are reported for the overall meta-analysis. In the left panels, each circle represents a study subgroup (i.e., study divided in age strata), with the circle size proportional to the inverse variance of the SNP main effect. In the right panels, forest plots display estimates obtained from each age-bin meta-analysis, with the symbol size proportional to the inverse variance of the SNP main effect.

**Figure 3.** Genetic correlations between TB-BMD and other traits and diseases. Calculation was based on the summary statistics of the only-European meta-analysis (N=56,284) and estimated by LD score regression implemented in LDHub. The diagram only show traits whose correlation with TB-BMD was significant (P<0.05).

**Figure 4.** Depict results for gene-set and cell/tissue enrichment analyses. Top panel: 25 Meta gene-sets were defined from similarity clustering of significantly enriched gene sets (FDR<5%). Each Meta gene-set was named after one of its member gene sets. The color of the Meta gene-sets represents the P-value of the member set. Interconnection line width represents the Pearson correlation ρ between the gene membership scores for each Meta gene-set (ρ < 0.3, no line; 0.3 ≤ ρ < 0.5,narrow width; 0.5 ≤ ρ < 0.7, medium width; ρ ≥ 0.7, thick width). Bottom panel: Bars represent the level of evidence for genes in the associated loci to be expressed in any of the 209 Medical Subject Heading (MeSH) tissue and cell type annotations. Highlighted in orange are these cell/tissue types significantly (FDR<5%) enriched for the expression of the genes in the associated loci.

**Tables**
Table 1. Index SNPs of loci not previously associated with BMD. Variants associated with TB-BMD in the all-ages combined meta-analysis that map outside +/- 500 Kb of known index SNPs of genetic associations with different bone traits. Genomic coordinates are on build 37 of the human genome. Notes refer to annotation based on the closest gene. Associations with Lumbar Spine (LS) and Femoral Neck (FN)-BMD. Beta coefficients and allele frequencies (EAF) are reported for the A1 allele.
<table>
<thead>
<tr>
<th></th>
<th>rsID</th>
<th>Chromosome</th>
<th>Position</th>
<th>Genotype</th>
<th>MAF</th>
<th>P-value</th>
<th>Log10(P-value)</th>
<th>Gene</th>
<th>Annotation</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Beta</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>39103882</td>
<td>20q12</td>
<td>0.30</td>
<td>0.035</td>
<td>3.50E-08</td>
<td>66497</td>
<td>intergenic</td>
<td>MAFB</td>
<td>Osteoclast differentiation</td>
<td>0.027</td>
<td>0.007</td>
<td>0.015</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>28773868</td>
<td>21q21.3</td>
<td>0.59</td>
<td>-0.035</td>
<td>1.74E-09</td>
<td>66489</td>
<td>intergenic</td>
<td>ADAMT5S</td>
<td>Endochondral Ossification</td>
<td>-0.029</td>
<td>0.001</td>
<td>-0.015</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>36970350</td>
<td>21q22.12</td>
<td>0.45</td>
<td>-0.038</td>
<td>8.01E-11</td>
<td>66514</td>
<td>intronic</td>
<td>RUNX1</td>
<td>Osteoclast differentiation</td>
<td>-0.019</td>
<td>0.031</td>
<td>-0.016</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>40350744</td>
<td>21q22.2</td>
<td>0.84</td>
<td>-0.043</td>
<td>2.99E-08</td>
<td>66298</td>
<td>intergenic</td>
<td>ETS2</td>
<td>Osteoblast maturation</td>
<td>-0.028</td>
<td>0.020</td>
<td>-0.028</td>
<td>0.007</td>
<td></td>
</tr>
</tbody>
</table>

* Monomorphic in European cohorts. ** Reported statistics from the meta-analysis in the 30-45 age-strata.