Disentangling the Heterogeneity of Bone Accrual

Maria Carolina Medina Gomez

The general design of the Generation R Study was made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Ministry of Health, Welfare and Sport, the Ministry of Youth and Families, the Netherlands Organization of Scientific Research(NWO) and the Netherlands Organization for Health Research and Development (ZonMw).

The work presented in this thesis was conducted at the department of Internal Medicine and the Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. The publication of this thesis was financially supported by the De Nederlandse Vereniging voor Calcium- en Botstofwisseling (NVCB) and the Erasmus University Rotterdam, the Netherlands.

ISBN: 978-94-6169-876-6

Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

Disentangling the Heterogeneity of Bone Accrual

Ontwarren van de heterogeniteit van bot opbouw

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 31 mei 2016 om 11:30 uur

door

Maria Carolina Medina Gomez
geboren te Bogotá, Colombia

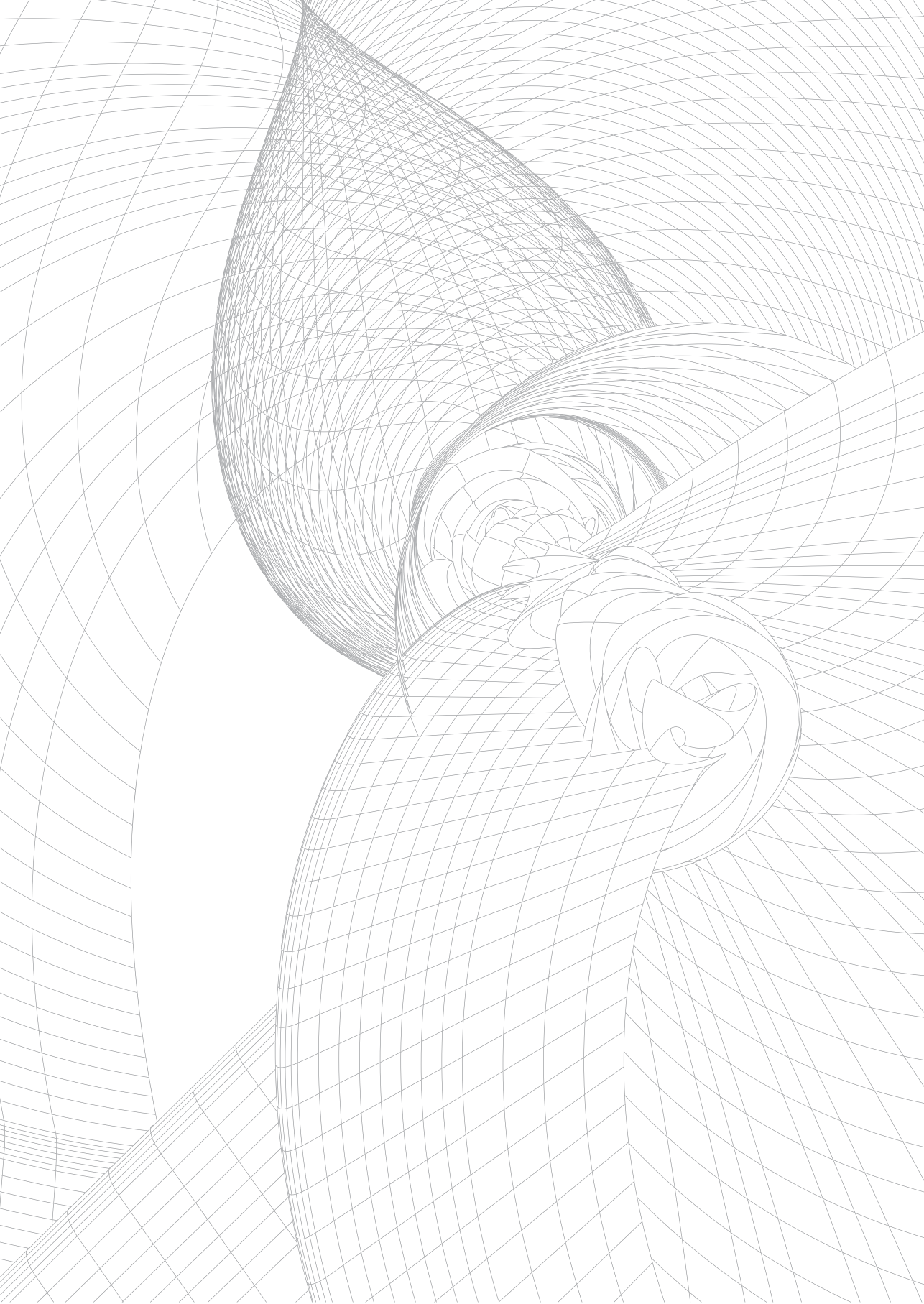
Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotor:	Prof.dr. A.G. Uitterlinden
Other members:	Prof.dr. B.S. Zemel
	Prof. dr. C.M. van Duijn
	Prof.dr. M.H. Kayser
Copromotor:	Dr. F. Rivadeneira

Paranimfen:	Katerina Trajanoska
	Jeroen van Rooij

CONTENT



Chapter 1

General Introduction

Bones play an important part in the overall function of the human body. Besides being a highly specialized supporting framework of the body, bones protect vital organs, provide an environment for marrow (both blood forming and fat storage), act as a mineral reservoir for calcium homeostasis and as storehouse for growth factors and cytokines [1]. One of the most important properties of bone is its ability to regenerate and repair; bones constantly undergo reshaping (modeling) during life to help them adapt to changing biomechanical forces; they also undergo remodeling to remove old, microdamaged bone and replace it with new, mechanically stronger tissue in order to help preserve bone strength [2]. In fact, this process is so dynamic that it is estimated that the entire adult human skeleton is replaced every ten years [3].

During childhood and adolescence, longitudinal growth of the skeleton takes place. The maximum amount of bone accrued at the end of the growth period is known as peak bone mass (PBM). PBM occurs by the end of the second or early in the third decade of life[4], yet more than 90% of PBM is already accrued by the age of 18 years [5]. Longitudinal growth occurs at the growth plates, where cartilage proliferates in the epiphyseal and metaphyseal areas of long bones, before subsequently undergoing mineralization to form primary new bone. This first type of bone, also known as woven bone is later replaced by secondary or lamellar bone (mature). Secondary bone is further classified as two types: cortical bone and trabecular bone (**Figure 1**). The cortical bone is dense, solid, and surrounds the marrow space while the trabecular bone that is made of a honeycomb-like network of trabecular plates and rods is interspersed in the bone marrow compartment. The mature skeleton is composed of 80% cortical bone and 20% trabecular bone overall, where different bones and skeletal sites within bones have different ratios of cortical to trabecular bone [2]. For example, the vertebra is composed of cortical to trabecular bone in a ratio of 25:75, while the ratio is 50:50 in the femoral head and 95:5 in the radial diaphysis [2].

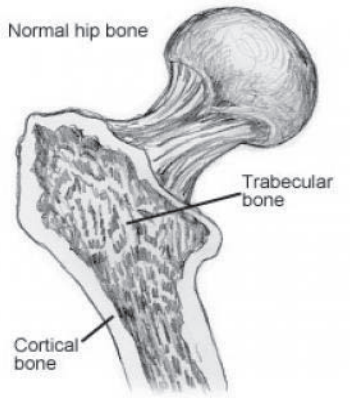


Figure 1. Mature bone composition in the hip. Cortical bone is compact having higher density while trabecular bone is spongy and has a greater surface area.

BONE FORMATION AND DEVELOPMENT

Skeletal formation occurs through two major mechanisms: intramembranous and endochondral ossification[6]. Osteochondral progenitors differentiate directly into osteoblasts to form membranous bone[6]. In contrast, during endochondral ossification, endochondral progenitors differentiate into chondrocytes to form a cartilage template of the future bone [6].

During pregnancy, the mother has to provide sufficient calcium, and optimal levels of vitamin D for the embryo to ensure its normal skeletal development [7]. Throughout skeletal growth, from fetal life until the PBM occurs, bone modeling is the process that shapes skeletal elements for appropriate bone morphology and mass in response to physiologic influences or mechanical forces [3]. Bones may widen or change axis by removal or addition of bone to the appropriate surfaces by independent action of osteoblasts and osteoclasts in response to biomechanical forces [2]. The process of bone maturation is of particular interest, as this thesis is dedicated to bone accrual in children of school age. After achievement of PBM, the main process in bone metabolism consists of bone remodeling. Normally, there is coupling between bone formation and bone resorption carried out by a tightly interconnected group of osteoclasts and osteoblasts, which sequentially replace old bone with new bone [8]. However, as age advances, there is a so-called "uncoupling" of these metabolic processes and less new bone is formed than resorbed in each site remodeled, resulting in bone loss and structural damage[9]. Interestingly, significant sexual dimorphism has been described in both bone accrual and loss, as shown in **Figure 2**.

Osteoporosis is a disease characterized by reduced bone mass and disruption of bone (micro)architecture with an ultimate clinical consequence of increased fracture susceptibility[10]. The increased risk of bone fractures has devastating outcomes for individuals and society in terms of decreased mobility, increased health care costs and ultimately, increased mortality[11]. The prevalence of osteoporosis in the European Union was estimated at 27.6 million subjects in 2010, with its greatest burden arising from osteoporotic fractures comprising the vertebrae, forearm, hip, and proximal humerus. Osteoporotic fractures incapacity and treatment in the 27 countries of the European Union, sum up to an estimated burden of 37 billion euros, costs that are expected to increase by 25% in 2025 [10]. People who have low PBM are at higher risk of osteoporosis forasmuch as any further loss of bone with aging and as a result of hormonal changes will make the skeleton even more fragile and susceptible to fracture. In fact, it is estimated that PBM accounts for more than half of the variability observed in adult bone mass [12], and an increase of PBM by one standard deviation would reduce fracture risk by 50% in women after menopause[13]. Consequently, childhood and youth are crucial periods to determine bone health, and, thus, the knowledge of determinants accounting for variations in bone accrual can be a key factor in the prevention of osteoporosis.

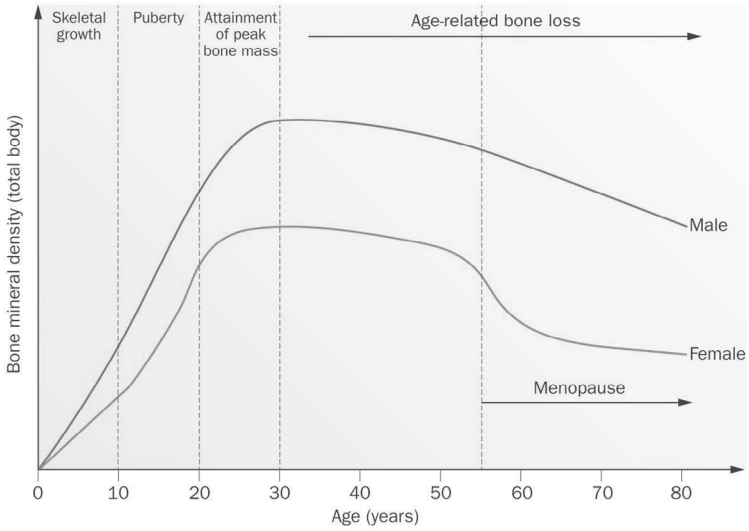


Figure 2. Change of bone mineral density in men and women through life. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Rheumatology 11, 462-474. Hendrickx, G. *et al.* Copyright 2015[14]

BONE MINERAL DENSITY (BMD) ASSESSMENT

Traditionally, bone health in children is assessed using BMD, which is also used to diagnose osteoporosis in adults. BMD is commonly measured by dual-energy X-ray absorptiometry (DXA), a non-invasive technique that directly measures BMD and derives bone area (BA); bone mineral content (BMC) can be quantified as the multiplication of the former parameters[15]. In children, total-body less head (TBLH) and lumbar spine DXA scans are preferred over the hip to evaluate BMD[16] (**Figure 3**). The skull is excluded as its relative contribution to bone mass is proportionally larger with respect to the rest of the body, and its inclusion can mask gains or losses at other skeletal sites making diagnostic interpretation difficult [17]. DXA measurements of the hip usually assessed in adults are not as reliable in children, as this region comprises a relatively small area influenced by changes in size during skeletal development and growth (i.e., bone modeling, remodeling, linear growth and bone expansion). Such limitation on bone age detection and the influence of changes in bone area have high repercussion in the BMD measurement leading to low reproducibility. The DXA lumbar spine measurement covers sufficient area to provide reliable information[15]. Despite this, normative DXA data for the adolescent hip, especially in girls [18], exist and this site might still be useful in addition to the TBLH and lumbar spine DXA measurements.

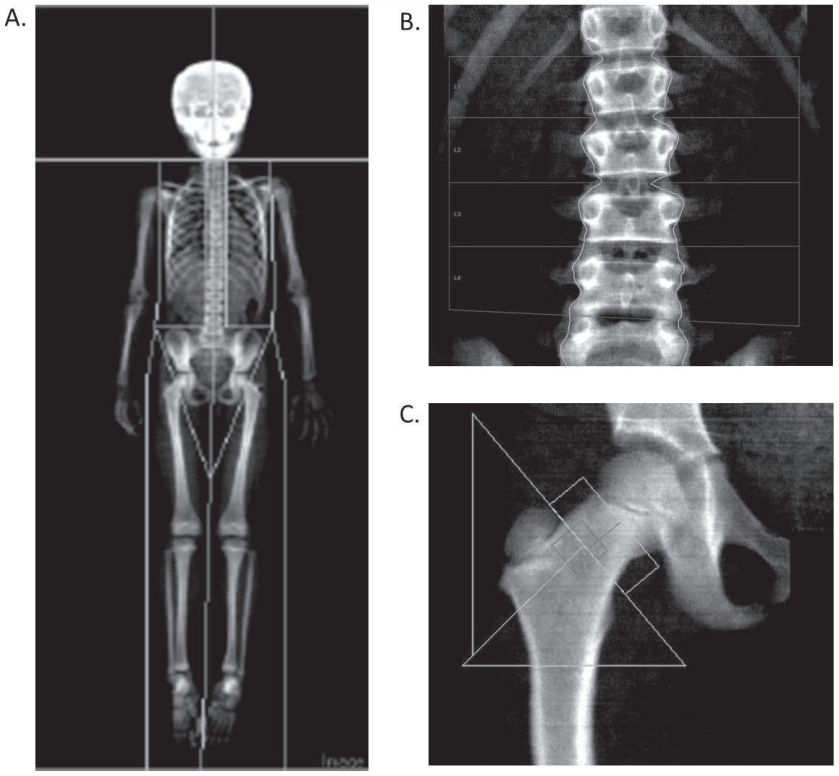


Figure 3. DXA scans commonly used in pediatric and adolescent populations. A. DXA image of Total body scan. Total body less head is the preferred measurement for BMD assessment in children. **B. DXA image of the lumbar spine.** The scan shows regions of interest from L1 to L4**. C. DXA image of the right hip.** The scan shows regions of interest as the femoral neck**.**

FACTORS INFLUENCING BONE GROWTH AND PEDIATRIC BMD

A complex interplay of genetic and environmental factors determines skeletal growth. Factors as intrauterine exposures[19,20], hormones (e.g., sex steroids, insulin-like growth factor I[21,22]), ethnicity[23,24], nutrition (e.g., calcium, vitamin D, protein)[25], physical activity[26], body composition[27,28] have been implicated in pediatric bone growth and density.

There is considerable disagreement in the literature as to the magnitude of the effect of these factors, probably related to the predominance of cross-sectional studies and small sample sizes. Further, twin and family studies indicate that BMD is a highly heritable trait, with heritability estimates ranging between 50 - 85%[29-32]. These estimates can vary due to the analytical model used in their calculation, the skeletal site measured and the population under study. Gueguen and colleagues assessed the heritability of BMD in 129 families and found that heritability estimates increase during childhood and reach a maximum of

85% in early adulthood, being of 63% and 34% at 14 and 53 years, respectively[33]. Hence, a substantial proportion of the variation in bone acquisition and peak bone mass is attributable to genetic factors.

BEYOND BMD, BONE STRENGTH

Bone strength, defined as the capacity of bone to respond to mechanical demands, is ultimately determined by its material composition, quantity and dimension, bone distribution and microarchitecture. BMD typically explains 60–80% of bone strength when bone samples are compared in a laboratory setting under controlled loading conditions[34]. Geometrical bone parameters might explain why many individuals who fracture do not have BMD in the osteoporotic range. For example, a low BMD can be compensated by specific bone structural geometry with no, or minimal compromise to mechanical strength. As DXA-derived images provide two-dimensional measurements, they have low spatial resolution, and no direct calculation of the three-dimensional structure can be made[35]. However, a method applied to DXA scans using biomechanical engineering principles known as hip structural analysis (HSA), allows the measurement of geometric and strength parameters (e.g., cortical thickness, bone width, cross-sectional area, section modulus among others) of the proximal femur. Studies examining the influence of bone geometry and strength on fracture risk have focused primarily on older adults. Even though there are some studies targeting the evolution of these parameters at young ages, most are throughout puberty.

During growth, bone modeling and remodeling optimize strength, by depositing bone where it is needed, and minimize mass, by removing it from where it is not, always in response to strain stimuli imposed by muscle forces. As stated by the Frost Mechanostat[36], the role of muscle in the development of bone strength is predominant, attesting to the importance of physical activity for bone health. In general, changes in bone structure rather than in bone mass have most often been associated with gains of bone strength as a response to physical activity in growing individuals[37]. Due to their rapid and highly dynamic changes, the peripubertal years are now recognized as the most critical period to modify bone structure and strength, traits that tend to track throughout life[38].

GENOME-WIDE ASSOCIATION STUDIES

Single Nucleotide Polymorphism (SNPs) are the most common type of genetic variation. Each SNP represents a difference in a single position on the DNA. The hypothesis-free genome-wide association study (GWAS) strategy enables population-based mapping of

genetic factors affecting complex traits and disease, whereby assessing the possible association of millions of genetic variants in one single experiment is made possible [39]. GWAS methods rely on a dense map of bi-allelic SNP markers that are distributed relatively evenly across the genome to capture most of the common variation. Two major advances have enabled GWAS, namely 1) the knowledge derived from the International HapMap Project [40], and 2) the technological development of high-throughput genotyping with the use of microarrays with dense genotype content. The GWAS process starts with the collection of DNA samples from a large number of individuals, the determination of hundreds of thousands of genotypes using the array technology, followed by methodical high standard steps of quality control. This is then followed by imputation of untyped genotypes. Imputation is a statistical process in which a reference panel of individuals genotyped (or sequenced) containing a dense set of SNPs is used to infer missing information of untyped genetic variants in individuals that have been genotyped for a subset of these SNPs [41]. Then, by means of statistical models, all imputed genetic variants are surveyed for association with disease or quantitative traits. Finally, usually within the setting of consortia, summary-level data of each study is combined to perform a GWAS meta-analysis. (**Figure 4**)

The genetic architecture of complex traits implies that hundreds if not thousands of associated SNPs exist, each contributing just a small amount to the overall trait population variance. It is for this reason that investigators are forced to collect large sample sizes and

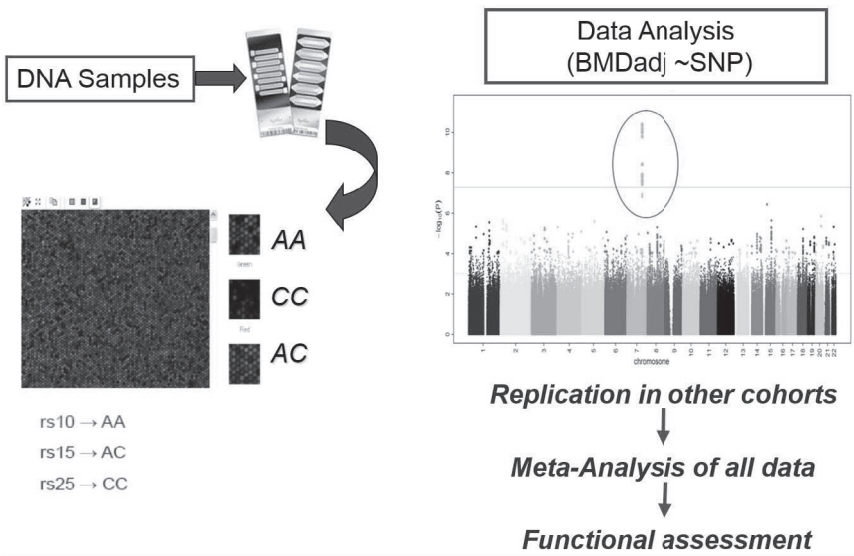


Figure 4. Genome-wide association study scheme. Genotyping of samples is followed by QC and imputation. Then, statistical models are used to stablish association of the genetic variants. Replication of results in other cohorts and meta-analysis are a must. Finally, functional follow-up allows to unveil the action mechanism of the positive findings.

collaborate in consortia bringing together research groups around the globe. So far, GWAS have focused on European populations nonetheless, the inclusion of other ethnic groups as well as admixed population studies is rapidly rising following the pressing need to increase statistical power and to extrapolate findings to non-European populations.

Genome-wide association studies of Bone Mineral Density

Up to date, more than 56 loci have been described as associated with BMD, as determined by GWAS, in adult North-European subjects [42-44], confirming the high complexity of this trait. Before 2012, the genetic analysis of pediatric BMD was restricted to only one analysis lead by the Avon Longitudinal Study of Parents and Offspring (ALSPAC). In that study, association results from a sample of ~5,300 children of white British ancestry were combined with summary data from previously reported GWAS in adult studies. This lead to the identification of an association between TBLH-BMD and genetic variants in the osteoblast transcription factor gene Osterix, an early-acting developmental gene influencing osteoblast differentiation [45]. The pediatric BMD genetic studies from 2012 onwards are presented throughout this thesis.

EVOLUTIONARY STUDIES

GWAS-derived information enables the investigation of the relationship between recent human evolution and human diseases and/or complex phenotypes. The discovery of alleles influencing different traits and their geographical distribution have allowed inquiries into the evolutionary parameters shaping the human natural history. Indeed, many common SNPs present significantly different frequencies among human populations or even appear to be polymorphic just in certain populations (i.e., they are population-specific SNPs) [46], shedding a light into the etiology of different traits. It is clear that the skeletal system as a whole has been shaped by evolutionary forces of natural selection. A good example is the bipedalism of humans, where the ability to walk and run upright on two feet was made possible by virtue of specialized adaptations of the skeleton and muscles [47]. Similarly, the analysis of ancient bones has unveiled important clues of the history of human evolution. Nature has selected materials and structural properties that meet the contradictory needs of strength and lightness, stiffness and flexibility [9]. Modern skeletons appear to be gracile as compared with earlier hominins, plausibly due to a decrease in daily physical activity as a result of technological and cultural innovations [48]. However, even modern populations show high skeletal diversity. Ancient migrations might have played a significant role in shaping the ethnic differences in bone characteristics of modern humans. This diversity might have emerged as different populations worldwide faced different demographic and environmental challenges once they left Africa [47]. Recent genetic and anthropologic studies

have provided evidence for positive selection driving high population differentiation in skeletal traits in modern human populations[47,49-51].

STUDY DESIGN

The majority of studies presented in this thesis arise from data of the Generation R Study, a population-based multi-ethnic prospective cohort study from fetal life until young adulthood, designed to identify early environmental and genetic determinants of growth, development and health[52]. The Generation R Study is conducted in Rotterdam, the Netherlands, within a multi-cultural metropolitan area including inhabitants of approximately 150 different ethnicities. In total, 9,778 women with a delivery date between April 2002 and January 2006 were enrolled in the study. Assessments were planned in early pregnancy (<18 weeks of gestation), mid-pregnancy (18- 25 weeks of gestation) and late pregnancy (≥25 weeks of gestation) and included parental physical examinations and self-administered questionnaires. Additionally, from birth to 6 years of age, data collection was performed in all children by questionnaires and visits to the routine child health centers. DXA scans of the total body and hip were performed on the participants at a mean age of 6 years. Measurements during this visit also included anthropometric assessments.

GENERAL AIM

This thesis contains studies that aimed to disentangle the multifactorial process of bone accretion in children including fetal exposures influencing early programming of the trait, possible effect of body size and composition during growth in bone status and innate genetic factors influencing bone mineral density. I have focused a great part of my research on genetic factors influencing pediatric BMD. Nevertheless, I also encountered methodological challenges intrinsic to running GWAS in general but particularly in multiethnic cohorts. Thus, I acknowledge several methodological aspects of the genetic association testing. In addition, I address major questions generated by the study of genetic variants as a source of phenotypic heterogeneity, taking up relevant issues as age- and site-specific effects, pleiotropy, and ethnic differences on the genetic regulation of bone mass attainment. Finally, I also use an explorative analysis to trace the evolutionary forces that could have shaped the geographic distribution of bone mineral density across populations.

OUTLINE OF THIS THESIS

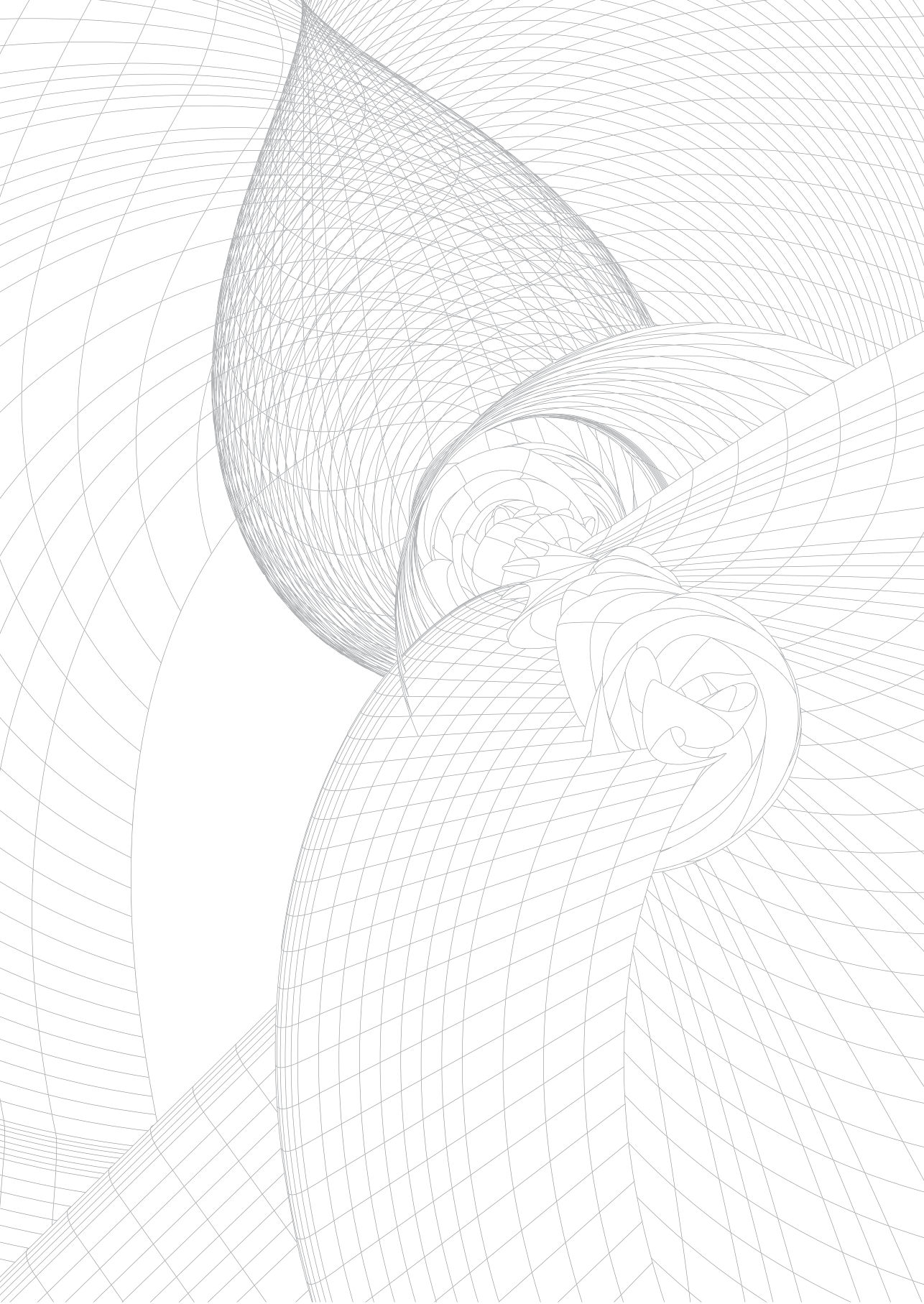
In **Chapter 2** the methodological framework that allowed us to follow up genetic association studies in an admixed population is described. In **Chapter 2.1** an exhaustive description of the generation of the GWAS data and the challenges of the GWAS approach in admixed studies is provided. In **Chapter 2.2** a novel approach to improve the quality of the imputation of rare variants, a crucial point in this new era in the genetic analysis of complex traits is presented. The potential of the genetic data of the multi-ethnic Generation R Study, as a resource in general for genetic studies is exemplified in **Chapter 2.3**, where it serves as control population to a collection of Behcet's cases from different ancestries in a GWAS study. **Chapter 3** focuses on the effect of different factors, from fetus life to pre-school age, influencing BMD measured at 6 years of age in the Generation R Study participants. **Chapter 3.1** is dedicated to the effect of maternal diet on childhood bone mass, and **Chapter 3.2** assesses the effect of fetal and early childhood growth on bone accrual at school age while **Chapter 3.3** describes sex differences in bone strength during childhood. Chapter 4 is dedicated exclusively to genetic studies on pediatric BMD. In **Chapter 4.1** the advantages of applying GWAS in pediatric populations is presented, and *WNT16/CPED1* locus is used to show an example of age-specific genetic effects. In **Chapter 4.2** we extend the analysis of pediatric populations to a general population context. We use DXA derived measures of the TB-BMD as a suitable phenotype for GWAS, discovering up to 18 novel loci associated with this trait in more than 50,000 individuals. We used the same data, meta-analyzed by age strata, to explore further the age-related heterogeneity in the effect of genetic variants on BMD in **Chapter 4.3**. Whereas in **Chapter 4.4** we analyze another type of heterogeneity in the study of bone, i.e., the study of different skeletal sites which differ in their composition of cortical and trabecular bone and are subject to differential loading forces and its relevance for genetic investigations. In **Chapter 4.5** the pleiotropic effects of genetic variants on both lean mass and BMD is studied. In **Chapter 5.1** we go beyond the discovery of new variants by GWAS, and use the GWAS discoveries to characterize at a worldwide level the geographical patterns of BMD-associated SNPs, assessing the possible role of evolutionary selective pressures shaping the ethnic differences in this trait. Finally, **Chapter 6** provides a general discussion in which the studies described in this thesis are placed in a broader context, providing additional implications of these findings and suggestions for future research.

REFERENCES

1. Kini U, Nandeesh BN. Physiology of Bone Formation, Remodeling, and Metabolism. In: Fogelman I, Gnanasegaran G, van der Wall H, editors. Radionuclide and Hybrid Bone Imaging: Springer Berlin Heidelberg; 2012. p. 29-57.
2. Clarke B. Normal bone anatomy and physiology. **Clin J Am Soc Nephrol**. 2008;3 Suppl 3:S131-9.
3. Sims NA, Martin TJ. Coupling the activities of bone formation and resorption: a multitude of signals within the basic multicellular unit. **Bonekey Rep**. 2014;3:481.
4. Baxter-Jones AD, Faulkner RA, Forwood MR, Mirwald RL, Bailey DA. Bone mineral accrual from 8 to 30 years of age: an estimation of peak bone mass. **Journal of Bone and Mineral Research**. 2011;26(8): 1729-39.
5. Rabinovich CE. Osteoporosis: a pediatric perspective. **Arthritis Rheum**. 2004;50(4):1023-5.
6. Yang Y. Skeletal Morphogenesis and Embryonic Development. In: Rosen CJ, editor. Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism. Eighth Edition ed. Ames, USA: John Wiley & Sons, Inc.; 2013.
7. Lieben L, Stockmans I, Moermans K, Carmeliet G. Maternal hypervitaminosis D reduces fetal bone mass and mineral acquisition and leads to neonatal lethality. **Bone**. 2013;57(1):123-31.
8. Baron R, Kneissel M. WNT signaling in bone homeostasis and disease: from human mutations to treatments. **Nat Med**. 2013;19(2):179-92.
9. Seeman E. Pathogenesis of bone fragility in women and men. **Lancet**. 2002;359(9320):1841-50.
10. Hernlund E, Svedbom A, Ivergard M, et al. Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). **Arch Osteoporos**. 2013;8(1-2):136.
11. Fonseca H, Moreira-Goncalves D, Coriolano HJA, Duarte JA. Bone Quality: The Determinants of Bone Strength and Fragility. **Sports Med**. 2014;44(1):37-53.
12. Hui SL, Slemenda CW, Johnston CC, Jr. The contribution of bone loss to postmenopausal osteoporosis. **Osteoporos Int**. 1990;1(1):30-4.
13. Bonjour JP, Chevalley T, Ferrari S, Rizzoli R. The importance and relevance of peak bone mass in the prevalence of osteoporosis. **Salud Publica Mexico**. 2009;51:S5-S17.
14. Hendrickx G, Boudin E, Van Hul W. A look behind the scenes: the risk and pathogenesis of primary osteoporosis. **Nat Rev Rheumatol**. 2015;11(8):462-74.
15. Binkovitz LA, Henwood MJ. Pediatric DXA: technique and interpretation. **Pediatr Radiol**. 2007;37(1): 21-31.
16. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporosis Int**. 2008;19(10):1369-78.
17. Bachrach LK, Sills IN, Section on E. Clinical report-bone densitometry in children and adolescents. **Pediatrics**. 2011;127(1):189-94.
18. Cromer BA, Binkovitz L, Ziegler J, Harvey R, Debanne SM. Reference values for bone mineral density in 12- to 18-year-old girls categorized by weight, race, and age. **Pediatr Radiol**. 2004;34(10):787-92.
19. Steer CD, Tobias JH. Insights into the programming of bone development from the Avon Longitudinal Study of Parents and Children (ALSPAC). **American Journal of Clinical Nutrition**. 2011;94(6 Suppl): 1861S-4S.
20. Sayers A, Tobias JH. Estimated maternal ultraviolet B exposure levels in pregnancy influence skeletal development of the child. **J Clin Endocrinol Metab**. 2009;94(3):765-71.
21. Manolagas SC, O'Brien CA, Almeida M. The role of estrogen and androgen receptors in bone health and disease. **Nat Rev Endocrinol**. 2013;9(12):699-712.
22. Cauley JA. Estrogen and bone health in men and women. **Steroids**. 2015;99(Pt A):11-5.

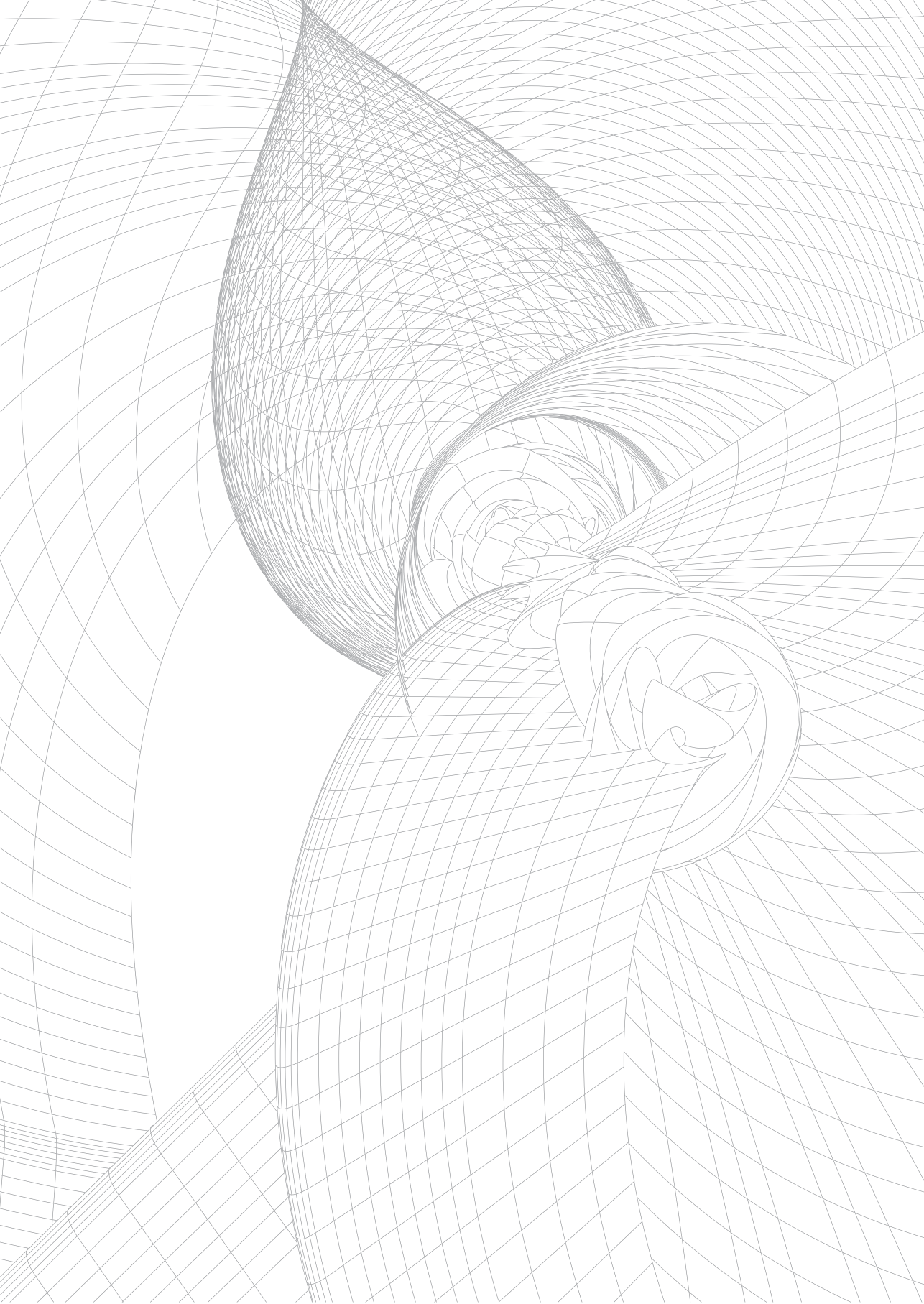
23. Micklesfield LK, Norris SA, Pettifor JM. Determinants of bone size and strength in 13-year-old South African children: the influence of ethnicity, sex and pubertal maturation. **Bone**. 2011;48(4):777-85.
24. Zemel BS, Kalkwarf HJ, Gilsanz V, et al. Revised reference curves for bone mineral content and areal bone mineral density according to age and sex for black and non-black children: results of the bone mineral density in childhood study. **J Clin Endocrinol Metab**. 2011;96(10):3160-9.
25. Moon RJ, Harvey NC, Davies JH, Cooper C. Vitamin D and skeletal health in infancy and childhood. **Osteoporos Int**. 2014;25(12):2673-84.
26. Gunter KB, Almstedt HC, Janz KF. Physical Activity in Childhood May Be the Key to Optimizing Lifespan Skeletal Health. **Exerc Sport Sci Rev**. 2012;40(1):13-21.
27. Ong T, Sahota O, Tan W, Marshall L. A United Kingdom perspective on the relationship between body mass index (BMI) and bone health: a cross sectional analysis of data from the Nottingham Fracture Liaison Service. **Bone**. 2014;59:207-10.
28. Ho-Pham LT, Nguyen UD, Nguyen TV. Association between lean mass, fat mass, and bone mineral density: a meta-analysis. **J Clin Endocrinol Metab**. 2014;99(1):30-8.
29. Arden NK, Baker J, Hogg C, Baan K, Spetor TD. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: A study of postmenopausal twins. **Journal of Bone and Mineral Research**. 1996;11(4):530-4.
30. Slemenda CW, Turner CH, Peacock M, et al. The genetics of proximal femur geometry, distribution of bone mass and bone mineral density. **Osteoporos Int**. 1996;6(2):178-82.
31. Smith DM, Nance WE, Kang KW, Christian JC, Johnston CC, Jr. Genetic factors in determining bone mass. **J Clin Invest**. 1973;52(11):2800-8.
32. Videman T, Levalahti E, Battie MC, Simonen R, Vanninen E, Kaprio J. Heritability of BMD of femoral neck and lumbar spine: a multivariate twin study of Finnish men. **Journal of Bone and Mineral Research**. 2007;22(9):1455-62.
33. Gueguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, Siest G. Segregation Analysis and Variance-Components Analysis of Bone-Mineral Density in Healthy Families. **Journal of Bone and Mineral Research**. 1995;10(12):2017-22.
34. Faulkner KG. Bone matters: are density increases necessary to reduce fracture risk? **Journal of Bone and Mineral Research**. 2000;15(2):183-7.
35. Beck T. Measuring the structural strength of bones with dual-energy X-ray absorptiometry: principles, technical limitations, and future possibilities. **Osteoporosis Int**. 2003;14:S81-S8.
36. Frost HM. Bone "mass" and the "mechanostat": a proposal. **Anat Rec**. 1987;219(1):1-9.
37. Tan VP, Macdonald HM, Kim S, et al. Influence of physical activity on bone strength in children and adolescents: a systematic review and narrative synthesis. **Journal of Bone and Mineral Research**. 2014;29(10):2161-81.
38. Farr JN, Laddu DR, Going SB. Exercise, hormones and skeletal adaptations during childhood and adolescence. **Pediatr Exerc Sci**. 2014;26(4):384-91.
39. Risch N, Merikangas K. The future of genetic studies of complex human diseases. **Science**. 1996; 273(5281):1516-7.
40. International HapMap C. The International HapMap Project. **Nature**. 2003;426(6968):789-96.
41. Hu YJ, Lin DY. Analysis of untyped SNPs: maximum likelihood and imputation methods. **Genet Epidemiol**. 2010;34(8):803-15.
42. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
43. Zheng HF, Forgetta V, Hsu YH, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. **Nature**. 2015;526(7571):112-7.
44. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. **Nat Genet**. 2009;41(11):1199-206.
45. Timpson NJ, Tobias JH, Richards JB, et al. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. **Hum Mol Genet**. 2009;18(8):1510-7.

46. Marigorta UM, Lao O, Casals F, et al. Recent human evolution has shaped geographical differences in susceptibility to disease. **BMC Genomics**. 2011;12:55.
47. Wu DD, Zhang YP. Positive selection drives population differentiation in the skeletal genes in modern humans. **Hum Mol Genet**. 2010;19(12):2341-6.
48. Chirchir H, Kivell TL, Ruff CB, et al. Recent origin of low trabecular bone density in modern humans. **Proc Natl Acad Sci U S A**. 2015;112(2):366-71.
49. Betti L, Cramon-Taubadel NV, Lycett SJ. Human pelvis and long bones reveal differential preservation of ancient population history and migration out of Africa. **Hum Biol**. 2012;84(2):139-52.
50. von Cramon-Taubadel N. Evolutionary insights into global patterns of human cranial diversity: population history, climatic and dietary effects. **J Anthropol Sci**. 2014;92:43-77.
51. Wu DD, Jin W, Hao XD, Tang NLS, Zhang YP. Evidence for Positive Selection on the Osteogenin (BMP3) Gene in Human Populations. **Plos One**. 2010;5(6).
52. Jaddoe VW, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. **Eur J Epidemiol**. 2012;27(9):739-56.



Chapter 2

GWAS methods: from data collection to analysis



Chapter 2.1

Practical and Methodological Challenges in Conducting Genome-Wide Association Studies in Highly Admixed Multi-Ethnic Populations: the Generation R Study

Carolina Medina-Gomez, Janine F Felix, Karol Estrada, Marjolein J Peters, Lizbeth Herrera, Claudia J Kruithof, Liesbeth Duijts, Albert Hofman, Cornelia M. van Duijn, André G Uitterlinden, Vincent WV Jaddoe and Fernando Rivadeneira

Eur J Epidemiol. 2015 Apr;30(4):317-30. doi: 10.1007/s10654-015-9998-4.

ABSTRACT

Genome-wide association studies (GWAS) have been successful in identifying loci associated with a wide range of complex human traits and diseases. Up to now, the majority of GWAS have focused on European populations. However, the inclusion of other ethnic groups as well as admixed populations in GWAS studies is rapidly rising following the pressing need to extrapolate findings to non-European populations and to increase statistical power. In this paper, we describe the methodological steps surrounding genetic data generation, quality control, study design and analytical procedures needed to ran GWAS in the multiethnic and highly admixed Generation R Study, a large prospective birth cohort in Rotterdam, the Netherlands. Furthermore, we highlight a number of practical considerations and alternatives pertinent to the quality control and analysis of admixed GWAS data.

INTRODUCTION

Genome-wide association studies (GWAS) analyze a large number of single nucleotide polymorphism (SNPs) across the genome in a large number of samples, aiming to identify loci associated with complex traits at the population level. Since 2007, well-designed studies have been able to comprehensively test common genetic variation across the genome [(1)]. Up to now, at least 11,680 SNPs have been robustly associated with one or more complex traits, providing biological insight in traits as different as Alzheimer's disease, prostate cancer, inflammatory bowel disease, obesity, stroke, diabetes, asthma, height, cholesterol levels and bone mineral density, to name only few examples of the successful performance of this approach [(2)].

Variants discovered by GWAS typically have small effects which is why minor sources of systematic or random error can lead to false positive associations or can mask real effects (false negative associations). In order to avoid bias, it is necessary to closely control the processes underlying the production of GWAS data, extending from laboratory processes (data generation) to imputation. It is also necessary to conduct statistical analyses, which incorporate factors into the models known to influence the trait of interest, as well as being appropriate to the characteristics of the study design.

Ethnicity is a confounder of epidemiological studies which incorporates cultural, geographical and biological dimensions. In the GWAS context spurious associations between genetic variants and a trait of interest occur when both allele frequencies and differences in trait distributions (disease prevalence or magnitude of quantitative traits), vary across ethnicities. From this perspective, adequate correction for potential population stratification is required for successful identification of genetics determinants of complex traits and diseases.

To date, GWAS have mainly focused on populations of European ancestry. Consequently, having another ethnic background is a common reason for exclusion of GWAS samples. As an illustration, from the 1,734 GWAS papers indexed in the GWAS catalogue, 66% included only individuals from European ancestry, 34% included Non-Europeans only (most of those carried out in Asian populations), and 12% included both Europeans and Non-European individuals [(3)]. Moreover, big consortia efforts such as Cohorts for Heart and Aging Research (CHARGE) or the Genetic Investigation of Anthropometric Traits (GIANT) have focused primarily on European populations, while efforts driven in populations of diverse ethnic background are of modest sample sizes. However, the inclusion of multiethnic and/or admixed populations in the analysis of GWAS can actually result in additional power. Firstly, larger datasets (representing higher power) can be assembled when the ancestry criterion

is not used for sample exclusion. Keeping such "ethnic outliers" in the study also represents a better use of resources considering the logistic and burden behind sample collection and genotyping, and their associated costs. Secondly, the European-only approach has little power to detect association for genetic variants segregating at low frequency in European populations and statistical power can be gained if those variants are more common in other ancestries included in the analysis [4]. Some examples are provided by Fu et al., who describe variants associated with type 2 diabetes mapping to *UBE2E2* and *KCNQ1* that have higher frequencies in East Asians (minor allele frequency (MAF) of 0.22 and 0.38, respectively) as compared with Europeans (MAF of 0.093 and 0.08, respectively) [5]. Similarly, Wu et al. showed examples of ethnic specificity in variants associated with lipid levels mapping to *APOA5* and *APOB*. These very rare variants identified in African-Americans were not detected in either East Asian or European populations[6]. Further, (rare) variants specific to a subpopulation (e.g., a diabetes susceptibility variant arising in Native-Americans) can be identified in a derived highly admixed population (i.e., Mexicans) as having the largest effect[7]. In addition, population admixture, due to interbreeding of individuals from different origins, would have brought together genomes from continental populations, which are a product of genetic drift and different selective pressures. Following this line of reasoning, it is expected that recently admixed populations are likely to harbor a larger number of genetic variants than the original populations they come from[8]. Theoretically, this will result in a higher yield in the discovery of genetic determinants of complex traits. Another important genetic approach, suitable in admixed population, to identify disease risk variants is admixture mapping, which is powerful when the ancestral populations differ both in allele frequencies and disease prevalence. Then, in the vicinity of a disease locus, an affected individual should have a higher proportion of alleles inherited from the most affected ancestral population[9].

As GWAS worldwide are expanding to include multi-ethnic and admixed populations, we describe here the steps used for genetic data generation, study design and analytical procedures applied in the Generation R Study. The Generation R Study is a population-based prospective cohort following children and their mothers from fetal life onwards, which comprises a multiethnic population, including a high proportion of highly admixed individuals. We here mainly focus on how this approach can allow analyzing the whole set of individuals independent of genetic background as a mean to increase sample size and power to identify loci underlying complex traits and diseases. The different considerations described here can be applied to other multiethnic studies, particularly those of admix nature such as Hispanics or African Americans.

MATERIALS AND METHODS

Study Population

The Generation R Study is a multi-ethnic population-based prospective cohort study, spanning from fetal life until young adulthood, designed to identify early environmental and genetic causes of normal and abnormal growth, development and health during fetal life, childhood and young adulthood. Study design, data collection in prenatal and postnatal phases, and ethical issues of this study have been previously described in detail [10]. The Generation R Study is conducted in Rotterdam, the Netherlands, within a multi-cultural metropolitan area. The study area includes inhabitants of approximately 150 different ethnicities [11]. Pregnant women with a delivery date between April 2002 and January 2006 were informed about the study and provided written informed consent through their prenatal care provider during their first visit of the pregnancy. The medical ethics committee of Erasmus University Medical Center approved the study. In total, 9,778 mothers were enrolled in the study.

The ethnic background of the children was defined by the parents' country of birth, which was obtained by questionnaire. The participating child was defined as of non-Dutch ethnic origin if one of her/his parents was born abroad, and further classified using a socio-demographic definition as described by Statistics Netherlands [11]. If both parents were born abroad, the country of birth of the mother decided the classification of the ethnic background of the child. The ethnic background of the mother and partner were obtained in the same manner, based on their parents' (the child's grandparents) country of birth.

Sample Collection, Biobanking and Genotyping

Blood samples of the children were collected from the umbilical cord at birth. Where an umbilical cord blood sample could not be collected at birth, a blood sample was obtained by venipuncture during the child's visit to the research center at a mean age of 6 years. All samples were coded with a unique laboratory number. Umbilical cord samples were collected in 10 ml EDTA tubes and stored immediately at -80°C, while samples obtained by venipuncture were collected in 5 ml EDTA tubes and stored directly after transport at -20°C. DNA was extracted manually from white blood cells using the Qiagen FlexiGene Kit (Qiagen Hilden, Germany). Normalisation and further processing of the DNA samples were performed on a Caliper ALH3000 pipetting robot. A detailed description of the Generation R Biobank has been previously published [12].

Genotyping was performed using Illumina HumanHap 610 or 660 Quad chips -depending on collection time- following manufacturer protocols, and intensities were obtained from the BeadArray Reader. Genotype calling was performed on normalized intensities using

the Genecall module from the Illumina Genome Studio software version 1.1.0.28426. A no-call threshold of 0.15 was applied to a manufacturer-provided cluster file. Illumina Genome Studio provides a quality metric used to identify low-quality samples and, we used a threshold of 97.5% for exclusion of samples.

DNA Quality Control (QC)

The two Genome Studio projects (one each for the HumanHap 610 array and for the 660 array), were merged using SNPs common to both arrays. The QC procedures were applied to the genotyped data using PLINK [13] in two phases: marker- and sample-based.

Marker QC included filters for: 1) marker call rate (calling <.2 - <0.05, --geno option), checked in two rounds, the initial with a threshold of 80% and the second one more stringent (95%), after inspection of sample quality, 2) minor allele frequency (MAF <=0.001, --maf option), 3) differential missingness between the two projects ($P<1x10^{-7}$, --test-missing option) and 4) deviation from Hardy-Weinberg equilibrium proportion ($P<10^{-7}$ --hwe option). Sample QC included: 1) duplicate detection (PLINK option IBS=1), 2) sex discordance rates (--check-sex option), comparing the reported sex of each participant with the sex predicted by the genetic data (expected chromosome X heterozygosity). When results

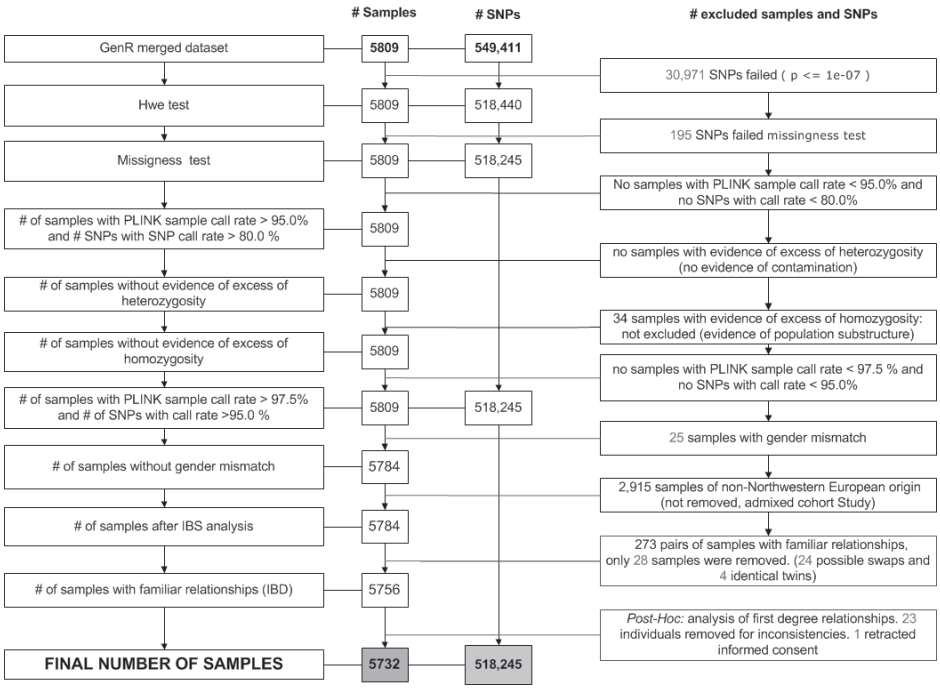


Figure 1. Flowchart overview of the entire GWAS QC process. Quality control of all samples from Generation R-1 and Generation R-2 after merging of the projects. Red font denotes exclusion of either SNPs or samples from the dataset in the different QC steps.

were inconclusive, the Genome Studio plots, log R ratios and B-allele frequencies, for both X and Y chromosomes were inspected. 3) Genotype call rate (<0.05 - <0.025--mind option) checked in two rounds, the initial with a threshold of 95% and the second one more stringent (97.5%), after inspection of marker quality and 4) high heterozygosity rate, over 4 SD of the mean heterozygosity of all samples (--het option). The step by step summary of the applied QC pipeline is presented in Figure 1, and Online Resources Figures 1 and 2.

Population sub-structure and family relationships

Additional sample QC assessments were applied to determine genetic-based ethnic background and to identify potential family relationships.

Genetic ancestry

To characterize the genetic ancestry of the children in the Generation R Study, all samples passing QC procedures were merged with the three genotyped panels from the HapMap Phase II release 22 build 36 including: Northwestern Europeans (CEPH collection or CEU), Sub-saharan West Africans (Yoruba or YRI) and Asians (Han Chinese from Beijing or CHB, and Japanese from Tokyo or JPT) [14,15] using only independent autosomal SNPs (r^2>0.05). In the merged dataset, pairwise identity-by-state (IBS) relations were calculated for each pair of individuals (representing the average proportion of alleles shared by those individuals) using PLINK (–genome option). In addition, principal axes of variation (or so-called genomic components equivalent to Principal Components (PCs)) were derived from this IBS matrix by Multi-dimensional Scaling (MDS), to characterize the variability present in the data using few variables (PLINK –-cluster -–mds-plot). Participants were defined as being of non- Northwestern European ancestry when deviating more than 4 standard deviations (SD) from the CEU panel mean value in any of the first four genomic components.

Sample relatedness

To identify cryptic family relationships within the Generation R samples, we first removed the HapMap samples, recalculated the IBS matrix including only participants of the Generation R Study and then determined pairwise, the proportion of shared IBS alleles. By using this information and the population allele frequency, PLINK is able to estimate the number of these alleles coming from the same ancestor, known as IBD (identity-by-descent), using the methods of moments [13]. These familial relationships detected using PLINK, were validated *post-hoc* using the recently released software REAP (Relatedness Estimation in Admixed Populations). REAP estimates IBD proportions in a similar way than PLINK. Nonetheless, it uses individual-specific allele frequencies that are calculated by conditioning on estimated individual genome-wide ancestry [16].

Genotype Imputation

A two-step genotype imputation, comprising a phasing step to resolve the haplotypes of the genotyped markers (using MACH software) and an imputation step in which unmapped SNPs are imputed to a reference panel (using Minimac software), was applied to the GWAS genotyped dataset after QC. Data was divided in marker sets across chromosomes to be processed using a parallel computing cluster. Thus far, two different reference panels have been used to impute the Generation R data: a) HapMap Project Phase II Release 22, build 36 phasing and b) 1000 Genomes Project (phase III release version), build 37 phasing. For the phasing of the haplotypes we used the standard parameters recommended by the MACH/minimac developers (http://genome.sph.umich.edu/wiki/Minimac) consisting of 20 iterations of the Markov sampler and 200 states (number of haplotypes that should be considered when updating each individual). Parallelization was achieved by splitting jobs on chunks across chromosomes. The window size was 2,100 markers, of which 100 were flanking markers, when using assembly build 36, and a window size of 2,500 markers, using 500 as flanking markers when using assembly build 37. Imputations were performed following the same chunking strategy of parallelization as mentioned for the phasing step. To evaluate genotype imputation quality we used the MACH r-squared (Rsq), metric based on the ratio of the empirically observed variance of the allele dosage to the expected binomial variance f(1 – f), assuming Hardy–Weinberg equilibrium, where f is the observed allele frequency. When imputations hold adequate information for predicting the unobserved genotypes from the observed haplotype backgrounds, this ratio should be distributed around unity[17]. By consensus an Rsq>0.3 has been used to define sufficiently good quality for analysis [18].

HapMap Imputations

Imputations of autosomal chromosomes to HapMap used all haplotypes available from Phase 2 of the International HapMap Project reference panel, in the so-called "cosmopolitan approach". This combined reference panel includes 210 individuals from the CEU, YRI and CHB/JPT panels [15].

1000 Genomes Imputations

A second round of imputations was performed using 1000 Genomes (1KG) data - phase 3 release (http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G.2012-03-14.html), which comprises the genomes of 1,092 individuals from 14 populations [19]. We employed the same parameters as described for the phasing procedure in build 37, and included autosomal and chromosome X markers. Chromosome X imputations were performed separately for males and females.

Genome Wide Association analysis in the Generation R Study

For illustration of possible pitfalls when using an admixed population, association analyses in the Generation R Study were performed with and without adjustment for population substructure. Additionally, we evaluated the distribution of the participant's ethnicity along the genomic components, in order to assess the adequacy of questionnaire-based ethnicity to correct for population structure in the association models. Finally, we contrasted the two most common approaches used for correction of population stratification: 1) the traditional method of inclusion of genomic components as covariates in the association model, and 2) linear mixed models, as implemented in the publicly available software, Efficient Mixed-Model Association eXpedited (EMMAX) [20]. The genome-wide significance (GWS) threshold for the association was established at $P<5x10^{-8}$. For illustration, we present here examples of association results in the whole Generation R population obtained for two model phenotypes: 1) the dichotomous red hair pigmentation (a highly stratified trait) and 2) the continuous bone mineral density measured at the skull. For the former example (n=5,731), logistic association analyses ran on directly genotyped markers were corrected by four genomic components. Additionally, we used the two imputed datasets – HapMap and 1KG- to show the fine mapping resolution improvement of the genome-wide signal. For the latter skull BMD analyses (n=4,086), linear association using HapMap imputed data including twenty genomic components as covariates in the model. Further details on collection and analysis of this phenotype have been reported elsewhere [21]. For further illustration, we ran GWAS for skull BMD with equal sample sizes (n=1,909) in both the non-European and a randomly selected sample of the European subgroup of the Generation R Study, adjusting for twenty genomic components and compared results for rs13223036, reported as the top-hit in a meta-analysis of more than 9,000 kids mainly from European ancestry [21].

All linear and logistic models were ran using the MACH packages (http://www.unc.edu/~yunmli/software.html) as available in the web-based tool GRIMP, which facilitates high-speed analysis of large imputed datasets making use of computational grid infrastructures [22].

RESULTS

Study Population

A summary of the ethnic classification based upon questionnaire of the 9,749 children participating in the Generation R Study is presented on Online Resource 3. Ethnic classification was missing in 6.7% of the population. The largest ethnic groups in the cohort were of Dutch (57%), other European (7.4%), Turkish (7.4%), Surinamese (7.3%), and Moroccan (6.4%).

Sample Collection, Biobanking and Genotyping

At birth, 5,908 samples were obtained from 30 ml cord blood (Generation R-1). Additionally, 320 samples were drawn during the visit to the research center at age 6 years (Generation R-2). Of the DNA samples from the collection of Generation R-1, 5,815 (98.4%) were genotyped using the Illumina HumanHap 610 Quad chip (including 620,901 markers, representing 592,532 SNPs and 28,369 copy number variation (CNV) probes). The 1.6% of samples not genotyped were discarded either for low quantity and/or low concentration of DNA in the stock solution as well as possible unresolved sample swaps. The extra 320 DNA samples in Generation R-2, were genotyped with the Illumina HumanHap 660 Quad chip (comprising 657,366 markers, 561,490 SNPs and 95,876 CNV probes). The genotype data were exported on forward strand for both collections. A total of 178 samples with genotyping rates lower than 97.5 % (Genome Studio sample call rate), likely arising due to low DNA quality, array problems or poor performance of agents, were excluded from the final projects (Generation R-1 and Generation R-2 sets).

DNA Quality Control

Marker QC

CNVs reported in the manifests of the arrays, together with SNPs which could not be called in at least 95% of the samples or with a MAF <=0.001, were eliminated (Online Resources 1 and 2) before merging the Generation R-1 and Generation R-2 sets. The combined dataset, merged using only SNPs common to both platforms (n=5,809), consisted of 549,511 SNPs. No SNP was excluded in any of the two call-rate inspections. One hundred and ninety-five SNPs were removed due to differential missingness, addressing possible bias induced by batch effects between the sets. Improvements to our quality control pipeline could be implemented, as to have more stringent standards. For example, although PLINK will report alleles incompatibilities when merging datasets, these would not be detectable in case of palindromic SNPs (A/T and G/C). Therefore, since strand issues would not be detected for these type of SNPs checking allele frequencies before merging is strongly recommended. In addition, 30,971 SNPs were excluded for deviations from Hardy-Weinberg Equilibrium (HWE) proportions ($P<1x10^{-7}$). While other causes of deviation exist, failure of this test is highly indicative of genotyping errors at a given marker [23].

Sample QC

Unique laboratory codes together with an anonymous person-unique study code were compared in order to identify duplicates. Fifteen duplicated samples were removed from analysis (10 from the Generation R-1 set and 5 from the Generation R-2 set). Sex inconsistencies were flagged by PLINK in 60 samples. Ten of them had incompatible sex data while the others were assessed as ambiguous. After revision of Genome Studio plots (Online

Resource 4), we identified discrepancies for 15 of those samples. In total, 25 samples were excluded during this sex check. A Sample genotyped call rate test, based on the remaining SNPs after merging projects, resulted in no samples exclusion. We found no individual samples with excess of heterozygosity of more than 4 SDs above the mean heterozygosity value of all samples, thus the presence of sample cross-contamination was unlikely. However, reduced heterozygosity (- 4SDs) was identified in 34 samples, possibly as result of the multiethnic background of the samples. Excess of homozygosity is typically seen in individuals from genetic isolates with large stretches of linkage disequilibrium (LD) or in populations with substructure, in which there is partial admixture as result of non-random mating, as is the case in the Generation R Study [24,25].

Population sub-structure

Genetic ancestry

Generation R and the three HapMap panels were merged based on a common set of 36,845 independent (LD-pruned) autosomal SNPs. After calculation of pairwise IBS genetic distances between all individuals, we derived genomic components, summarizing the structure of the data into main genomic components explaining the genetic variation (Figure 2). Approximately 50.5% of the samples deviated more than 4 SDs from the mean CEU panel cluster on the main four components and were classified as of "Non- Northwestern

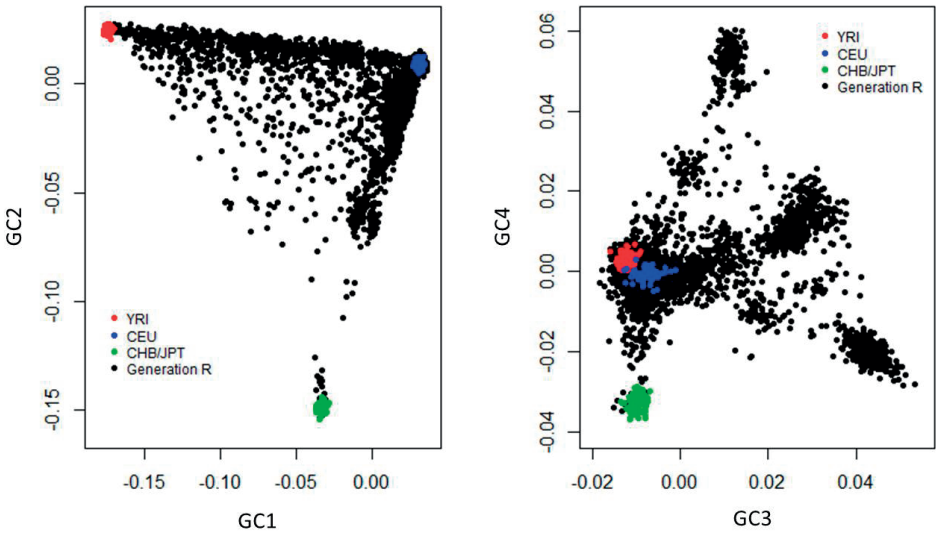


Figure 2. Genetic substructure of the Generation R Study. Two-dimensional plots from MDS analyses of the Generation R Study and the three initial Panels form the HapMap Project. Left panel: First two components explaining most of the variability of the data. Right panel: Third and Fourth components explaining some of the remaining variability.

European" origin. A previous release of the Generation R data from 2009, included only individuals whose samples were collected at birth, and who were classified as of North-western European origin (N=2,661) following the same steps mentioned above, and has been used in some publications [10,26,27].

Cryptic family relatedness

Two-hundred and eighty-nine possible pairwise sib-ships were found by IBS-sharing using PLINK (0.35<PI_HAT<1). Sixteen pairs of individuals shared two alleles at every locus corresponding either to monozygotic twins or a single sample processed twice. Twelve of these relations were conflicting with the registry, and thus most likely correspond to the same sample being processed twice. In these cases, both samples were removed from the dataset. The four remaining pairs were twins when traced back to registries. For these true twin pairs, the sample with a lower call rate was removed from the dataset. First-degree relationships discrepant with registry (13 samples) identified using PLINK were not initially excluded. Nevertheless, they were excluded after confirmation by REAP. Visualization of kinship coefficients obtained from REAP revealed that Generation R participants are (to a large extent) unrelated. Sibling pairs are represented by the small peak around a kinship coefficient of 0.25. Yet another peak (0.025< kinship coefficient <0.0635) evidence the presence of third and fourth degree related individuals (Online Resource 5). Related individuals were not removed from the dataset to allow exclusion/inclusion in association analyses to be done specifically by phenotype availability. In addition, one more individual was recently removed for retracted informed consent. In summary, the current GWAS collection for the Generation R Study consists of samples from 5,732 children.

Genotype Imputation

HapMap imputations

Using the three HapMap panels combined 3,021,329 SNPs were imputed. The MAF distribution of imputed SNPs is shown in Figure 3. The mean Rsq for all the imputed data was 0.883, (median 0.972, IQR = 0.127); when markers with MAF<0.01 were excluded (comprising 313,593 SNPs or 10.38% of the markers), the mean Rsq was 0.914, (median 0.979, IQR=0.083). Figures 3A and 3B show how the increase of Rsq is proportional to the increase in the MAF of the markers. When grouping the markers into MAF bins, 83% of the SNPs with MAF<0.01 achieved sufficient quality, while for the other bins more than 95% of the SNPs were well imputed. Nonetheless, there is a broad range of quality scores for SNPs in each MAF bin. Statistical dispersion is decreasing with MAF as seen by the interquartile range represented by the size of the box in each bin. Patterns of imputation quality by chromosome are shown in Online Resource 6. In general, larger chromosomes tended to be better imputed. Imputation quality was visually checked across chromosomes and

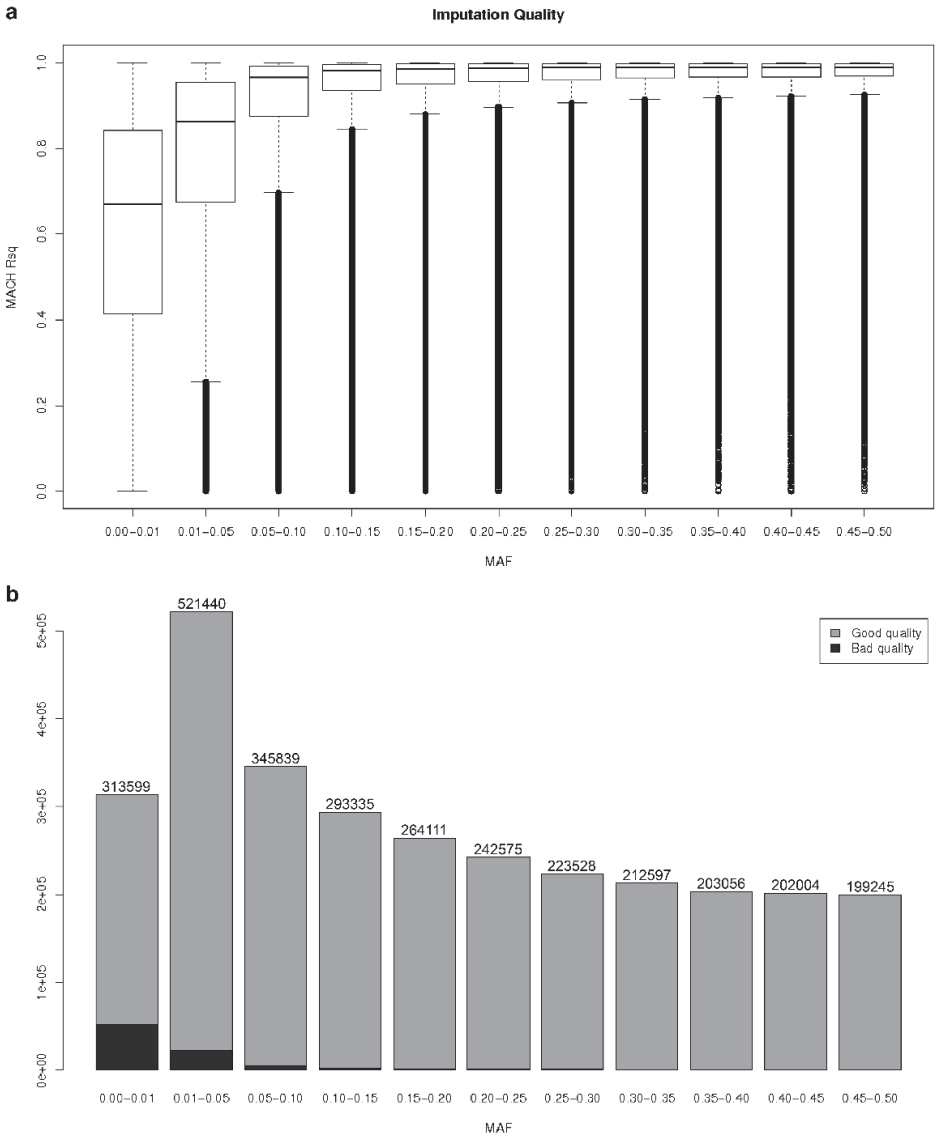


Figure 3. Imputation Quality metrics evaluation HapMap. a. Boxplots of the MACH Rsq in function of the MAF of the imputed SNPs. **b.** Imputation quality distribution per MAF category. Blue and green denotes the poorly and well imputed SNPs based in a 0.3 quality score as threshold. 88,625 out of 3,021,329 (2,93%) are poorly imputed SNPs (Rsq<0.3).

the only notorious fall in Rsq was at centromeres and extremes of the telomeres, where the density of markers is low. Markers on the sex chromosomes were not imputed to the HapMap reference panel.

1KG Imputations

We were able to impute 30,072,738 autosomal variants using the 1KG reference panel, in which 28,681,763 are SNPs and 1,390,975 are insertion/deletions. The mean Rsq for all variants was 0.574 (median 0.622, IQR = 0.636); when markers with MAF<0.01 were

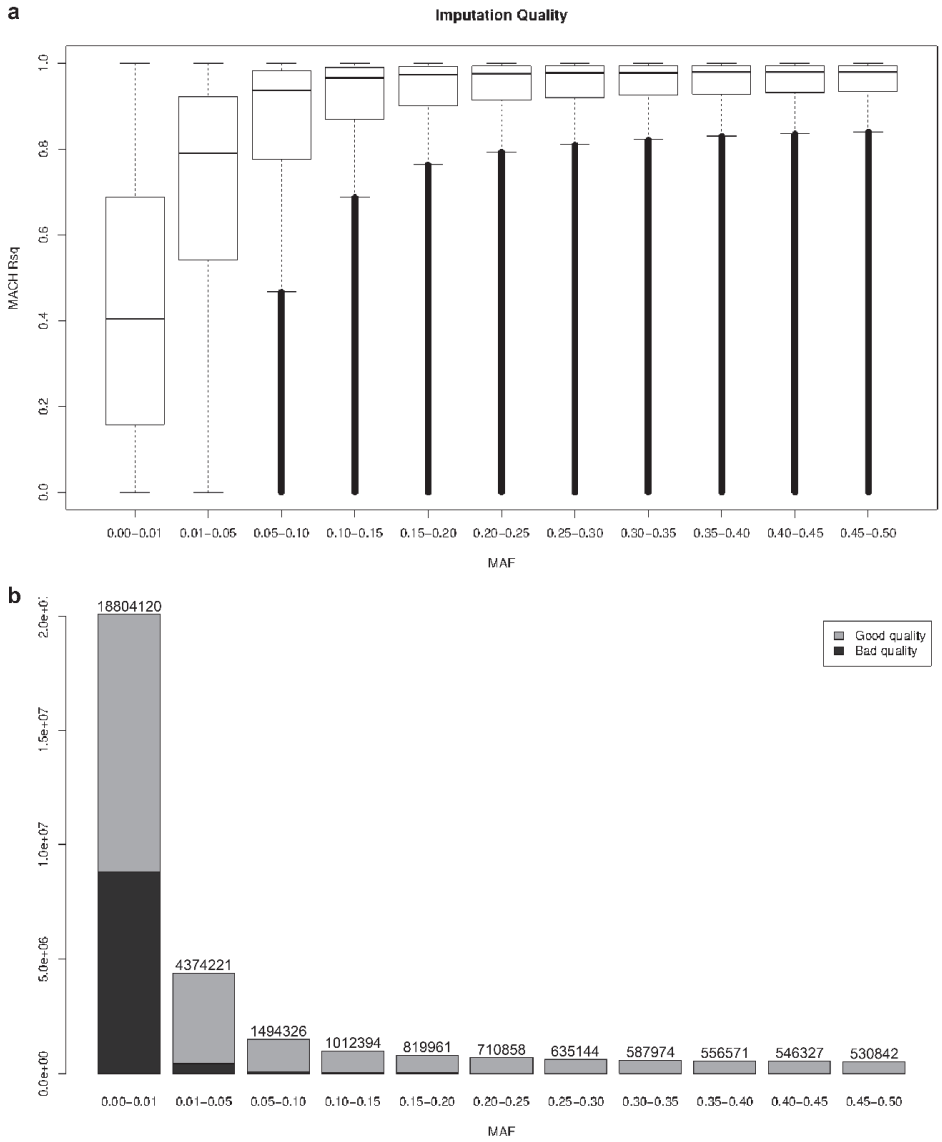


Figure 4. Imputation Quality metrics evaluation 1KG. a. Boxplots of the MACH Rsq in function of the MAF of the imputed SNPs. **b.** Imputation quality distribution per MAF category. Blue and green denotes the poorly and well imputed SNPs based in a 0.3 quality score as threshold. 8,263,752 out of 30,072,738 (27.4%) are poorly imputed SNPs (Rsq<0.3)

excluded (comprising 18,804,120 SNPs or 62.52% of the markers), the mean Rsq increased to 0.815 (median 0.929, IQR= 0.244). Figure 4 shows an assessment of imputation accuracy by MAF. Although imputation quality was poor in the lower spectrum of allele frequencies (MAF<0.05), 15,164,960 markers had an Rsq >= 0.3 and were suitable for analysis. Moreover, the number of markers comprising bins of common frequency (6,894,397 markers with MAF>0.05) is much lower than the number of markers comprising bins of low frequency (23,178,341 markers with MAF<0.05), which usually have low imputation quality. Online Resource 6 summarizes the performance of the imputation per chromosome. The number of SNPs imputed on chromosome X was 1,264,877, of which 903,868 (71.5%) were rare (MAF<0.005). As expected, quality was not as high as for the autosomal chromosomes, as a consequence of the lower number of haplotypes contributed by men in this chromosome. Considering markers of sufficient imputation quality (Rsq>=0.3) on the autosomal chromosomes only, the 1KG imputation resulted in 18,874,123 more markers than those arising from the HapMap imputations including 7,892,440 markers with a MAF>0.01. There are minimal differences in imputation quality when comparing the 2,972,940 SNPs common across the two datasets (mean Rsq, 0.886 (median=0.972, IQR=0.123) for the HapMap imputed dataset against 0.903 (median=0.978, IQR=0.097) in the 1KG imputed dataset). When further filtering markers for MAF>0.01 and Rsq>=0.3, (resulting in 2,671,742 SNPs) the concordance rate, based on best guess genotypes, between the Hapmap and the 1KG imputed datasets was 0.983 as calculated by PLINK (using the --merge-mode 7 option).

Correcting Genome Wide Association Analysis for ethnic background in the Generation R Study

The socio-demographic ethnic definition in the Generation R Study is based on country of birth of the parents of the participants. To evaluate the degree of potential misclassification between definitions, we assessed the agreement of the questionnaire definition with that of genetic ancestry, using genomic components (Online Resource 7). Groups classified as being of European and Dutch origin have historically undergone high waves of migration during the 1960s, 1970s or early 1980s. As a consequence, a scattered distribution across the genomic components axes was observed instead of a uniform one. A similar pattern was also observed for participants of Surinamese origin, arising from two clearly differentiated ethnic groups, the Hindustani and the Creoles.

Statistical approaches based on EMMAX and genomic components were tested for two different traits. There is no evidence of major degrees of residual population stratification in the GWAS results for red hair color (Figure 5 and Online Resource 8), within the Generation R Study (196 children with red hair (3.4%) as gauged in the QQ-plots (no early deviation from the test statistic or p-value distribution) and genomic inflation factors (GIF) close to unity for both EMMAX (GIF= 0.994) and genomic components correction (GIF=0.999). In

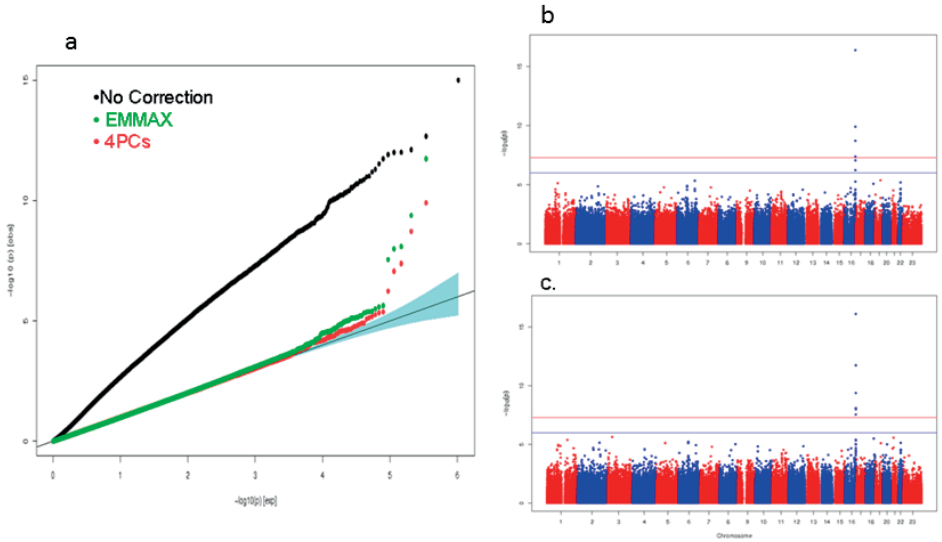


Figure 5. Genome-wide association of red-hair pigmentation in the Generation R cohort. a. Q-Q plot showing the inflation of the test statistics when correction for data structure is not applied (black dots) and the slightly lower power when Genomic Components correction is applied (red dots) in comparison with the EMMAX model (green dots) and **b.** Manhattan Plots of the red-hair pigmentation GWAS in the Generation R Study using adjustment for Genomic Components **c.** Manhattan Plots of the red-hair pigmentation GWAS in the Generation R Study using a Linear Mixed Model as implemented in EMMAX

contrast, when no adjustment for population stratification was employed, very early (artefactual) deviation was seen in the QQ plot, erroneously indicating that the vast majority of markers across the genome were associated with red hair pigmentation (Figure 5). After correction for population stratification, only the markers on chromosome 16q24.3 mapping in the vicinity of *MCR1* reached GWS, variants in this gene largely explain the presence of red hair pigmentation [28]. GWAS based on the imputed data gave rise to similar results, but showed an even higher number of SNPs underlying the *MCR1* associated signal. Furthermore, the leading SNP on these analyses was a missense variant rs1805007, $P < 1x10^{-20}$, reported previously as associated with this trait [29], which was not present in the genotyped data (Online Resource 9). QQ-plots from the skull BMD GWAS show adequate correction for population stratification (Online Resource 10). Power for both EMMAX and genomic components is similar in the two tested traits, as gauged by the number of GWS signals and their significant level (Online Resources 8 and 11). Moreover, the effect size of skull BMD associated SNPs is practically identical across the two approaches.

Skull BMD analysis for equal sets of European and Non-European children shows a GWS signal in the *WNT16/CPED1* locus only in the Non-European children, although similar direction for the leading SNPs was found in both sets (Online Resource 12). We compared the association results for rs13223036 in this locus. The frequency of the effect allele in

Europeans was 0.622, while it was 0.695 in non-Europeans. The effect size differed by ~27% of the effect size in the European group (beta=0.15, $P=6.8x10^{-6}$), being stronger and more significant in the non-European set (beta= 0.19, $P=2.5x10^{-8}$).

DISCUSSION

In summary, we have described the methodology used to genotype, impute, and analyze data for association with phenotypes in the multiethnic Generation R Study, addressing a number of practical issues that arise in implementing imputation-based association for a multiethnic cohort. Our genome-wide genotyped data, ready for analysis after quality control (QC), comprises information for 518,245 markers in 5,732 individuals of different ethnic backgrounds, which is available in the most common genome builds (i.e., 36 and 37). Enrichment by imputation of our genotypes, following a cosmopolitan approach, resulted in an increment of the number of markers of about 5.7 times for the HapMap imputed data and 46 times for the 1KG imputed data (Rsq>0.3).

The Generation R Study withholds a special setting determined by the admixed nature of its population confined within a restricted area. This encompasses analytical challenges as well as opportunities to design genetic studies, which take advantage of such characteristics. A joint analysis including all Generation R participants represents a considerable increase in power of the design, as about half of the study population is of non-Northwestern European background. While, increment in sample size will in principle boost the power of the study, differences in allele frequency or LD relationships between the variants merit further interpretation, as shown in the example of skull BMD with equal number of individuals for both European and non-European sets. Decrease in power due to the use of an admixed population can appear when the association is confined to one of the subpopulations (especially if small) either because of differential tagging or due to the effect of secondary signals [30].

GWAS meta-analyses are expanding to include Non-European populations (i.e., Latinos, African-Americans, etc) with adequate methodology lagging behind due to scarce available software for the processing and analysis of multiethnic data. For example, as the most used software, PLINK [13], relies on the assumption of homogeneous populations it cannot be applied directly to establish family relatedness in multiethnic cohorts. PLINK routine results in an overestimation of relatedness between ancestrally similar individuals. Alternatively, REAP [16] employs a routine that considers the presence of more than one ancestral population and accounts for it in the calculation of IBD probabilities. Nevertheless, in our range of interest for QC purposes -greater than second degree relatives- we found no misclassification of the degree of relationship in the samples. Yet, in situations where high

sensitivity is required (e.g., for the assessment of distant relatedness and/or fine pedigree structure), REAP is recommended in studies with admixed populations.

Choosing the optimal panel to impute the GWAS data of a multiethnic population is critical. For the Generation R Study, we have employed the so-called "cosmopolitan approach", which has become the preferred approach after the release of the 1000 Genomes Project panel [19]. Notably, nowadays all studies are being imputed to the whole 1KG reference panel regardless of the background of the population. Introducing such a combination of reference panels, which achieve very large sample sizes of sequencing reference sets, has been shown to improve imputation accuracy [31-33]. This is mainly beneficial for the imputation of rare variants, which have probably arisen recently and are highly population specific.

New denser reference panels for imputation are becoming available achieving a better characterization of human genetic variation [7,19,34]. The 1KG project data significantly increased the genomic coverage providing more variants suitable to be analyzed in a new phase of the GWAS era [19], with already few reports of novel findings [6,35,36], yet to be embraced at a larger scale. Despite the higher density of markers in the 1KG, only ~63% of the markers achieved good quality as compared to 97% of the HapMap imputed markers. Nonetheless, the low imputation performance observed in 1KG markers is a consequence of the large amount of low-frequency and rare markers in the panel in low LD with the tagging SNPs in the array, which are thus, difficult to impute. When the analysis is limited to common variants (MAF>0.05) present in both datasets (n=2,144,906) the imputation quality was somewhat higher in the 1KG (mean Rsq=0.954, median=0.989, IQR=0.037) than the HapMap panel (mean Rsq=0.947, median=0.986, IQR=0.046), a slight improvement reflecting better imputation arising from a more dense set of markers and a larger reference panel. Special methods for imputation of admixed populations such as MACH-Admix, have also emerged [37], claiming better performance in admixed population and should be part of future studies.

Ethnic background, as assessed by questionnaire did not match the distribution of the samples in the genomic components, mainly because it does not allow for the identification of the third generation participants, i.e., the grand children of those who originally migrated to the country, and thus groups together children that are genetically divergent as shown in Online Resource 7. This comparison together with analysis of different traits, indicate that the genetic structure of studies, such as the Generation R Study, cannot be accounted for by considering the ethnic group definitions based on questionnaire data alone. Another practical advantage when using genomic components to adjust GWAS, is the possibility to include participants even when no information of the parents' country of origin is available. Although in other studies this percentage might be larger, in the population under study, less than 7% of the information on ethnicity from the questionnaire was missing. The ethnic

distribution of the remaining children is in agreement with the ethnic demography of the city of Rotterdam [11] and thus we found no evidence of a systematic non-response.

We chose red hair pigmentation as an example of a highly stratified trait since it is more common in countries in the north of Europe and selected against in Africa due to higher sensitivity of its carriers to UV rays. As shown in Figure 5, if adjustment for population stratification is not used, alleles with different frequencies in Africa and the North of Europe would spuriously show an association with the trait. Instead our association results show that both genomic components and linear mixed models strategies cope well with the substructure of the data and yield similar results. This conclusion can also be derived from the skull BMD GWAS, where even magnitude of effect sizes can be reliably compared. Thus, across all tested scenarios subtle differences emerged not justifying the use of the more computational intensive EMMAX approach. Tests done on other traits such as height, fractional exhaled nitric oxide, site-specific BMD (total body, skull, arms and legs) produced similar results using EMMAX as those published before using genomic components [21,38,39].

It is important to emphasize possible drawbacks of both association strategies. Those of using EMMAX include: 1) for discrete traits obtained betas cannot be translated to odd ratios, given the statistical model applied and 2) its requirement for PLINK files, prevents the use of allelic dosages for analysis. On the other hand, since genomic components are calculated based on the variability of the input data, it is important to generate specific sets of components for particular subsets of the data when working with structured populations. Moreover, while using between two and four genomic components is common practice, the number of genomic components needed to control for population stratification is trait-specific (i.e., dependent on the actual genetic architecture of the trait) [40]. The GIF is an indicator of the degree of inflation of the test statistic due to true signals, cryptic relatedness, assay bias and/or population stratification. Hence, assessment of the GIF is instrumental to determine the needed number of genomic components to be used as covariates in the models. This strategy is not confined to admixed populations and should be assessed even in homogenous populations. In the examples mentioned above, both height and BMD needed up to ten genomic components (data not shown) to reach an acceptable GIF<1.1 [18].

Although the general problem of stratification, differential ethnic allele frequencies, has been successfully addressed in our cohort by the use of genomic components or linear mixed models, the ethnic differences in patterns of correlation between the underlying casual variant and the surrounding SNPs which are under study (LD), can still induce to false-negative findings [41].

As single-center GWAS are usually underpowered, the standard strategy in the field is meta-analysis, the combination of results from multiple independent studies, increasing

sample size and reducing false-positive findings. Frequently, pooling studies from ethnically diverse populations within a single transethnic meta-analysis can be challenging. To cope with this, specialized software such as MANTRA, which allows effect size to vary across different populations, has been developed [42]. The same strategy could also be applied to multiethnic studies such as Generation R, if clear boundaries between different ethnic groups forming part of the study population could be established. However, this is not plausible in our highly admixed population.

The complex structure of the Generation R Study, where admixture of individuals cannot be easily discerned just by assessing the combination of two ancestral populations, constrains the application of admixture mapping, which is an important limitation of our study. Further, in the current setting of the Generation R Study, the small sample size resulting from defining well characterized ethnic groups (of non-European background) is insufficient to allow fine mapping of variants underlying complex traits, typically withholding weak genetic effects. Yet, with new approaches being developed (39), this analytical methodology should be further implemented.

The Generation R Study is unusual in the international arena due to its size, age range, quality of data and longitudinal study design, but particularly due to its multiethnic nature. These characteristics represent the main strengths of the cohort, allowing among others, the generalizability of findings and ethnic comparisons in epidemiological research, although complex routines might be required for genetic association analysis.

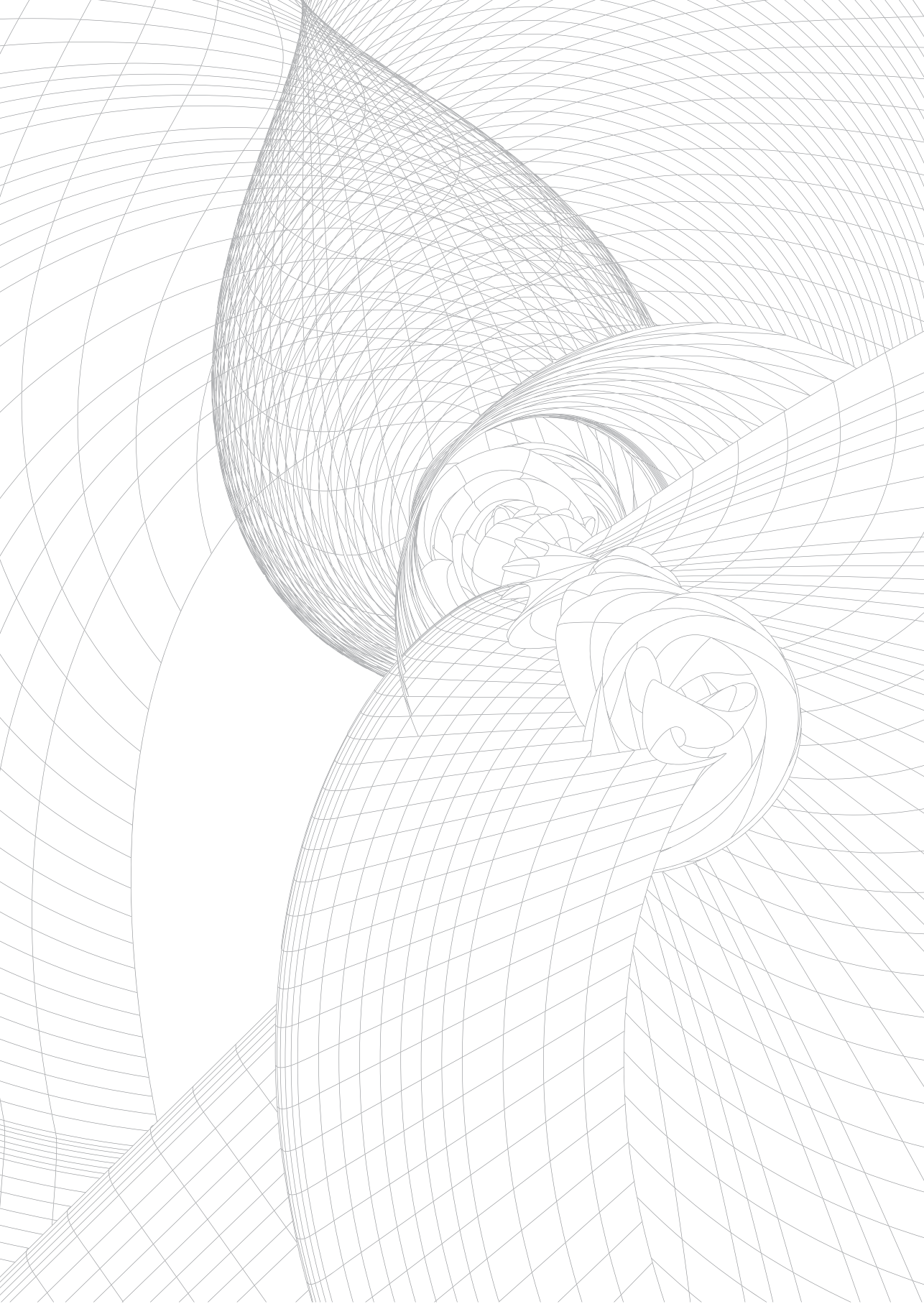
In summary, we have described the methods used for generating the GWAS data of the Generation R Study, as well as general strategies for imputation and analysis within a multiethnic setting. Such strategies have allowed the Generation R Study to take part in several consortia and collaborations, which have successfully identified genetic factors underlying an ample range of complex traits.

Detailed acknowledgements and online resources can be found in the published article online: http://link.springer.com/article/10.1007%2Fs10654-015-9998

REFERENCES

1. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. **Nature**. 2007;447(7145):661-78.
2. Hindorff LA MJ, Morales J,Junkins HA, Hall PN, Klemm AK, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. NIH; 2013.
3. Hindorff LA MJ, Morales J,Junkins HA, Hall PN, Klemm AK, and Manolio TA. A Catalog of Published Genome-Wide Association Studies.: NHGRI.
4. Pulit SL, Voight BF, de Bakker PI. Multiethnic genetic association studies improve power for locus discovery. **PLoS One**. 2010;5(9):e12600.
5. Fu J, Festen EA, Wijmenga C. Multi-ethnic studies in complex traits. **Hum Mol Genet**. 2011;20(R2): R206-13.
6. Wu C, Li D, Jia W, et al. Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. **Nat Genet**. 2013;45(6): 632-8.
7. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. **Eur J Hum Genet**. 2014;22(2):221-7.
8. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. **Nat Rev Genet**. 2011;12(8):523-8.
9. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. **Am J Hum Genet**. 2004;74(5):979-1000.
10. Bradfield JP, Taal HR, Timpson NJ, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. **Nat Genet**. 2012;44(5):526-31.
11. Center for research and statistics RC. 2012.
12. Jaddoe VW, Bakker R, van Duijn CM, et al. The Generation R Study Biobank: a resource for epidemiological studies in children and their parents. **Eur J Epidemiol**. 2007;22(12):917-23.
13. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**. 2007;81(3):559-75.
14. International HapMap C. The International HapMap Project. **Nature**. 2003;426(6968):789-96.
15. International HapMap C. A haplotype map of the human genome. **Nature**. 2005;437(7063):1299-320.
16. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. **Am J Hum Genet**. 2012;91(1):122-38.
17. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. **Hum Mol Genet**. 2008;17(R2):R122-8.
18. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. **Nat Protoc**. 2014;9(5):1192-212.
19. Gao X, Haritunians T, Marjoram P, et al. Genotype Imputation for Latinos Using the HapMap and 1000 Genomes Project Reference Panels. **Front Genet**. 2012;3:117.
20. Kang HM, Sul JH, Service SK, et al. Variance component model to account for sample structure in genome-wide association studies. **Nat Genet**. 2010;42(4):348-54.
21. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **PLoS Genet**. 2014;10(6):e1004423.
22. Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. **Bioinformatics**. 2009;25(20):2750-2.
23. Hosking L, Lumsden S, Lewis K, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. **Eur J Hum Genet**. 2004;12(5):395-9.
24. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. **Nat Protoc**. 2010;5(9):1564-73.

25. Overall AD, Nichols RA. A method for distinguishing consanguinity and population substructure using multilocus genotype data. **Mol Biol Evol**. 2001;18(11):2048-56.
26. Cousminer DL, Berry DJ, Timpson NJ, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. **Hum Mol Genet**. 2013;22(13):2735-47.
27. van der Valk RJ, Duijts L, Kerkhof M, et al. Interaction of a 17q12 variant with both fetal and infant smoke exposure in the development of childhood asthma-like symptoms. **Allergy**. 2012;67(6):767-74.
28. Healy E, Flannagan N, Ray A, et al. Melanocortin-1-receptor gene and sun sensitivity in individuals without red hair. **Lancet**. 2000;355(9209):1072-3.
29. Sulem P, Gudbjartsson DF, Stacey SN, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. **Nat Genet**. 2007;39(12):1443-52.
30. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. **PLoS Biol**. 2013;11(9): e1001661.
31. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. **G3 (Bethesda)**. 2011;1(6):457-70.
32. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. **Genet Epidemiol**. 2010;34(8):816-34.
33. Huang L, Li Y, Singleton AB, et al. Genotype-imputation accuracy across worldwide human populations. **Am J Hum Genet**. 2009;84(2):235-50.
34. What is UK10K? : Wellcome Trust Sanger Institute; 2011.
35. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. **Eur J Hum Genet**. 2012;20(7):801-5.
36. Kou I, Takahashi Y, Johnson TA, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. **Nat Genet**. 2013;45(6):676-9.
37. Liu EY, Li M, Wang W, Li Y. MaCH-admix: genotype imputation for admixed populations. **Genet Epidemiol**. 2013;37(1):25-37.
38. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **PLoS Genet**. 2012;8(7):e1002718.
39. van der Valk RJ, Duijts L, Timpson NJ, et al. Fraction of exhaled nitric oxide values in childhood are associated with 17q11.2-q12 and 17q12-q21 variants. **J Allergy Clin Immunol**. 2014;134(1):46-55.
40. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. **Nat Genet**. 2006;38(8):904-9.
41. Li H, Teo YY, Tan EK. Patterns of Linkage Disequilibrium of LRRK2 across Different Races: Implications for Genetic Association Studies. **PLoS One**. 2013;8(9):e75041.
42. Morris AP. Transethnic meta-analysis of genomewide association studies. **Genet Epidemiol**. 2011; 35(8):809-22.



Chapter 2.2

Improving accuracy of rare variant imputation with a two-step imputation approach

Eskil Kreiner-Møller, Carolina Medina-Gomez, André G Uitterlinden, Fernando Rivadeneira and Karol Estrada

Eur J Hum Genet. 2015 Mar;23(3):395-400. doi: 10.1038/ejhg.2014.91.

ABSTRACT

Genotype imputation has been the pillar of the success of genome-wide association studies (GWAS) for identifying common variants associated with common diseases. However, most GWAS have been run using only 60 HapMap samples as reference for imputation, meaning less frequent and rare variants not being comprehensively scrutinized. Next-generation arrays ensuring sufficient coverage together with new reference panels, as the 1000 Genomes panel, are emerging to facilitate imputation of low frequent single-nucleotide polymorphisms (minor allele frequency (MAF) <5%). In this study, we present a two-step imputation approach improving the quality of the 1000 Genomes imputation by genotyping only a subset of samples to create a local reference population on a dense array with many low-frequency markers. In this approach, the study sample, genotyped with a first generation array, is imputed first to the local reference sample genotyped on a dense array and hereafter to the 1000 Genomes reference panel. We show that mean imputation quality, measured by the r^2 using this approach, increases by 28% for variants with a MAF between 1 and 5% as compared with direct imputation to 1000 Genomes reference. Similarly, the concordance rate between calls of imputed and true genotypes was found to be significantly higher for heterozygotes ($P<1x10^{-15}$) and rare homozygote calls ($P<x10^{-15}$) in this low frequency range. The two-step approach in our setting improves imputation quality compared with traditional direct imputation noteworthy in the low-frequency spectrum and is a cost-effective strategy in large epidemiological studies.

INTRODUCTION

The genome-wide association studies (GWAS) approach has been useful in identifying thousands of single-nucleotide polymorphisms (SNPs) associated with hundreds of complex traits and human diseases [1-3]. This has been made possible by the increase in power achieved by the meta-analysis of different studies where imputation of missing genotypes is essential to harmonize and share data [4, 5, 6]. Low-frequency variants have not been scrutinized by most GWAS based on HapMap content [7]. Using newer reference panels for imputation will allow this (e.g., the 1000 Genomes project based on resequenced data sets) with more variants with lower frequencies for the association analysis with traits [8] hereby increasing resolution and improving power [9, 10] within the so-called next-generation GWAS. Newer arrays have been designed including marker content in the low-frequency spectrum[11] These arrays are expensive (at the time of writing this paper, the approximate price for 5 M array was around 600$ per sample) and it is not clear so far if worth the investment in already GWAS'ed populations.

Here, we propose a two-step imputation approach seeking the optimization of imputations of low-frequency variants when using arrays based on HapMap content. This approach consists of first genotyping a local reference population with an array with dense marker content and a high coverage of markers in the low-frequency spectrum. Second, the whole study population on arrays with HapMap content is imputed to the array content of the local reference set, before imputing to the 1000 Genomes reference panel. This strategy results in improved imputation accuracy and quality in what constitutes a cost-effective strategy for genotyping very large populations.

MATERIALS AND METHODS

Design

This study was nested within the Rotterdam study, a prospective study of 14,926 participants over 45 years of age living in a suburb of Rotterdam. The study has been approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. Over 11000 samples have been genotyped either with Illumina (San Diego, CA, USA) HumanHap microarray beadchips 550 or 610 K [12].

Strategy overview

A subset of individuals from the Rotterdam study was chosen as the study sample for this project. This subset consisted of trios genotyped on two different platforms: the Illumnia

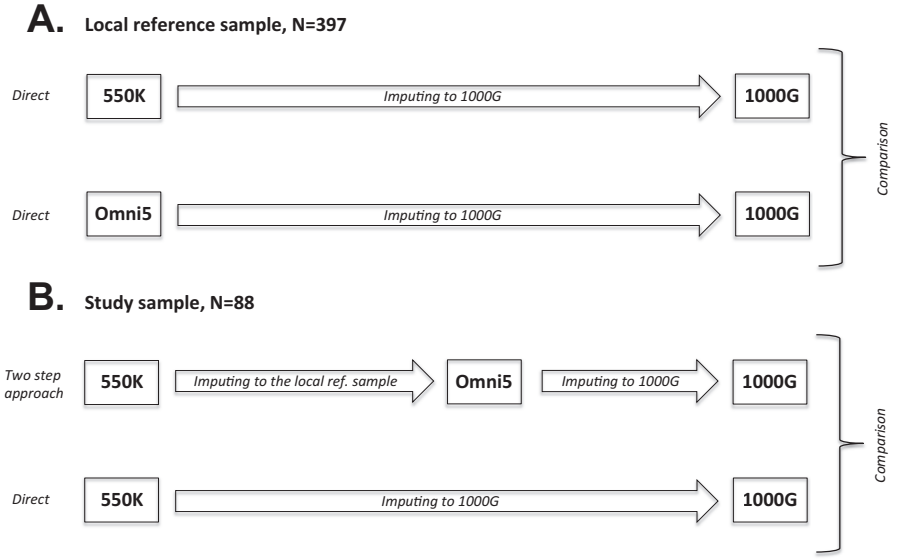


Figure 1. Study Diagram Flow. Panel A illustrates the baseline comparison of the local reference sample genotyped on both the 550K and Omni5 arrays imputed to the 1000 Genomes reference panel. **Panel B** illustrates imputation of the study sample applying the two step approach and comparison to traditional one step imputation approach.

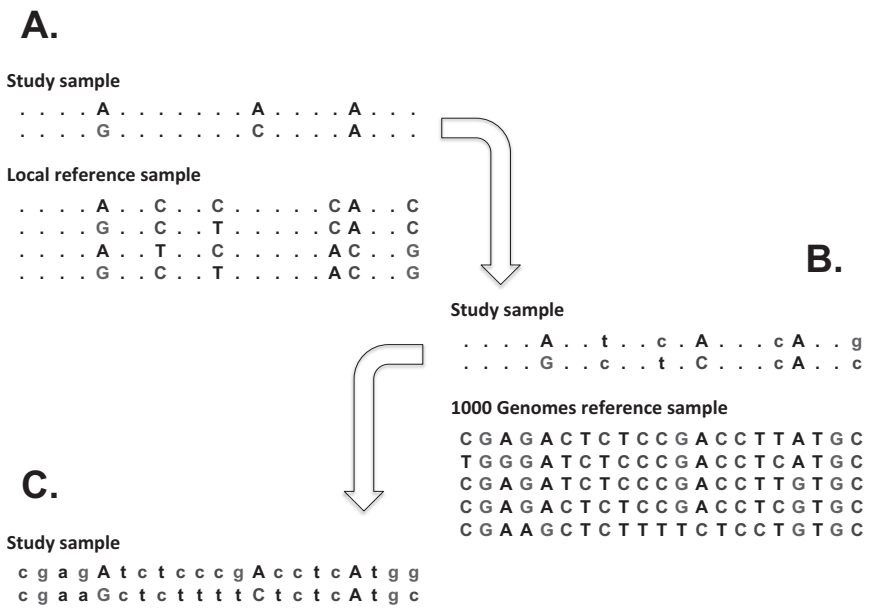


Figure 2: Two step imputation approach. Uppercase characters represent genotyped SNPs while lowercase characters represent imputed SNPs. **Panel A** illustrates the first step: imputation of the study sample using the local reference panel. **Panel B** illustrates the second imputation step: imputing the output best-guess imputed genotypes from step 1 by using the 1000 Genome reference panel. The final output data set is illustrated in **panel C**. Illustration adapted from Li *et al* [6].

550 K and the Omni2.5 (N=88). Another subset of individuals, consisting of women without major disease conditions during follow-up (N=397), was genotyped on the Illumnia Omni5 array and used as the local reference panel as described below. These individuals were also genotyped using the Illumina 550 K array. For baseline comparison of the Omni5 and 550 K arrays, these two data sets were imputed to 1000 Genomes (1000 G) in a traditional direct one-step approach (Online Resource Figure S1).

For the two-step approach, we first used our local reference panel and imputed our study sample using the 550 K data, while the Omni2.5 data for the same individuals were used as gold standard for later accuracy analyses. After this first imputation, the best-guess imputed genotypes were hereafter used as input for imputation using the 1000 G reference panel (Version 3, 2010-11-23, European panel) (Figures 1 and 2). For comparison, we also directly imputed our study sample to 1000 G in a traditional direct one-step approach.

Imputation software

We employed the MACH/Minimac software [13] using a pre-phasing step before the actual imputation as described at the Minimac website (http://genome.sph.umich.edu/wiki/Minimac) adjusted with an extra intermediate imputation step as described above. The local reference sample was phased with MACH and hereafter used as a reference panel.

Samples QC

We used regular QC procedures before imputation steps as described in Online Resource Table S1, including removal of markers out of Hardy–Weinberg Equlibrium ($P<1.0x10^{-6}$) with very low MAF (<0.001) and low SNP call rate (<0.98). We further removed individuals from the study sample related to individuals in the local reference population based on IBD estimation (proportion IBD >0.2). The study sample consisted of 88 samples genotyped on both Illumina 550 K and Omni2.5. The concordance rate between these two arrays was higher than 99.9%. The local reference sample consisted of 397 individuals genotyped on the Omni5 platform.

Imputation quality assessment

The r^2 statistic from Minimac was used as a quality measure with $r^2>0.3$ considered as sufficient imputation quality [4] and high imputation quality defined as $r^2>0.8$. After the first imputation step in the two-step strategy, SNPs were removed using three different filters for comparison: (1) no filter; (2) sufficient quality (r^2 filter on 0.3) and (3) high quality (r^2 filter on 0.8). Accuracy was defined as the proportion of correct imputed genotypes over the total number of imputed genotypes and estimated by comparing the imputed best-guess data to the gold standard data (Omni2.5) using Calcmatch (http://genome.sph.umich.edu/wiki/CalcMatch). This analysis was stratified per genotype class as the allelic concordance

rate for homozygotes major allele (major–major), heterozygotes (major–minor) and for homozygotes minor allele (minor–minor). Mendelian errors were calculated per study sample family in PLINK [14]. Only imputed markers (i.e., those markers not genotyped in the 550 K array) were taken into account while assessing and comparing the imputation quality. To evaluate the influence of number of individuals included in the local reference sample and its impact on imputation quality, we randomly selected two subsets of individuals with 100 and 200 individuals, respectively and compared these with the full local reference using all 397 samples.

The performance of our approach using different imputation parameters was assessed by analyzing chromosomes 2 and 20. Information on samples genotyped across different platforms and allele frequencies is presented in Table 1. After the evaluation of imputation parameters, we imputed all autosomal chromosomes using the two-step approach (r^2 filter on first step 0.3) as well as using the direct approach. We then compared number of SNPs with sufficient imputation quality in the low-frequency spectrum by both approaches. Annotation was done using publicly available data (ftp://share.sph.umich.edu/1000genomes/fullProject/2010.11.23/20101123.annotation.v2.tgz).

When comparing quality between the two-step and the direct imputation strategy in means of r^2 or concordance rates, we applied paired *t*-tests.

Table 1. Description of number of markers and samples per array used in this study

	Local reference sample	Study Sample	
Array	Omni5	550K	Omni2.5
Samples	397	88	
SNPs, CHR2+20	309320	51050	157924
SNPs, MAF >0-1%	61815	413	7738
SNPs, MAF 1-5%	83588	3130	25912
SNPs, MAF 5-10%	33718	5681	20814
SNPs, Overlap*	-	119965	

For the study sample the 550K data was used for imputation. *Overlap: refers to the common SNPs between 550K imputed data and Omni2.5 genotypes used in gold standard analysis

RESULTS

Imputation Quality (as measured by r^2 statistic)

Description of number of markers per array is presented in Table 1. We compared the effect of the array's coverage in imputation quality by direct imputation of the local reference sample (*N*=397) using both the Omni5 and the 550 K array data (Online Resource Figure S1). This plot revealed improvement in imputation quality measured by r^2 using the Omni5

array as compared with the 550 K with a mean r^2=0.91 for the Omni5 and r^2=0.71 for 550 K data in the minor allele frequency range between 1 and 5%.

To assess the effect of the two-step approach, we compared the imputation quality, as measured by the mean r^2 across all imputed SNPs (chromosome 2 and 20), over the whole-allele frequency spectrum (Figure 3). First, our results indicate a gain in mean r^2 while using the two-step approach compared with the direct imputation, especially in the low-frequency ranges. Second, evaluating the results when applying different imputation quality (r^2) cutoffs on the first imputation showed that using no filter and using a filter on r^2>0.3 increased the quality as compared with applying the stricter filter on r^2>0.8.

The comparison of the statistical differences in r^2 values revealed highly significant improvement in imputation quality for the two-step approach using paired t-test (all P<$1x10^{-15}$) (Table 2). In the 1–5% MAF range, the two-step approach using a r^2 filter on 0.3 (mean r^2=0.87) had a 28% higher mean r^2 compared with the direct imputation (mean r^2=0.68). For

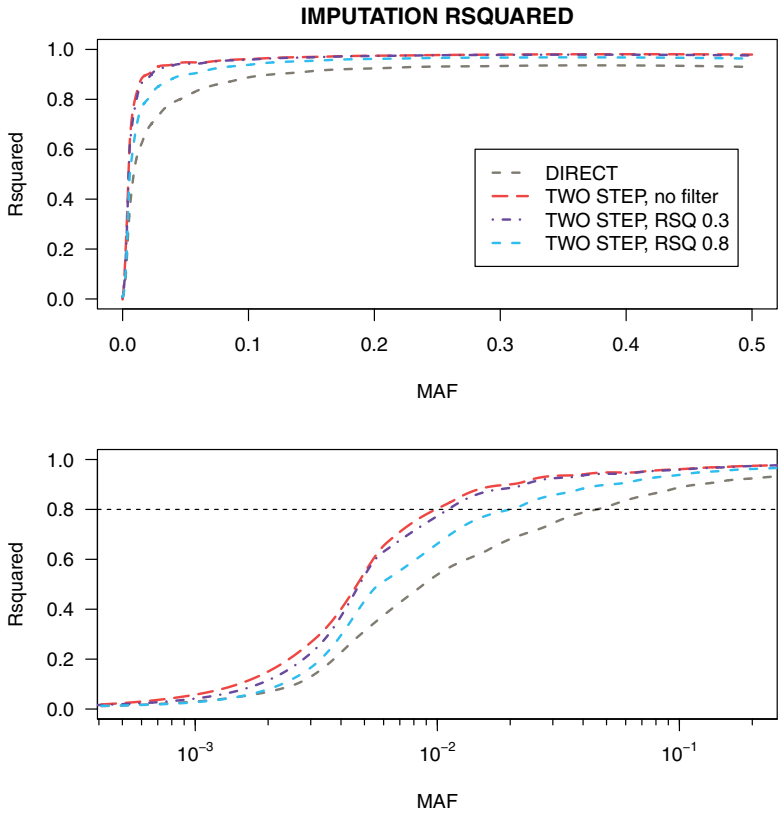


Figure 3. Imputation quality metric (r^2). Imputation r^2 per MAF for the two different imputation strategies, chromosome 2 and 20 for the study sample. The two step strategy further stratified by first imputation step filter. A grey horizontal dashed line at r^2= 0.8 is included to indicate high imputation quality. Lines were created by meaning r^2 per 100 SNPs and applying local polynomial regression fitting (loess function in R-project).

Table 2. Quality comparison traditional one step and proposed two step approaches

MAF	TWO STEP*		DIRECT	
	N	r^2, mean (SD)	N	r^2, mean (SD)
>0-1%**	684334	0.16 (0.28)	839141	0.10 (0.20)
1-5%**	309759	0.87 (0.23)	303608	0.68 (0.31)
5-10%**	144718	0.94 (0.17)	143065	0.85 (0.24)
10-50%**	538142	0.97 (0.11)	539322	0.92 (0.17)

*First imputation filter on r^2 <0.3

** Comparing two step and direct imputation in each MAF bin revealed p-values < 10^{-15}

this analysis, filtering out SNPs genotyped in the local reference panel resulted in a slightly lower r^2 increase between approaches on 21%.

Once we set a fixed threshold of r^2>0.3 (sufficient quality) to remove badly imputed markers in the first imputation step, we imputed all autosomal chromosomes using both the two-step and direct imputation approaches. This analysis revealed 2,528,598 SNPs (intergenic: 78.49%, intronic: 20.47%, exonic: 1.04%) with MAF 1–5% and r^2 above 0.3 using the two-step approach (r2>0.3 on first step) compared with 2,147,168 (intergenic: 78.42%, intronic: 20.58%, exonic: 1.00%) when imputing directly to the 1000 G. This represents an increment of 18% in the number of low-frequency SNPs using the two-step approach (synonymous: 21.10%, non-synonymous: 25.89% and UTR 20.39% more variants). The r^2 distribution of the extra imputed SNPs for the two-step approach in this low-frequency spectrum was highly right-skewed toward an r^2 of 1 depicted in the supplement (Online Resource Figure S2).

Accuracy

Overall accuracy estimates are presented in Online Resource Table S2. The accuracy is also presented per genotype class in Table 3 and in Online Resource Figure S3 shows higher accuracy for the heterozygotes and rare homozygotes when using the two-step approach in low frequencies (MAF<10%). Again, using a filter for the first step on r^2>0.3 performed best. The mean accuracy was statistically significantly higher in the lower MAF (1–5% and 5–10%) bins when using the two-step approach (first-step filter r^2>0.3) compared with the direct imputation. Comparing the accuracy of the heterozygotes-imputed genotypes from the two-step approach using the r^2 filter on 0.3 with the approach without filter also revealed a significant difference (P<1x10^{-15}) and better accuracy of applying a filter on 0.3.

Mendelian consistency

There were no differences between direct and two-step imputation in the number of Mendelian errors per family in the imputed SNPs (Online Resource Table S3).

Table 3. Accuracy comparison traditional one step and proposed two step approaches

MAF	N	TWO STEP* Accuracy, mean (SD)	DIRECT Accuracy, mean (SD)	P-value**
	Common homozygotes (Major-Major)			
>0-1%	25087	0.998 (0.005)	0.999 (0.004)	-
1-5%	25087	0.998 (0.006)	0.998 (0.006)	-
5-10%	18613	0.996 (0.009)	0.996 (0.01)	-
	Heterozygotes (Minor-Major)			
>0-1%	25087	0.830 (0.237)	0.806 (0.244)	3.4e-6
1-5%	25071	0.915 (0.145)	0.886 (0.172)	< 1e-15
5-10%	18613	0.955 (0.083)	0.942 (0.106)	< 1e-15
	Rare homozygotes (Minor-Minor)			
>0-1%	0	-	-	-
1-5%	1248	0.895 (0.241)	0.826 (0.294)	< 1e-15
5-10%	6746	0.925 (0.204)	0.888 (0.245)	< 1e-15

*First imputation filter on r^2>0.3

** Comparing groups, heterozygotes and rare homozygotes in each MAF bin

Evaluating the size of the local reference panel

Accuracy estimates are presented in Figure 4, and Online Resources Figures S4 and S5 for different sizes of the local reference panel. As expected, the larger the local reference panel, the better the quality of the imputation. In the low-frequency spectrum with MAF from 1 to 5%, accuracy measures (two step, r^2>0.3 at first step) dropped in the heterozygotes when lowering the number from all 397 samples (accuracy=0.915) to 200 (accuracy=0.901) and 100 (accuracy=0.879) individuals in the local reference panel. Similar drops were seen in the rare homozygotes in the same MAF bin from all 397 samples (accuracy=0.895) to 200 (accuracy=0.868) and 100 (accuracy=0.838) individuals used in the local reference panel.

DISCUSSION

Principal findings

In this study, we have compared the traditional direct one-step imputation procedure with a proposed two-step imputation approach to improve the quality of the imputed data obtained after the use of dense reference panels, as the ones emerging nowadays. Our most important finding was that the two-step approach increases both the quality and the accuracy of imputation, even increasing the number of high-quality markers in 18% for the lower frequency spectrum (MAF 1–5%) as compared with the traditional direct imputation

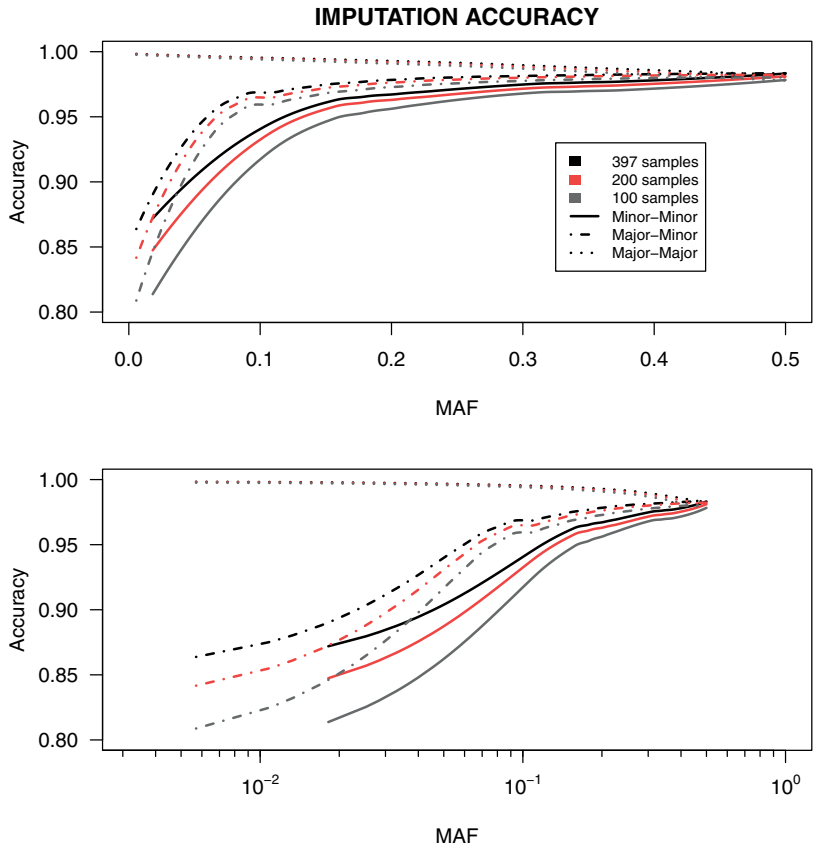


Figure 4. Imputation Accuracy for different local reference sizes. Imputation accuracy stratified per genotype for the different sizes of the local reference panel. This is presented using a first step r^2 filter on 0.3.

approach. Accuracy improvement was primarily seen for the low-frequency variants and only in the heterozygotes and rare homozygotes genotype classes.

Limitations and strengths

We evaluated the two-step imputation approach by means of different measures of quality, specifically r^2 and imputation accuracy. Comparison to a gold standard data (in our case: 88 samples genotyped with two different arrays) must be seen as one of the primary strengths of this study, enabling the evaluation of the approaches in low-frequency markers. Nonetheless, one limitation is the relative low number of samples in our study sample.

To evaluate the gain of power and improvement in means of novel findings in the GWAS setting, a future effort should focus on applying the two-step approach to entire data sets, as earlier applied when analyzing the difference between HapMap and 1000 G imputa-

tion[9]. This comparison would surely require large sample sizes and the combination of several studies before an assessment of its value at meta-analysis level can be done.

Further estimation of minimum marker content of the input for the study sample would also be necessary to evaluate and analyze by testing if this two-step approach would surpass the quality of direct imputation in studies that are already genotyped in arrays with lower content than the 550 K array.

Meaning of the study

We found that the proposed two-step imputation approach excels the imputation quality (r^2) and accuracy (concordance) as compared with the direct imputation to 1000 G. One of the main reasons for this could be that we used a dense array with many low-frequency markers at the intermediate imputation step. As these low-frequency markers can create population-specific haplotypes [8], we would expect that the markers imputed in our study sample using the local reference panel provide more precise haplotypes of both common and rare variants over which imputation to a more comprehensive panel such as the 1000 G can be improved. The main reasoning to this improvement being that a local reference panel must be closer to the study sample in means of ancestry and that the overall number of haplotypes for some markers are doubled (794 haplotypes in the local reference panel in step 1 + 758 from the 1000 G panel in step 2). However, we did not apply the two-step approach imputing other European populations and using this Dutch local reference panel. The reasoning on closeness in ancestry between the study sample and local reference panel is therefore speculations.

It is possible that the increasing quality of future panels of the 1000 G (or other sequencing projects used as reference) may decrease the boost provided by our proposed two-step approach and thus be as efficient or even surpass the accuracy reached by the two-step approach suggested in this manuscript. Nevertheless, the rationale behind our approach is that by first imputing from a local reference panel can result in more accurate imputation of variants in the lower minor allele frequency spectrum, independent of the quality of the reference panel.

The size of the local reference panel influenced the imputation quality of the two-step approach. Nonetheless, for heterozygous genotypes of variants in the MAF 1–5% range, the increase in quality seems to level-off with increasing sample size, as observed from doubling the local reference panel from 100 to 200 samples. This increment showed a bigger leap in imputation quality, than the one after doubling the number from 200 to 397 samples (Online Resource Figure S5). Probably, further increment in the local reference panel size (beyond the 397 samples) could still add to the imputation accuracy but in a less pronounced manner. Moreover, the current imputation quality metrics are within reason-

able values expected across studies, raising the question if adding more samples will be a cost-efficient approach.

One other study by Sampson *et al.* [15] has assessed a two-platform approach, genotyping a subset of samples on a denser array (Omni2.5). These samples were combined with the 1000 G reference set before imputation. Increasing the number of samples to combine with the 1000 G improved r^2 [15]. However, by using standard imputation tools, the combination of reference panels is only implemented in IMPUTE2,5 software not used in the current study. We were therefore not able to compare the two-step approach to a direct one-step approach with the combined use of the 1000 G reference panel and our local reference panel. Nonetheless, in a recent study we showed that combining 1000 G with a Dutch sequenced panel (GoNL) only increased imputation r^2 by 1.4% compared with the GoNL alone, for low-frequency markers imputing a Dutch GWA data set[16] . Thus, we would not expect that combining our local reference panel with the 1000 G panel would surpass the increase showed in our previous study.

Our study highlights a possible cost-effective strategy for large studies undertaking genome-wide genotyping or a possible upgrade of existing genotyped data sets. We expect that in large studies (with thousands of individuals), a substantial reduction in costs would be achieved by genotyping only a subset of the samples on these relatively expensive arrays with many low-frequency markers (e.g., Omni5) and subsequently use these samples as a local reference set in a two-step imputation approach. This extra load will represent an increase in imputation quality and hence, the uncertainty of imputation calls will decrease, yielding more precise dosages what will be reflected in the power to detect associations in future GWAS on complex diseases. Nonetheless, discovering low-frequency variants associated with complex traits in GWAS does not depend only on imputation quality, but also on sufficiently powered settings (large sample sizes). Thus, the proposed two-step approach needs to be applied by all participating studies of a GWAS meta-analysis in order to observe an increase in the power needed to detect rare variants.

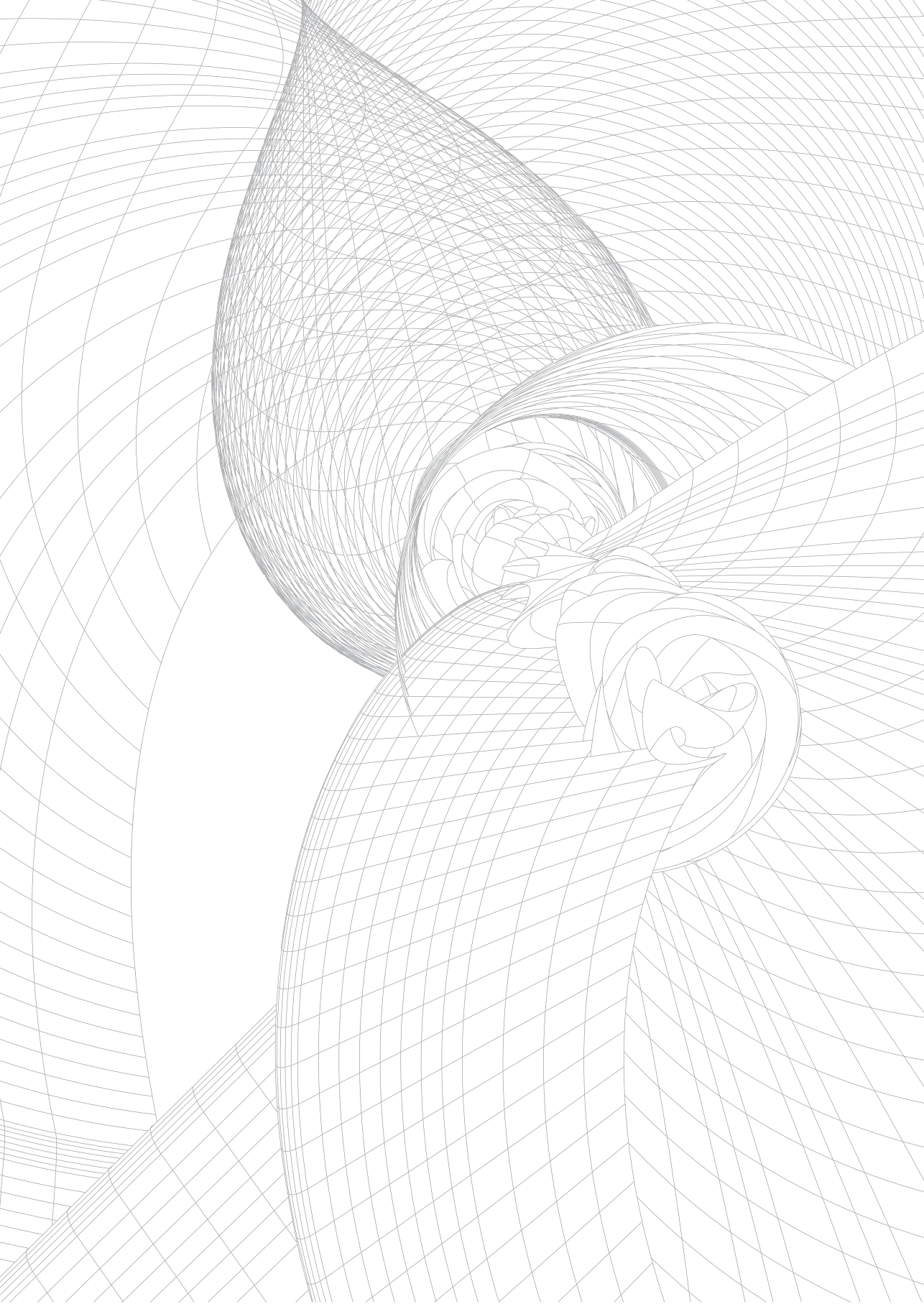
CONCLUSION

Imputation is a useful approach to improve both coverage and power in genetic association studies. The two-step approach in our setting increased imputation quality compared with direct imputation especially in the low-frequency spectrum. Further, this imputation methodology is a cost-effective strategy for improving imputation quality in large samples.

Detailed acknowledgements and online resources can be found in the published article online: http://www.nature.com/ejhg/journal/v23/n3/full/ejhg201491a.html

REFERENCES

1. Hindorff, L., Junkins, H., Hall, P., Metha, J., and Manolio, T. A Catalog of Published Genome-Wide Association Studies.
2. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five Years of GWAS Discovery. **Am. J. Hum. Genet**. 90, 7–24.
3. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. **Nature** 467, 832–838.
4. De Bakker, P.I.W., Ferreira, M.A.R., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. **Hum. Mol. Genet**. 17, R122–128.
5. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. **Nat. Rev. Genet**. 11, 499–511.
6. Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype Imputation. **Annu. Rev. Genom. Hum. G.** 10, 387–406.
7. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. **Nature** 461, 747–753.
8. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. **Nature** 467, 1061–1073.
9. Huang, J., Ellinghaus, D., Franke, A., Howie, B., and Li, Y. (2012). 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. **Eur. J. Hum. Genet**. 20, 801–805.
10. Sung, Y.J., Wang, L., Rankinen, T., Bouchard, C., and Rao, D.C. (2012). Performance of genotype imputations using data from the 1000 Genomes Project**. Hum. Hered**. 73, 18–25.
11. Baker, M. (2010). Genomics: The search for association. **Nature** 467, 1135–1138.
12. Hofman, A., van Duijn, C.M., Franco, O.H., Ikram, M.A., Janssen, H.L.A., Klaver, C.C.W., Kuipers, E.J., Nijsten, T.E.C., Stricker, B.H.C., Tiemeier, H., et al. (2011). The Rotterdam Study: 2012 objectives and design update. **Eur. J. Epidemiol**. 26, 657–686.
13. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. **Nat. Genet**. 44, 955–959.
14. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am. J. Hum. Genet**. 81, 559–575.
15. Sampson, J.N., Jacobs, K., Wang, Z., Yeager, M., Chanock, S., and Chatterjee, N. (2012). A two-platform design for next generation genome-wide association studies. **Genet. Epidemiol**. 36, 400–408.
16. Deelen, P., Menelaou, A., van Leeuwen, E.M., Kanterakis, A., van Dijk, F., Medina-Gomez, C., Francioli, L.C., Hottenga, J.J., Karssen, L.C., Estrada, K., Kreiner-Møller, E., Rivadeneira, F., van Setten, J., Gutierrez-Achury, J., Westra, H.J., Franke, L., van Enckevort, D., Dijkstra, M., Byelas, H., van Duijn, C.M., Genome of Netherlands Consortium, de Bakker, P.I., Wijmenga, C., Swertz, M.A. (2014). **Eur J Hum Genet**. 22, 1321-6.



Chapter 2.3

Genome-wide association study in an admixed case series reveals *IL12A* as a new candidate in Behçet disease.

Jasper H. Kappen, Carolina Medina-Gomez, Martin van Hagen, Lissete Stolk, Karol Estrada K, Fernando Rivadeneira, André G. Uitterlinden, Miles R. Stanford, Eldad Ben-Chetrit, Graham R. Wallace, Merih Soylu, Jan A. M. van Laar

PLoS One. 2015; 10(3): e0119085. doi:10.1371/journal.pone.0119085

ABSTRACT

Introduction: The etiology of Behçet's disease (BD) is unknown, but widely considered an excessive T-cell mediated inflammatory response in a genetically susceptible host. Recent genome-wide association studies (GWAS) have shown limited number of novel loci-associations. The rarity and unequal distribution of the disease prevalence amongst different ethnic backgrounds have hampered the use of GWAS in cohorts of mixed ethnicity and sufficient sample size. However, novel statistical approaches have now enabled GWAS in admixed cohorts. **Methods**: We ran a GWAS on 336 BD cases and 5,843 controls. The cases consisted of Western Europeans, Middle Eastern and Turkish individuals. Participants from the Generation R study, a multiethnic birth cohort in Rotterdam, The Netherlands were used as controls. All samples were genotyped and data was combined. Linear regression models were corrected for population stratification using Genomic Principal Components and Linear Mixed Modelling. Meta-analysis was performed on selected results previously published. **Results**: We identified SNPs associated at genome-wide significant level mapping to the 6p21.33 (HLA) region. In addition to this known signal two potential novel associations on chromosomes 6 and 18 were identified, yet with low minor allele frequencies. Extended meta-analysis revealed a GWS association with the *IL12A* variant rs17810546 on chromosome 3. **Discussion**: We demonstrate that new statistical techniques enable GWAS analyses in a limited sized cohort of mixed ethnicity. After implementation, we confirmed the central role of the HLA region in the disease and identified new regions of interest. Moreover, we validated the association of a variant in the *IL2A* gene by meta-analysis with previous work. These findings enhance our knowledge of genetic associations and BD, and provide further justification for pursuing collective initiatives in genetic studies given the low prevalence of this and other rare diseases.

INTRODUCTION

Behçet's disease (BD) is a systemic auto-inflammatory occlusive vasculitis presenting with oral and genital aphthous ulcers, skin lesions, ocular inflammation and a pathognomonic pathergy test [1-3]. Although the etiology of the disease is largely unknown, it is considered to be an excessive inflammatory response possibly triggered by an infectious antigen in a genetically susceptible host. Prevalence varies from 110–420 per 100,000 in the Middle East (Turkey) to about 2 per 100,000 in Western countries [3, 4]. In patients of Turkish origin there is a positive family history in 12% of the cases with a sibling risk ratio between 11–52 [3, 4]. The heritability of BD has been estimated to range between 20 to 60%, with the strongest genetic association embracing variants in HLA-B51, explaining about 20% of the disease heritability [3,4]. The combination of epidemiological and genetic data suggest causal involvement of both genetic and environmental factors [5].

Over the years, numerous single nucleotide polymorphisms (SNPs) have been found associated with BD [6-9]. In 2010 the first genome-wide association study (GWAS) in BD cohorts of Turkish and Japanese origin demonstrated association of various variants in the known HLA-B51 domain, and two new association signals, mapping to the interleukin-10 (*IL10*), and the IL-23 receptor–IL-12 receptor Beta2 (*IL23R–IL12RB2*) locus [10,11]. These associations later have been confirmed in an Iranian cohort [12]. Yet another study in Algerian individuals only replicated the associations from the *IL10* variants [13]. GWAS in Chinese and Korean individuals reported associations of SNPs mapping to two other loci, with regions containing the signal transducer and activator of transcription 4 (*STAT4*) and GTPase of immune associated protein GIMAP genes, respectively. Polymorphisms in *IL10* and *IL23R–IL12RB2* were not found to be significantly associated with BD in neither of those studies [14,15]. More recently, associations with *CCR1*, *STAT4* and *KLRC4* were identified by means of imputation and GWAS meta-analysis[14]. This study also identified the *IL12A* region as suggestively associated with BD, but genome wide significance (GWS) was not reached. Discrepant association results, as those observed for variants mapping to the GIMAP locus, might be explained by the different ethnic origin of the cohorts studied[15,16]. Therefore, studying cohorts of diverse ethnic background may be useful to understand the origin of these differences. Moreover, the inclusion of multiple ethnicities results in larger datasets (representing higher power) crucial for these analyses given the typically small effects of common genetic variants discovered by the GWAS approach[17]. Nonetheless, GWAS results can be confounded by ancestry differences between cases and controls, potentially leading to spurious associations. This challenge in multi-ethnic studies has so far been approached by methods that take into account confounding by genetic ancestry, such as correction for Genomic Principal Components (GPC) [18]. More recently, linear mixed models (LMM) have been introduced as a reliable method to correct not only for ancestry differences, but also for family structure and/or cryptic relationships [19].

We have therefore set up a multi-center GWAS including BD patients of both Middle Eastern and Western background. We demonstrate that by means of novel statistic approaches it is feasible to run a GWAS in a small multiethnic cohort and use these results for meta-analysis.

MATERIALS AND METHODS

Study populations

A total of 369 unrelated BD patients from 18 different geographic origins were included in our study (Table 1). The age range was 18 to 73 years (mean 47 years). All patients fulfilled the International Study Group (ISG) criteria of BD diagnoses [20]. No exclusion criteria were applied.

Table 1. Cases before and after QC. Collection at [1] Erasmus MC at Rotterdam, The Netherlands, [2] University Hospital of Cukurova, Turkey, [3] St John's Ophthalmic Hospital, Jerusalem, Israel, [4] St Thomas' Hospital, London, UK, [5] University Hospital, Damascus, Syria, Ethnicity of cases after QC.

Ethnicity	Number of patients	Collection	Number of patients after QC
Afghanistan	1	[1]	1
Iranian	4	[1]	4
Lebanese	1	[1]	1
Cape Verde	2	[1]	2
Curacao	1	[1]	1
Dominican Republic	1	[1]	1
Dutch caucasian	24	[1]	20
Greece	1	[1]	1
Israel	1	[1]	1
Jordan	1	[1]	1
Morocco	16	[1]	12
Surinam	2	[1]	1
Thailand	1	[1]	1
China	1	[1]	1
Turkey	35	[1]	32
Turkey	91	[2]	87
Arab Jerusalem ancestry	110	[3]	98
UK caucasian	38	[4]	33
Syrian	38	[5]	38
Total	**369**		**336**

As controls, 87 Syrian healthy volunteers (22-73 years, mean 51 years) and 5,756 participants from the Generation R study, a multi-ethnic birth cohort in Rotterdam (4-9 years, mean 6 years), The Netherlands, were used [21]. Replication was pursued in a cohort of 82 BD patients who met the ISG criteria, and 98 ethnically matched controls in Western European, collected from Birmingham and Midland Eye Centre, Birmingham, UK. The ethics committees of the involved centers (Erasmus MC (METC), Ethical Committe at Çukurova University and Sandwell and West Birmingham Hospitals Trust Ethics Committee) approved the study, and written informed consent was obtained from all participants.

Genome wide genotyping

Genotyping of cases and controls (including the Generation R Study) was performed using the Illumina HumanHap 610K and/or 660 K arrays, following manufacturers protocols. Quality Control (QC) was performed individually for each set following a standard protocol to exclude samples and SNPs with low quality genotyping. Markers with minor allele frequency (MAF) ≤ 1%, missing genotypes ≥5% or which failed an exact test of Hardy-Weinberg equilibrium proportions in the controls ($P<1x10^{-7}$) were excluded from the datasets. Samples with gender discrepancy, excess of heterozygosity, duplicates or samples with relatedness or other inconsistencies were also removed. In a subsequent stage, we extracted SNPs common to both platforms which passed individual QC and then applied the QC protocol for the combined dataset.

Phenotype-Genotype Analyses

Heritability

To characterize the extent to which common genetic variants play a role in BD susceptibility, we applied a recently proposed approach for estimating heritability based on genome-wide sharing between distantly related individuals, as implemented in the GCTA software in combination with the determination of ancestry-aware kinship coefficients by REAP software [22, 23]. No pairs exceeded the standard cutoff coefficient of 0.025 for genetic relatedness, confirming that no two individuals in the analysis were closer than third degree cousins. Relatedness coefficients calculated from REAP were used in the relatedness matrix before implementing GCTA. A disease prevalence of 0.42% was used in the modelling. [3, 4]

Genome-wide association analysis

Association between BD susceptibility and 553,224 genome-wide SNPs was carried out using a regression framework adjusting for population stratification. GWS threshold for the association was established at $p<5x10^{-8}$. Two different strategies were followed to adjust for population sub-structure in the data.

Genomic Principal Components

In the first approach, 20 GPC were obtained for all pairs of individuals in the combined dataset based on the similarities of individual genotypic profiles for 37,060 independent autosomal SNPs contained in the HapMap Phase II release 22 build 36 [24] using Multi-Dimensional Scaling (MDS) as implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/). Afterwards, we performed logistic regression adjusting for those 20 GPC.

Linear Mixed models

In the second approach we used LMM in order to account for possible population structure in the association analysis. This methodology estimates the level of relatedness even between independent individuals using the genotyped markers. By modeling the dissimilarity between genotypes of the subjects, it is able to correct the association results for stratification. This approach is implemented in the publicly available EMMA eXpedited (EMMAX) software[25]. Since the effect sizes reported by EMMAX are based on Mantel–Haenszel statistics, the odd ratios for associations of the significant SNPs were estimated from the regular logistic regression models performed in PLINK.

Meta-analysis

Meta-analysis of our own study results with published results for the BD suggestive associated SNP rs17810546 [14] was performed using an inverse variance fixed-effects approach in METAL[26].

Replication analysis

Four SNPs: rs8187722, rs439033, rs11969661 and rs17087141 were genotyped using TaqMan assays designed by Applied Biosystems, Warrington, UK in 82 BD patients and 98 control samples. In short, PCR was performed in 384-well plates with a 10 µl total reaction volume containing 20 pg/ml DNA (samples) or water (controls) and 2x LC-480 probe master (Roche, West Sussex UK) followed by endpoint genotyping analysis using the LC-480 system (Roche, West Sussex UK). Genotypes were determined as either homozygous (e.g., AA or GG) or heterozygous (e.g., AG) according to the presence or absence of fluorescence for each genotype.

RESULTS

An overview of the ethnicity of the included BD cases is presented in Table 1; all controls were collected in St John's Ophthalmic Hospital, Jerusalem and had Arab ancestry. In order to gain power and match all remaining ethnicities, this dataset was combined with the genotyped samples from the multiethnic Generation R Study.

Genome wide Genotyping

Genotyping was performed in 456 samples (369 cases and 87 controls) collected from the different participating hospitals (Erasmus MC at Rotterdam, The Netherlands, University Hospital of Cukurova, Turkey, St John's Ophthalmic Hospital, Jerusalem, Israel and by St Thomas' Hospital, London, UK). After quality control, 553,224 genotyped SNPs in 423 of these samples remained. We excluded 33 patients (including 1 control) because of low Illumina call rate (< 97.5%) and one more individual was excluded for gender mismatch. After QC of this combined dataset, 336 cases and 5,843 controls were available for analysis.

Genomic Principal Components

The representation of the first GPC for the combined dataset projected together with the three reference panels of the International HapMap Project Phase II known as YRI, CEU and CHB/JPT representing populations of African, European and Asian background respectively [27] is shown in Figure 1. As shown by the overlapping distribution across the four first GPC, BD cases (represented in black) are well matched by genetic ancestry to controls (represented in yellow Syrian volunteers and grey for Generation R participants).

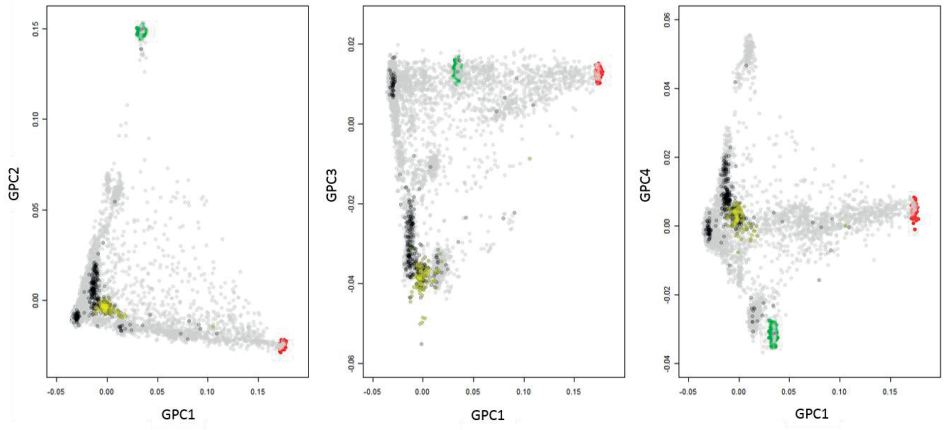


Figure 1. Genetic substructure of the Combined GWAS dataset. Two-dimensional scatterplots from multi-dimensional scaling analyses of the Generation R Study and Behçet collected data together with the three initial Panels form the HapMap Project. Each dot represents an individual in the dataset. Color codes: Grey=Generation R, Black= Behcet Cases, Yellow = Jordan controls. Blue= CEU, Red=YRB, Green= CBH/JPT.

Phenotype Genotype Analyses

Heritability

The heritability estimate, based on the 505,454 genotyped SNPs and 4,855 unrelated individuals (296 cases) was 32.0% (95%CI 21.0-44.0%;P= $5.5x10^{-17}$). Thus a considerable percent of the BD risk is explained by the additive effect of the common SNPs analyzed.

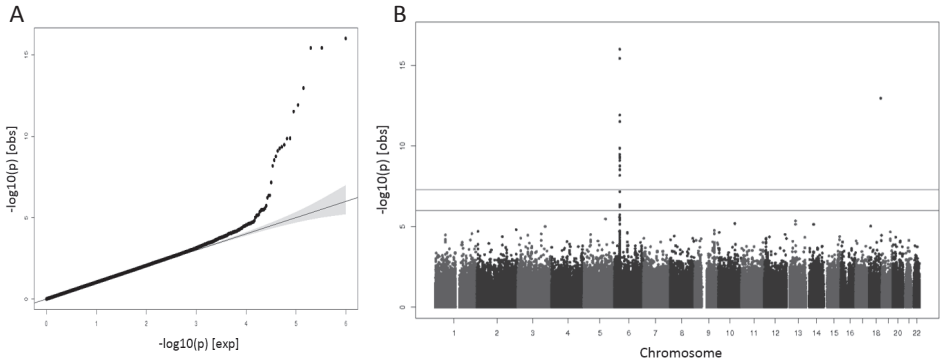


Figure 2. Behçet GWAS results using Genomic Principal Components (GPC) adjustment. Each dot represents an SNP in the dataset. A. QQ-plot. Associated SNPs deviating from the null hypothesis of no association (identity line). **B.** Manhattan plot. SNPs showing association with the disease map to chromosome 6 and a singleton in chromosome 18.

Table 2. GWS SNPs associated with Behçet's disease susceptibility using GPC and EMMAX approaches. Results for the association analysis when correcting for 20 PCs to adjust for stratification. In italic font the SNPs which show association only by one of the two approaches

CHR	SNP	BP	A1	A2	MAF	N	20PCs		EMMAX* P
							OR	P	
6	rs9380215	31157634	A	G	0.09	6178	2.132	1.32E-10	3.15E-15
6	rs4947296	31166157	C	T	0.09	6178	2.131	1.38E-10	4.98E-15
6	rs9263509	31174398	T	C	0.16	6179	1.82	6.62E-09	2.69E-10
6	rs2233984	31187243	A	G	0.09	6179	2.064	3.35E-10	1.08E-15
6	rs4495304	31188697	C	T	0.09	6028	2.069	7.92E-10	3.84E-14
6	rs4959053	31207556	A	G	0.09	6175	2.348	1.20E-12	3.90E-17
6	*rs2523589*	*31435313*	*C*	*A*	*0.43*	*6177*	*0.5466*	*2.92E-09*	5.83E-08
6	rs2844575	31442924	G	A	0.48	6173	1.81	5.62E-10	5.92E-08
6	rs9266409	31444547	C	T	0.25	6178	2.183	3.66E-16	2.62E-18
6	rs2253907	31444849	A	G	0.48	6175	1.797	4.41E-10	1.46E-09
6	rs7770216	31448590	T	G	0.26	6072	2.198	9.75E-17	2.25E-18
6	rs6933050	31451611	C	T	0.25	6178	2.183	3.69E-16	2.59E-18
6	rs1131896	31487094	A	G	0.31	6141	1.647	6.86E-08	2.66E-09
6	rs2256028	31487177	T	G	0.21	6177	1.784	1.72E-09	7.49E-12
6	rs2848713	31492458	A	G	0.1	6173	2.118	3.03E-12	7.08E-22
6	*rs11969661*	*160626409*	*T*	*C*	*0.02*	*6179*	*2.696*	*1.30E-03*	*5.15E-09*
6	*rs439033*	*160739406*	*C*	*T*	*0.02*	*6163*	*2.016*	*2.30E-03*	*1.14E-08*
6	*rs8187722*	*160784748*	*G*	*A*	*0.01*	*6179*	*1.696*	*1.00E-04*	*2.87E-15*
18	rs17087141	69036707	C	T	0.03	6167	4.925	1.08E-13	2.81E-18

* Odd ratios cannot be calculated by EMMAX approach

Genome wide association analyses

Association analyses were performed for 505,454 SNPs. As expected in such structured population, results from analysis without any type of stratification adjustment, showed association through the whole genome and a Genomic Inflation Factor (λ) of 5.25 (Online Resource Figure S1). After GPC correction, the association analysis showed minimal inflation of the test statistics with a λ of 1.035 (Figure 2). The top associated SNPs mapped to the HLA region on chromosome 6 with rs7770216 being the most significant (Table 2). Yet, another GWS signal in chromosome 18 driven by rs17087141 mapped to a region containing the uncharacterized miscRNA LOC400655.

Additional association analyses using different strategies including reduction of number of matching controls (to different case/control ratios) and incorporating different number of genomic principal components in the models yielded similar results for effect coefficients and genomic inflation factors (data not shown).

Results from the GWAS analysis using EMMAX yielded three independent GWS signals (Table 2). In addition to those mapping to the two previously described BD loci from chromosomes 6 and 18, an additional signal was found on chromosome 6q25.3 6 in the solute carrier family 22 member 3 (*SLC22A3*) gene region. The three most significant markers underlying this GWAS signal included rs11969661, rs439033 and rs8187722 (Table 2). Using EMMAX resulted in adequate adjustment for population stratification as corroborated by λ=1.0 (Table 2) and the absence of early deviation from the identity line in the QQplot (Figure 3).

The most significant SNP identified by the EMMAX approach maps also to the HLA region in chromosome 6. The regional association plot of this HLA-B region is shown in Online Resource Figure S2. This HLA-B region has been associated with BD susceptibility in previous GWAS analyses [10,22]. Figures 4 and 5 show the association plots for the other

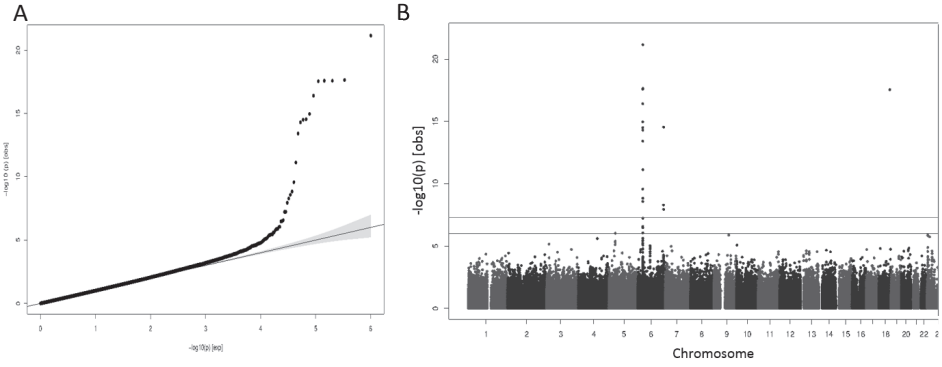


Figure 3. Behcet GWAS results using Linear Mixed Models Genomic approach. Each dot represents an SNP in the dataset. **a.** QQ-plot. Associated SNPs deviating from the null hypothesis of no association (identity line). **b.** Manhattan plot. SNPs showing association with the disease map to two different signals in chromosome 6 and a singleton in chromosome 18.

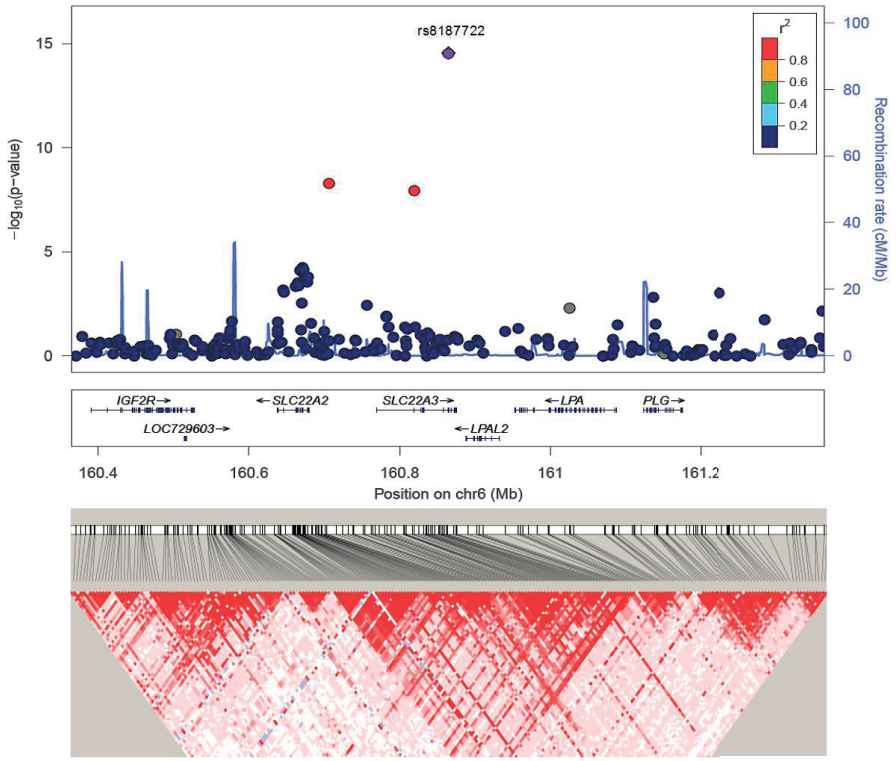


Figure 4. SNP association plot for Behçet's susceptibility-associated region of chromosome 6q25.3. Dots represent GWAS P-values (EMMAX approach) and positions of SNPs found within the 6q25.3 locus. The top SNP, i.e., rs8187722, is denoted by a diamond. Different colors indicate varying degrees of pair-wise linkage disequilibrium (1000 Genomes Nov 2010 CEU) between the top SNP and all other genotyped SNPs. Genetic coordinates are per 1000 Genomes Nov 2010-CEU. **Bottom:** LD heat map based on D' values from the combined population under study including all SNPs in the 500Kb region.

two GWS signals in 6q25.3 (*SCLA22A3* locus) and 18q22.3, which have not been previously reported as associated with BD.

Meta-analysis

By scanning published BD GWAS, we identified a variant suggestive of association ($P=6.01x10^{-7}$) [14], showing strong evidence for association with BD in our study, although not at GWS level. After meta-analysis of our results with those published by Kirino *et al.* [14] in two Turkish cohorts the leading SNP (rs17810546) in *IL12A* reached GWS. The G-allele conferred a 1.7 increased risk for BD (OR= 1.66; 95%CI [1.42,1.93]; $P= 1.12x10^{-10}$ (Table 3)).

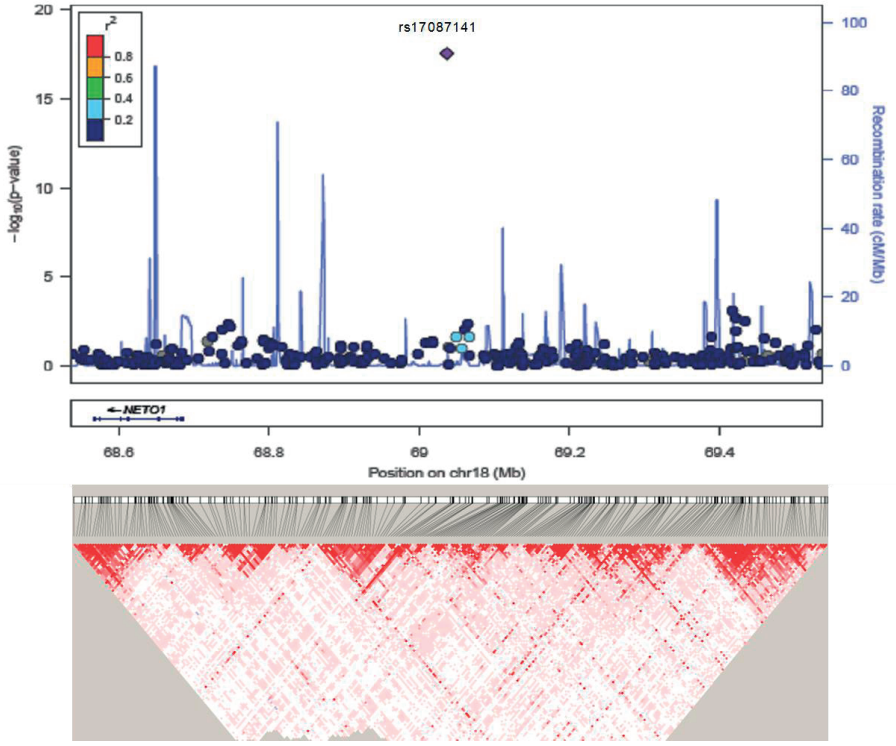


Figure 5. SNP association plot for Behçet's susceptibility-associated region of Chromosome 18q22.3. Dots represent GWAS P-values (EMMAX approach) and positions of SNPs found within the 18q22.3 locus. The top SNP, i.e., rs17087141, is denoted by a diamond. Different colors indicate varying degrees of pair-wise linkage disequilibrium (1000 Genomes CEU) between the top SNP and all other genotyped SNPs. Genetic coordinates are per 1000 Genomes Nov 2010-CEU. **Bottom**: LD heat map based on D′ values from the combined population under study including all SNPs in the 500Kb region.

Table 3. Meta-analysis of leading SNP in IL12A. Association results for rs17810456 (G-allele) in two Turkish cohorts [14] and the current study.

Study	N	N cases	OR	CI 95%	P value
Turkish replication*	2487	1209	1.64	[1.31 - 2.04]	1.20E-05
Turkish discovery*	1447	821	1.41	[1.06 - 1.89]	2.00E-02
Current study	6179	336	2.06	[1.49 - 2.84]	9.31E-06
Combined	10113	2366	1.66	[1.42 - 1.93]	1.12E-10

* From Table 1 Kirino et al. 2013

Replication Analysis

No significant evidence for replication of the leading SNPs from the GWAS signals of chromosome 6 and 18 was found, while the MAF of the genotyped markers were very similar to those reported in healthy individuals.

DISCUSSION

Enabled by innovative novel methodology we could run a GWAS in a rare condition like BD within a unique case collection of multiethnic background. We identified variants associated with BD mapping to the well-established *MICA-HLA-B* locus and to two regions on chromosome 6 in the *SLC22A* gene region and on chromosome 18 in an uncharacterized region. Moreover, meta-analysis with previous published results enabled the identification of a GWS association with variants in the *IL12A* region. All together, these common variants across these four loci explained up to 32% of the variance in BD risk.

Methods for the calculation of heritability estimates from SNP microarray data of population-base studies have recently emerged [22,23]. We could determine that the narrow sense heritability of BD explained by common variants in our study was 32%, estimate in line with those obtained by previous reports based on family data [3,4,28], albeit smaller. This could be explained by the overestimation of heritability estimates in family-based design, as a result of biases due to epistatic interactions or shared environment. In addition, GCTA estimates should be interpreted as the lower bound of the true additive genomic influence on heritability, since the genetic variance is limited to the common SNPs present in the arrays [29]. Therefore variance explained by rarer variants (MAF<0.01) and/or causal variants that were not genotyped or are not tagged by the SNPs on the genotyping array will be missed. As GCTA is intended for studies on homogeneous populations it is necessary to use appropriate methods to calculate the relatedness matrix in studies with admixed individuals to avoid overestimation of the heritability. Therefore, to obtain unbiased estimates we used REAP for the estimation of the kinship coefficient [30].

Methods based on LMM enabled the analysis of GWAS data in a cohort of mixed ethnicity of relative small sample size, maximizing power. The fact that our GWAS identified well-established variants in the *HLA-B* locus, provides certainty that the methodological approach was sound and in fact applicable to other rare diseases. Expanding the current case collection or performing meta-analysis with other collections is warranted. Efforts including datasets of different ancestries, should be analyzed with methodologies which take into account different genomic architecture based on ethnic background. Such methods can also be applied at the meta-analysis level as implemented in MANTRA [31].

In addition to the signal in the well-known HLA locus on chromosome 6, we also identified two genomic regions of interest. Nevertheless, all variants had relatively low MAF (2-3%) and thus their significance should be interpreted with caution. We could not find evidence for replication of the associations in an independent yet underpowered case/control set. Considering the low MAF and the very limited power of the replication further scrutiny in an expanded dataset is required to confirm these variants as real or spurious associations.

Yet another GWS association with rs17810546 in the *IL12A* locus was identified by meta-analyses of our results with those reported previously in the literature. These results support the role of this locus in the susceptibility of BD. Kirino et al., described this locus as suggestive for association with BD. In that study the variants did not were reach GWS in the combined analysis likely due to the variant being monomorphic in the samples of Japanese origin included in the study. Variants in *CCR1-CCR3, STAT4* and *KLRC4* which were polymorphic in both Turkish and Japanese populations surpassed GWS thresholds, suggesting that the lower power likely contributed to the non-significant findings in the *IL12A* locus.

BD is a rare disease, even in countries with the highest prevalence. Collecting the patient material for large multi-ethnic genetic studies proved to be challenging. Not surprisingly, the first two GWAS studies reported in BD were in two separate cohorts with a single ethnicity, a Turkish and a Japanese cohort [10,11]. After meta-analyses of the data of both studies, polymorphisms in *IL10* and *IL23R–IL12RB2* proved to be associated at GWS level. In our study, these two polymorphisms were not associated with BD, nor in two other recent studies in Chinese and Korean individuals [32,33]. Moreover, the GIMAP variants presented in the Chinese and a Korean cohort were not identified in the other studies. Recently, Ortiz-Fernandez et al., reported no association of GIMAP variants in a Spanish study but also acknowledged lack of power as a likely cause for this negative results [16]. Considering that most of our cases are of admixed ethnicity differences in ethnic background, specific pathology or clinical features of the cases are a more plausible explanation for the lack of replication of those variants in our study on top of limited power [32, 33].

In conclusion we have shown that appropriate methodological approaches enable performing GWAS in cohorts of admixed ethnicity, combining cases of different ethnic origin as is often required in rare conditions like BD. As such, previously reported SNPs in the *MICA-HLA-B* locus reached GWS in the current approach. Additionally, we have shown that results of a mixed cohort can successfully be used in meta-analysis, by which we could establish the involvement of previously putative variant in the *IL12A* locus in the susceptibility of the disease.

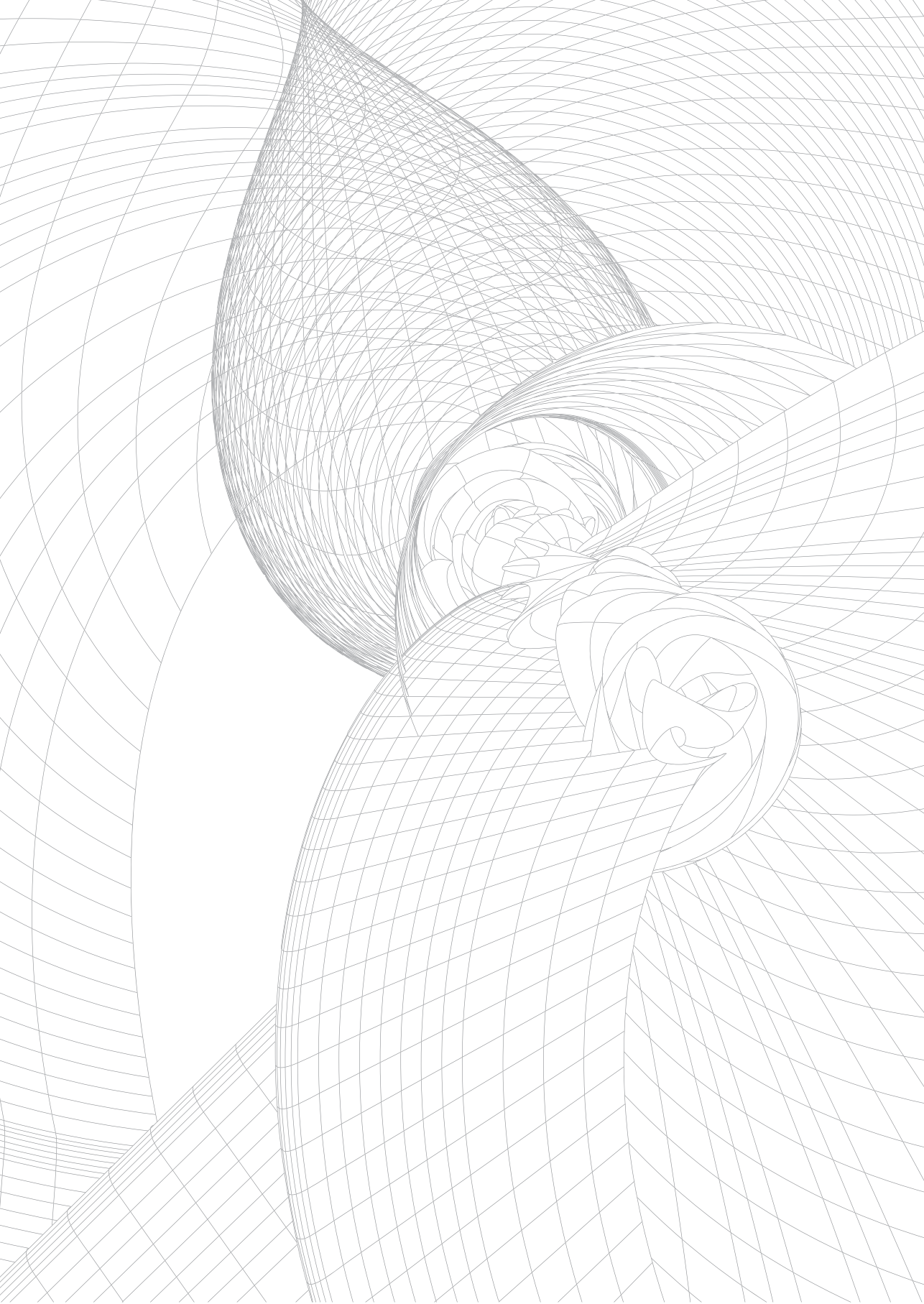
Our results emphasize the need of collaborative efforts intending to enlarge the collection of BD cases, which can enable a well-powered setting for detection of genetic associations with the disease, especially when interrogating low-frequency variants. Potentially novel loci need to be further explored in additional studies, and ultimately coupled with functional studies, allowing a better understanding of the pathophysiology and the genetic architecture underlying the risk for BD.

Detailed acknowledgements and online resources can be found in the published article online: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119085

REFERENCES

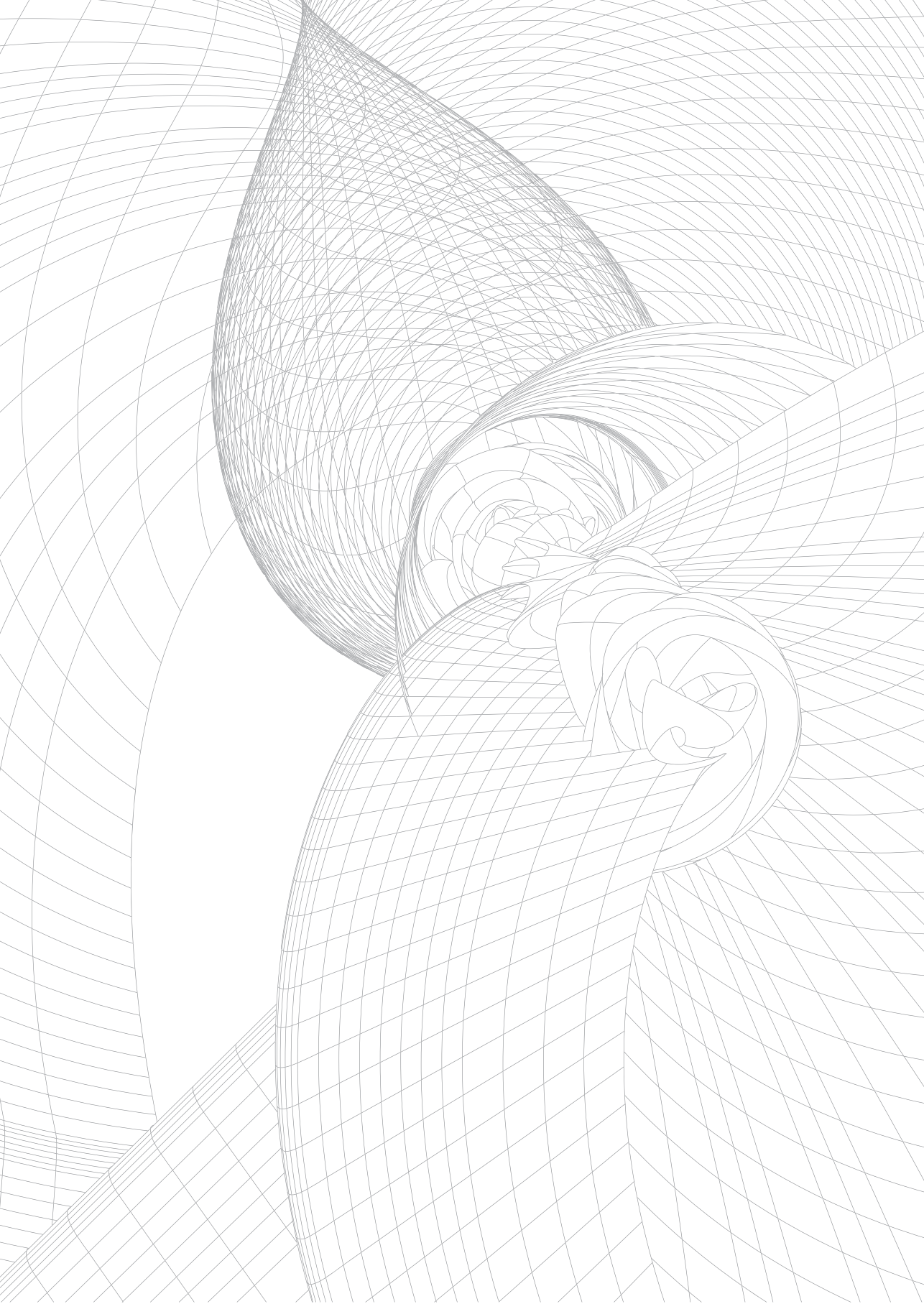
1. Sakane T, Takeno M, Suzuki N, Inaba G. Behcet's disease. **The New England journal of medicine**. 1999;341(17):1284-91.
2. Gul A. Behcet's disease: an update on the pathogenesis. **Clinical and experimental rheumatology**. 2001;19(5 Suppl 24):S6-12.
3. Yazici H, Fresko I, Yurdakul S. Behcet's syndrome: disease manifestations, management, and advances in treatment**. Nature clinical practice Rheumatology**. 2007;3(3):148-55.
4. Altenburg A, Mahr A, Maldini C, Kneifel CE, Krause L,et al. [Epidemiology and clinical aspects of Adamantiades-Behcet disease in Gemany. Current data]. Der Ophthalmologe : Zeitschrift der Deutschen Ophthalmologischen Gesellschaft. 2012;109(6):531-41. Epub 2012/06/16. Epidemiologie und Klinik des Morbus Adamantiades-Behcet in Deutschland. Aktuelle Daten.
5. Gul A, Inanc M, Ocal L, Aral O, Konice M. Familial aggregation of Behcet's disease in Turkey. **Annals of the rheumatic diseases**. 2000;59(8):622-5.
6. Baranathan V, Stanford MR, Vaughan RW, Kondeatis E, Graham E, et al. The association of the PTPN22 620W polymorphism with Behcet's disease. **Annals of the rheumatic diseases**. 2007;66(11):1531-3.
7. Kappen JH, Wallace GR, Stolk L, Rivadeneira F, Uitterlinden AG, et al. Low prevalence of NOD2 SNPs in Behcet's disease suggests protective association in Caucasians. **Rheumatology (Oxford)**. 2009; 48(11):1375-7.
8. Touma Z, Farra C, Hamdan A, Shamseddeen W, Uthman I, et al. TNF polymorphisms in patients with Behcet disease: a meta-analysis. **Archives of medical research**. 2010;41(2):142-6. Epub 2010/05/18.
9. Hu K, Yang P, Jiang Z, Hou S, Du L, Li F. STAT4 polymorphism in a Chinese Han population with Vogt-Koyanagi-Harada syndrome and Behcet's disease. **Human immunology**. 2010;71(7):723-6.
10. Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behcet's disease. **Nature genetics**. 2010;42(8):698-702. Epub 2010/07/14.
11. Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, et al. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behcet's disease susceptibility loci. **Nature genetics**. 2010;42(8):703-6. Epub 2010/07/14.
12. Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, et al. Association study of IL10 and IL23R-IL12RB2 in Iranian patients with Behcet's disease. **Arthritis and rheumatism**. 2012;64(8):2761-72.
13. Khaib Dit Naib O, Aribi M, Idder A, Chiali A, Sairi H, Touitou I, et al. Association Analysis of IL10, TNF-alpha, and IL23R-IL12RB2 SNPs with Behcet's Disease Risk in Western Algeria. **Frontiers in immunology**. 2013;4:342. Epub 2013/10/24.
14. Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, et al. Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. **Nature genetics**. 2013;45(2):202-7.
15. Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, et al. Genome-wide association study identifies GIMAP as a novel susceptibility locus for Behcet's disease. **Annals of the rheumatic diseases**. 2013;72(9): 1510-6.
16. Ortiz-Fernandez L, Conde-Jaldon M, Garcia-Lozano JR, Montes-Cano MA, Ortego-Centeno N, et al. GIMAP and Behcet disease: no association in the European population. **Annals of the rheumatic diseases.** 2014.
17. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. **Human molecular genetics**. 2008;17(R2):R156-65.
18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature genetics**. 2006;38(8):904-9.
19. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. **Nature reviews Genetics**. 2010;11(7):459-63.

21. Criteria for diagnosis of Behcet's disease. International Study Group for Behcet's Disease. **Lancet**. 1990;335(8697):1078-80.
21. Jaddoe VW, van Duijn CM, Franco OH, van der Heijden AJ, van Iizendoorn MH et al. The Generation R Study: design and cohort update 2012. **European journal of epidemiology**. 2012;27(9):739-56.
22. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. **American journal of human genetics**. 2011;88(1):76-82.
23. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, et al. Estimating kinship in admixed populations. **American journal of human genetics**. 2012;91(1):122-38.
24. International HapMap C. The International HapMap Project. **Nature**. 2003;426(6968):789-96.
25. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, et al. Variance component model to account for sample structure in genome-wide association studies. **Nature genetics**. 2010;42(4):348-54.
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. **Bioinformatics**. 2010;26(17):2190-1.
27. A haplotype map of the human genome. **Nature**. 2005;437(7063):1299-320.
28. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. **PLoS genetics**. 2013;9(5):e1003520.
29. Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. **Molecular psychiatry**. 2014.
30. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. **American journal of human genetics**. 2013;92(6):904-16.
31. Morris AP. Transethnic meta-analysis of genomewide association studies. **Genetic epidemiology**. 2011;35(8):809-22.
32. Hou S, Yang Z, Du L, Jiang Z, Shu Q, et al. Identification of a susceptibility locus in STAT4 for Behcet's disease in Han Chinese in a genome-wide association study. **Arthritis and rheumatism**. 2012;64(12): 4104-13.
33. Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, et al. Genome-wide association study identifies GIMAP as a novel susceptibility locus for Behcet's disease. **Annals of the rheumatic diseases**. 2013 Sep 1; 72(9):1510-6



Chapter 3

Epidemiological studies of pediatric BMD



Chapter 3.1

Maternal first-trimester diet and childhood bone mass: the Generation R Study

Denise HM Heppe, Carolina Medina-Gomez, Albert Hofman, Oscar H Franco, Fernando Rivadeneira, and Vincent WV Jaddoe

Am J Clin Nutr. 2013 Jul;98(1):224-32. doi: 10.3945/ajcn.112.051052

ABSTRACT

Background: Maternal diet during pregnancy has been suggested to influence bone health in later life. Objective: We assessed the association of maternal first-trimester dietary intake during pregnancy with childhood bone mass. **Design**: In a prospective cohort study in 2,819 mothers and their children, we measured first-trimester daily energy, protein, fat, carbohydrate, calcium, phosphorus, and magnesium intakes by using a food-frequency questionnaire and homocysteine, folate, and vitamin B-12 concentrations in venous blood. We measured childhood total body bone mass by using dual-energy X-ray absorptiometry at the median age of 6 years. **Results**: Higher first-trimester maternal protein, calcium, and phosphorus intakes and vitamin B-12 concentrations were associated with higher childhood bone mass, whereas carbohydrate intake and homocysteine concentrations were associated with lower childhood bone mass (all P-trend, 0.01). Maternal fat, magnesium intake, and folate concentrations were not associated with childhood bone mass. In the fully adjusted regression model that included all dietary factors significantly associated with childhood bone mass, maternal phosphorus intake and homocysteine concentrations most strongly predicted childhood bone mineral content (BMC) [β = 2.8 (95% CI: 1.1, 4.5) and β = 21.8 (95% CI: 23.6, 0.1) g per SD increase, respectively], whereas maternal protein intake and vitamin B-12 concentrations most strongly predicted BMC adjusted for bone area [β = 2.1 (95% CI: 0.7, 3.5) and β = 1.8 (95% CI: 0.4, 3.2) g per SD increase, respectively]. **Conclusion**: Maternal first-trimester dietary factors are associated with childhood bone mass, suggesting that fetal nutritional exposures may permanently influence bone development.

INTRODUCTION

Osteoporosis represents a major public health problem [1–3]. Previous studies have shown the relevance of early life factors for the development of bone health and osteoporosis during the life course [4]. The maternal diet during pregnancy, which is the main determinant of fetal nutrition, has been suggested to influence childhood bone mass [5–9]. In mother-offspring cohorts in Tasmania, India, and the United Kingdom, maternal milk, magnesium, calcium, and folate intakes during pregnancy were positively associated with bone mass in their offspring [5–8], whereas fat intake showed an inverse association [8].

Although these finding may reflect an effect of shared genetic or environmental determinants, they support the hypothesis of bone-health programming via direct or indirect effects of nutrient availability in early life. It is unknown which mechanisms may underlie such a programming effect. Nutrients such as calcium and phosphorus, which are the main bone-forming minerals, may enhance fetal bone accrual in a direct manner [10]. In addition, indirect effects of nutrient availability may act through changes in the growth hormone/insulin-like growth factor I axis or alterations in glucocorticoid concentrations, such as cortisol [11,12]. Two other candidate hormonal mediators of bone-diet interactions are the adipokine leptin and osteocalcin [13, 14]. In addition, epigenetic changes that are a result of altered DNA methylation or histone modification have been suggested to be key players in this programming process and should also be considered [15].

Maternal folate intake and blood concentrations were positively associated with childhood bone mass in 2 studies that focused on this nutrient [5,7]. Folate and other B vitamins (riboflavin, vitamin B-6, and vitamin B-12) are involved in the homocysteine metabolism [16] and may reduce adverse effects of hyperhomocysteinemia, which is a strong risk factor for osteoporotic fractures in older adults [17]. Alternatively, by acting as methyl donors, a beneficial effect of high maternal folate and B-vitamin concentrations on childhood bone mass may support the hypothesis of epigenetic changes that program later bone health. Because homocysteine and B-vitamin concentrations are potentially modifiable, we investigated the influence of maternal dietary factors on the bone health of their children with particular focus on the homocysteine metabolism. We assessed associations of maternal first-trimester dietary nutrient intake and homocysteine, folate, and vitamin B-12 concentrations with childhood bone mass in a population-based prospective cohort study that included 2,819 Dutch mothers and 6-y-old children.

MATERIALS AND METHODS

Study population

This study was embedded in the Generation R Study, which is a population-based prospective cohort study from fetal life onwards in Rotterdam, Netherlands [18]. All mothers who were resident in the study area at their delivery dates and had a delivery date from April 2,002 until January 2,006 were eligible. We aimed for enrollment in first trimester, but we allowed enrollment until the delivery of the child. In total, 75% of all mothers enrolled before a gestational age of 18 wk. Of all eligible children in the study area, 61% of children participated at birth in the study. We conducted the study according to the guidelines of the Helsinki Declaration, and it was approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam (MEC 198.782/2001/31). We obtained written informed consent from all participants. We restricted the study population to 4,097 mothers of Dutch ethnicity because the food-frequency questionnaire (FFQ) was validated for the assessment of dietary intake in a Dutch population [19]. We defined the ethnicity of the mother according to the classification of Statistics Netherlands [20,21]. Of these Dutch mothers, 4,016 mothers enrolled in the first or second trimester of pregnancy. We measured dietary intake

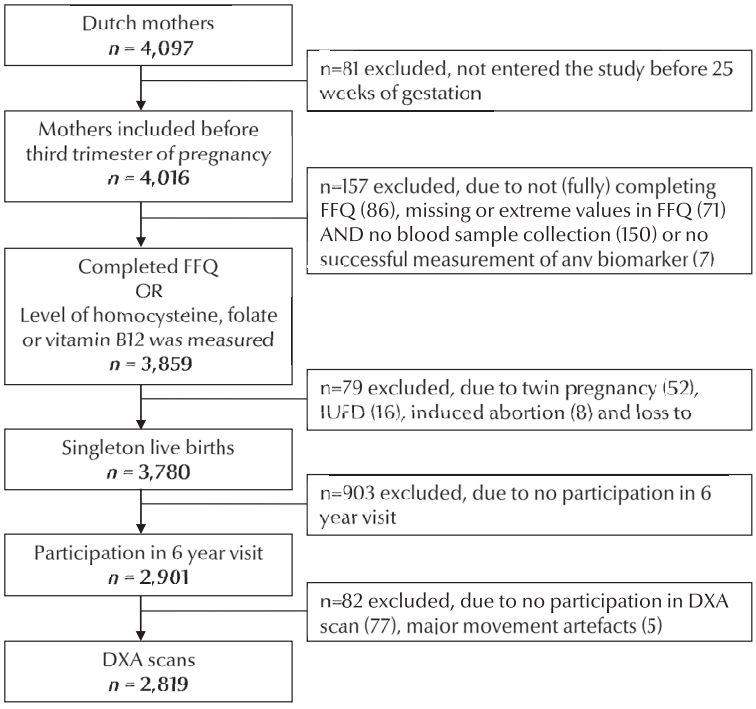


Figure 1. Flowchart of participants included for analysis in the Generation R Study, Rotterdam, Netherlands. DXA, dual-energy X-ray absorptiometry; FFQ, food-frequency questionnaire; IUFD, intrauterine fetal death.

and/or homocysteine, folate, or vitamin B-12 concentrations in 3,859 of these mothers. Their pregnancies resulted in 3,780 singleton live births. Of these births, 2,901 children participated in the 6-y visit, and we performed dual-energy X-ray absorptiometry (DXA) scans in 2,819 children (69%). Because the exclusion of mothers with a second (n = 241) or third (n = 2) pregnancy in the study did not change our results, we included these pregnancies in the analyses (Figure 1).

Maternal dietary intake

We assessed maternal dietary intake at enrollment in the study (median: 13.5 wk of gestation, IQR: 3.3 wk) by using a modified version of the validated semiquantitative FFQ of Klipstein-Grobusch et al [19]. We validated this modified version of the FFQ against 3-d 24-h recalls in 82 Dutch pregnant women visiting their midwives in Rotterdam, Netherlands. We considered the FFQ valid for studying macronutrient and micronutrient intakes in Dutch pregnant women because the intraclass correlation coefficients for the (energy-adjusted) nutrients were 0.65 for protein, 0.54 for fat, 0.54 for carbohydrate, 0.75 for calcium, and 0.78 for phosphorus intakes, whereas for total energy intake, it was only 0.33. The FFQ considered food intake over the previous 3 mo, which mostly covered dietary intake within the first trimester of pregnancy. The FFQ consisted of 293 items structured according to the meal pattern. Questions included consumption frequency, portion size, preparation method, and additions. We estimated portion sizes by using Dutch household measures and photographs of foods that showed different portion sizes [22]. To calculate average daily nutritional values, we used the Dutch Food-Composition Table 2006 [23]. In the current study, we analyzed intakes of macronutrients and intakes of calcium, phosphorus, and magnesium following results of earlier studies [5–8].

Maternal homocysteine, folate, and vitamin B-12 concentrations

Venous blood samples were drawn at enrollment in early pregnancy (median: 12.9 wk of gestation; IQR: 2.3 wk) and stored at room temperature for a maximum of 3 h. Blood samples were transported to a dedicated laboratory facility of the regional laboratory in Rotterdam, Netherlands (Star-Medisch Diagnostisch Centrum) for additional processing and storage at −80°C (24). To analyze homocysteine, folate, and vitamin B-12 concentration, we picked and transported the serum (vitamin B-12) and plasma (homocysteine and folate) samples to the Department of Clinical Chemistry at the Erasmus University Medical Centre, Rotterdam in 2008. After thawing, we analyzed homocysteine, folate, and vitamin B-12 concentrations by using an immunoelectrochemoluminence assay on the Architect System (Abbott Diagnostics BV). Between-run CVs for plasma measurements were as follows 3.1% at 7.2 μmol/L, 3.1% at 12.9 μmol/L, and 2.1% at 26.1 μmol/L with an analytic range of 1–50 μmol/L for homocysteine; 8.9% at 5.6 nmol/L, 2.5% at 16.6 nmol/L, and 1.5% at 33.6 nmol/L with an analytic range of 1.8–45.3 nmol/L for plasma folate; and 3.6% at 142 pmol/L,

7.5% at 308 pmol/L, and 3.1% at 633 pmol/L with an analytic range of 44–1,476 pmol/L for serum vitamin B-12.

Bone mass measurements

We measured the total body bone mineral density (BMD), bone mineral content (BMC), and bone area by using a dual-energy X-ray absorptiometry (DXA) scan (iDXA; General Electrics–Lunar) at the median age of 6.0 y (IQR: 0.37 y). Well-trained research assistants obtained DXA scans by using the same device and software (enCORE, version 13; GE Healthcare) following standard manufacturer protocols. We performed daily quality assurance by using a spine phantom. In 2,011, the longitudinal CV for BMD was 0.23. Before the scan procedure, we asked participants to take off their shoes, heavy clothes, and metal accessories. We measured weight to the nearest 0.2 kg by using an electronic personal scale (Seca) and height to the nearest 0.1 cm by using a Harpenden stadiometer (Holtain Ltd). Thereafter, we placed participants in a supine position on the scan and instructed them to hold their hands flat with the palms down on the scanner table and arms alongside the body. To ensure that the lines between adjacent subregions of the body were placed correctly, scans were evaluated twice, directly after the scanning procedure and at a later time point by a second well-trained research assistant.

In our analyses, we used areal total-body-less-head BMD and BMC as recommended by the International Society for Clinical Densitometry [25]. Bone-size effects should be considered in studies focused on BMC and areal BMD measured during periods of growth [26]. In subjects with large bones, areal BMD alone can overestimate the volumetric BMD. In studies focused on BMD, Prentice et al [27] suggested to use BMC corrected for bone area, weight, and height to correctly adjust for size effects. Therefore, in a sensitivity analysis, we further adjusted BMC models for bone area because weight and height were already included.

Covariates

We registered maternal age and measured height and weight without shoes and heavy clothing at enrollment. We also collected information about weight just before pregnancy, maternal education, marital status, parity, ethnicity, and ethnicity of the grandparents and father by using a questionnaire at enrollment in the study. Because the enrollment in our study was during pregnancy, we were not able to measure maternal weight before pregnancy. However, the correlation of pre-pregnancy weight obtained by questionnaire and weight measured at enrollment was high ($\rho = 0.97$, $P < 0.01$). Maternal smoking and alcohol habits were assessed by questionnaires in each trimester. We obtained information about offspring sex, gestational age, and birth weight from medical records and hospital registries and about breastfeeding from questionnaires at 2, 6, and 12 mo after delivery (28). We as-

sessed the child's diet around the age of 14 mo by using a modified version of a validated semi-quantitative FFQ [29] that consisted of 211 food items. This FFQ has been validated against 3-d 24-h recalls in Dutch children aged 14 mo and showed the following intraclass correlation coefficients: 0.7 for total protein; 0.4 for total fat; 0.4 for carbohydrate, and 0.5 for calcium intakes. Because this FFQ was implemented in a later stage in the study, it was only available in a subgroup of the population (59%).We obtained information about the frequency of participation in sports from a questionnaire at 6 y of age.

Statistical analysis

We used t-tests and chi-square tests to compare differences in subject characteristics between boys and girls. We categorized dietary nutrient intake (fat, protein, carbohydrate, calcium, phosphorus, and magnesium) and blood concentrations (homocysteine, folate, and vitamin B-12) into quintiles and used the lowest category as the reference category in all models. We performed multiple linear regression analyses to assess associations of dietary factors with childhood bone measures at the age of 6 y. We performed tests for trends by using dietary intakes and blood concentrations as continuous variables in the models. We log- and square root–transformed blood concentrations that were not normally distributed and lacked goodness of fit. On the basis of previous studies, we adjusted models for maternal age, weight, height, parity, educational level, marital status, alcohol use, and smoking and sex, birth weight, ethnicity, breastfeeding, participation in sports, and age and height at the time of the measurement of the child [5–8]. We adjusted for weight at the time of the measurement by adding lean mass plus fat mass to the model, which thereby excluded the contribution of bone mass to the child's weight. Because maternal total energy intake and macronutrient intakes were strongly correlated (Online Resource Table S1), we used the energy-partition method to adjust for total energy intake in the macronutrient analyses [30]. In the micronutrient analysis, we used the residual nutrient method. We entered dietary factors shown to be associated with childhood bone mass at a $P < 0.10$ significance level in a backward-selection regression model to identify the major dietary factors of bone mineral accrual in childhood ($P < 0.10$). We assessed potential multicollinearity between nutrients by exploring the change in effect estimates and calculating the variance inflation factor [31]. We performed a sensitivity analysis to assess whether associations of maternal nutrient intake with childhood bone mass could be explained by the later diet of the child rather than maternal diet. We added available information on protein, fat, carbohydrate, or calcium intake of the child at 14 mo to each corresponding model. We did not add the nutrient intake of the child to our final models because data were missing in 41% of the study population because nutrient intake of the child was only assessed in a subgroup [18]. In another sensitivity analysis, we further adjusted models that concerned BMC for the total-body bone area to assure the correct size adjustment [27]. To prevent bias associated with missing data, we used multiple imputations (5 imputations) for covariates with missing values on the basis

of the correlation of missing variables with other participant characteristics, according to the Markov Chain Monte Carlo method [32]. In addition, we added variables related to covariates as predictors to the imputation model to increase the plausibility of the missing-at-random assumption. The amount of missing values ranged from 1% to 14%, except for breastfeeding duration (26%). We report the pooled results of the analyses performed in each of the 5 imputed datasets. P values (P) were 2-sided. We performed analyses with the Predictive Analytic Software (version 17.0 for Windows; PASW Inc).

RESULTS

Characteristics of study participants

Boys were born with a higher birth weight, were taller, and more often participated frequently in sports at age 6 y than did girls (Table 1). Bone measures did not differ between boys and girls. The average maternal nutrient intake and homocysteine, folate, and vitamin B-12 concentrations in the first trimester of pregnancy are shown in Table 2. See Online Resource Table S2 for characteristics of the imputed study population.

Table 1. Characteristics of study participants in the Generation R Study Cohort, Rotterdam, Netherlands. All values are means ± SDs for continuous variables and absolute numbers (percentages) for categorical variables. P values were obtained by using Student's t tests for continuous variables and chi-square tests for categorical variables.

	Boys (n = 1409)	Girls (n = 1410)	P
Maternal characteristics			
Height (cm)	170.9 ± 6.2	170.9 ± 6.6	0.99
Prepregnancy weight (kg)	68.1 ± 12.5	68.3 ± 12.2	0.59
BMI (kg/m2)	23.3 ± 4.0	23.4 ± 4.0	0.5
Age (y)	31.6 ± 4.3	31.5 ± 4.1	0.41
Parity ≥1			
No	877 (62)	866 (61)	0.25
Yes	531 (38)	539 (38)	
Missing	1 (0)	5 (0)	
Single motherhood			
No	1279 (92)	1296 (93)	0.56
Yes	85 (6)	74 (5)	
Missing	45 (3)	40 (3)	
Educational status			
Primary	24 (2)	38 (3)	0.16
Secondary	495 (35)	519 (37)	
Higher	866 (62)	835 (59)	

Table 1. (continued)

	Boys (n = 1409)	Girls (n = 1410)	P
Missing	24 (2)	18 (1)	
Smoking during pregnancy			
Never	958 (68)	988 (70)	0.1
Until pregnancy was known	116 (8)	136 (10)	
Continued	207 (15)	184 (13)	
Missing	128 (9)	102 (7)	
Alcohol use during pregnancy			
No	597 (42)	640 (45)	0.07
Yes	694 (49)	680 (48)	
Missing	118 (8)	90 (6)	
Paternal characteristic			
Ethnicity			
Dutch	1115 (79)	1108 (79)	0.9
Other	266 (19)	271 (19)	
Missing	28 (2)	31 (20)	
Infant characteristics			
Gestational age at birth (wk)	40.0 ± 1.6	40.0 ± 1.8	0.26
Birth weight (g)	3564 ± 540	3434 ± 539	<0.001
Birth length (cm)	50.9 ± 2.4	50.1 ± 2.3	<0.001
Breastfeeding			
Never	103 (7)	106 (8)	0.8
≤3 mo	365 (26)	347 (25)	
>3 mo	583 (41)	587 (42)	
Missing	358 (25)	370 (26)	
Child characteristics at 6-y visit			
Age (y)	6.15 ± 0.46	6.14 ± 0.42	0.37
Height (cm)	119.9 ± 5.7	119.2 ± 5.7	<0.001
Weight (kg)	23.0 ± 3.4	22.7 ± 3.7	0.05
BMI (kg/cm2)	15.9 ± 1.4	15.9 ± 1.6	0.91
Participation in sports			
Never	683 (49)	636 (45)	<0.001
1/wk	445 (32)	613 (44)	
≥2/wk	183 (13)	52 (4)	
Missing	98 (7)	109 (8)	
Bone mineral density (g/cm2)	0.551 ± 0.046	0.548 ± 0.045	0.12
Bone mineral content (g)	520 ± 95	519 ± 93	0.78
Bone area (cm2)	940 ± 111	942 ± 107	0.56

Table 2. Maternal first-trimester dietary intake and blood concentrations in the Generation R Study Cohort, Rotterdam, Netherlands. Dietary intake has been energy adjusted.

Dietary intake	n	Values **Mean ± SD**
Energy (kcal/d)	2580	2152 ± 5042
Fat (g/d)	2580	73.6 ± 10.7
Protein (g/d)	2580	78.5 ± 11.7
Carbohydrate (g/d)	2580	236 ± 30
Calcium (mg/d)	2580	1108 ± 311
Phosphorus (mg/d)	2580	1443 ± 241
Magnesium (mg/d)	2580	339 ± 56
		Median; IQR
Blood concentrations		
Homocysteine (μmol/L)	2260	7.0 (13.3)
Folate (nmol/L)	2282	19.5 (1.9)
Vitamin B-12 (pmol/L)	2173	175 (100)

Maternal dietary intakes and childhood bone mass

Higher maternal protein intake in the first trimester of pregnancy was associated with a higher childhood BMC (P-trend = 0.02) and density (P-trend < 0.001). As shown in Table 3, the difference in BMC between highest and lowest (reference) quintiles of maternal protein intake was 7.0 g (95% CI: 0.1, 13.9 g). Maternal carbohydrate intake was inversely associated with childhood BMC and density (both P-trend = 0.02). The difference in BMC between highest and lowest quintiles was −6.9 g (95% CI: −12.8, −0.9 g). Neither maternal total energy intake nor fat intake were associated with the bone mass of the child (P-trend > 0.05) (data for total energy not shown).

Maternal homocysteine, folate, and vitamin B-12 concentrations and childhood bone mass

Higher maternal first-trimester homocysteine concentrations were associated with a lower childhood BMC (P-trend = 0.03) and showed a borderline significant association with BMD (P-trend = 0.06) (Table 5). The difference in BMC between highest and lowest quintiles was −5.2 g (95% CI: −10.4, 0.1 g). Maternal folate concentrations were not associated with childhood bone outcomes. Higher maternal vitamin B-12 concentrations were not associated with absolute bone mineral accrual but were associated with BMD (P-trend < 0.001), which suggested an effect on bone size rather than bone mineral accrual.

The first sensitivity analysis showed that effect estimates did not change when corresponding nutrient intakes of the child at 14 mo were added to each of the models (see Online Resource Table S3). Likewise, when we further adjusted BMC models for bone area in a second sensitivity analysis, results hardly changed (see Online Resource Table S4).

Table 3. Associations of maternal first-trimester daily macronutrient intake with childhood bone mass in the Generation R Study Cohort, Rotterdam, Netherlands. Values are based on multiple linear regression models and reflect the difference (95% CI) for each quintile of daily macronutrient intake compared with the lowest reference quintile. Models were adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, and smoking; paternal ethnicity; and sex, birth weight, birth length, breastfeeding, participation in sports, and age, weight measured as lean mass plus fat mass, and height at measurement of the child and mutually for other macronutrients following the energy-partition method (30). *P < 0.05, **P < 0.01, ***P < 0.001.

Macronutrient intake (kcal/d)	n	Bone mineral content g **ß (95% CI)**	Bone mineral density mg/cm^2 **ß (95% CI)**
Energy from proteins			
Quintile 1	516	Reference	Reference
Quintile 2	516	6.4 (1.3, 11.4)*	6.1 (2.0, 10.2)**
Quintile 3	516	6.3 (0.9, 11.7)*	8.8 (4.4, 13.2)***
Quintile 4	516	10.1 (4.0, 16.1)**	11.4 (6.5, 16.3)***
Quintile 5	516	7.0 (0.1, 13.9)*	11.9 (6.3, 17.5)***
P-trend	—	0.02	<0.001
Energy from fat			
Quintile 1	516	Reference	Reference
Quintile 2	516	−0.1 (−5.1, 4.9)	−2.9 (−7.0, 1.1)
Quintile 3	516	−0.1 (−5.5, 5.3)	0.0 (−4.4, 4.3)
Quintile 4	516	−1.3 (−7.1, 4.5)	−3.7 (−8.3, 1.0)
Quintile 5	516	0.9 (−5.7, 7.5)	−3.8 (−9.1, 1.5)
P-trend	—	0.46	0.31
Energy from carbohydrate			
Quintile 1	516	Reference	Reference
Quintile 2	516	−2.6 (−7.6, 2.3)	−1.8 (−5.8, 2.2)
Quintile 3	516	-5.4 (-10.6, −0.2)*	−3.7 (−7.9, 0.5)
Quintile 4	516	−3.9 (−9.5, 1.6)	-4.9 (-9.3, −0.5)*
Quintile 5	516	-6.9 (-12.8, −0.9)*	−4.6 (−9.4, 0.2)
P-trend	—	0.02	0.02

To identify the major dietary factors that influencing bone mineral accrual in childhood and take into account the effect of other dietary factors, we entered all dietary factors associated with childhood bone mass at the P < 0.10 significance level in a backward-selection regression model. When simultaneously analyzed, maternal phosphorus intake was the only dietary factor significantly associated with childhood BMC (P-trend = 0.002), whereas maternal homocysteine concentrations showed a borderline significant association (P-trend = 0.058) (Table 6). Maternal protein intake (P-trend = 0.003) and vitamin B-12 concentrations (P-trend = 0.01) showed the strongest association with childhood BMD. The other maternal dietary factors did not add significantly to the explained variation. Multicollinearity

Table 4. Associations of maternal first-trimester daily calcium, phosphorus, and magnesium intakes with childhood bone mass, the Generation R Study Cohort, Rotterdam, Netherlands. Values are based on multiple linear regression models and reflect the difference (95% CI) for each quintile of daily micronutrient intake compared with the lowest reference quintile. Models were adjusted for maternal total daily energy intake, age, weight, height, parity, educational level, marital status, alcohol use, and smoking; paternal ethnicity; and sex, birth weight, birth length, breastfeeding, participation in sports, and age, weight measured as lean mass plus fat mass, and height at measurement of the child. *P < 0.05, **P < 0.01, ***P < 0.001.

Micronutrient intake (g/d)	n	Bone mineral content g **ß (95% CI)**	Bone mineral density mg/cm^2 **ß (95% CI)**
Calcium			
Quintile 1	516	Reference	Reference
Quintile 2	516	5.4 (0.6, 10.3)*	4.1 (0.2, 8.0)*
Quintile 3	516	5.7 (0.8, 10.6)*	5.6 (1.7, 9.6)**
Quintile 4	516	5.6 (0.7, 10.6)*	4.0 (0.0, 8.0)*
Quintile 5	516	9.1 (4.1, 14.1)***	6.4 (2.4, 10.4)**
P-trend	—	<0.001	0.005
Phosphorus			
Quintile 1	516	Reference	Reference
Quintile 2	516	8.6 (3.6, 13.5)***	6.5 (2.5, 10.5)**
Quintile 3	516	4.3 (−0.7, 9.3)	3.8 (−0.2, 7.8)
Quintile 4	516	7.0 (2.0, 12.1)**	6.1 (2.0, 10.1)**
Quintile 5	516	8.6 (3.5, 13.6)***	7.5 (3.4, 11.5)***
P-trend	—	<0.001	<0.001
Magnesium			
Quintile 1	516	Reference	Reference
Quintile 2	516	−0.8 (−5.7, 4.2)	−0.1 (−4.0, 3.9)
Quintile 3	516	−0.4 (−5.4, 4.7)	0.5 (−3.6, 4.5)
Quintile 4	516	4.4 (−0.7, 9.5)	0.6 (−3.5, 4.7)
Quintile 5	516	2.3 (−3.0, 7.6)	0.4 (−3.9, 4.6)
P-trend	—	0.23	0.76

between maternal protein, phosphorus, and calcium intakes seemed present. Therefore, we explored potential collinearity between nutrients in the final model. We observed that effect estimates for the associations of protein, phosphorus, and calcium intakes with bone outcomes varied largely when mutually adjusted for one another, and the variance inflation factors were 4.8, 3.8, and 8.0 for protein, calcium, and phosphorus intakes. These findings suggested that collinearity between these nutrients is present.

Table 5. Associations of maternal first-trimester homocysteine, folate, and vitamin B-12 concentrations with childhood bone mass in the Generation R Study Cohort, Rotterdam, Netherlands. Values are based on multiple linear regression models and reflect the difference (95% CI) for each quintile of blood concentration compared with the lowest reference quintile. Models are adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, smoking, and gestational age at blood collection; paternal ethnicity; and sex, birth weight, birth length, breastfeeding, participation in sports, and age, weight measured as lean mass plus fat mass, and height at measurement of the child. *P < 0.05, **P < 0.01, ***P < 0.001.

Micronutrient intake (g/d)	n	Bone mineral content g ß (95% CI)	Bone mineral density mg/cm^2 ß (95% CI)
Homocysteine			
Quintile 1	485	Reference	Reference
Quintile 2	454	-0.9 (-6.0, 4.2)	-3.3 (-7.4, 0.8)
Quintile 3	441	-3.1 (-8.3, 2.0)	-2.1 (-6.3, 2.1)
Quintile 4	436	-3.8 (-9.0, 1.3)	-7.1 (-11.3, -2.9)***
Quintile 5	444	-5.2 (-10.4, 0.1)	-4.7 (-9.0, -0.5)*
P-trend	—	0.03	0.06
Folate			
Quintile 1	457	Reference	Reference
Quintile 2	458	-0.1 (-5.4, 5.1)	0.9 (-3.4, 5.2)
Quintile 3	456	0.7 (-4.6, 6.1)	0.4 (-3.9, 4.7)
Quintile 4	456	2.9 (-2.5, 8.4)	0.6 (-3.8, 5.0)
Quintile 5	455	-0.1 (-5.6, 5.3)	1.1 (-3.3, 5.6)
P-trend	—	0.95	0.52
Vitamin B-12			
Quintile 1	439	Reference	Reference
Quintile 2	430	-1.6 (-6.9, 3.7)	-0.7 (-5.0, 3.6)
Quintile 3	442	1.8 (-3.5, 7.1)	2.3 (-1.9, 6.6)
Quintile 4	434	-1.4 (-6.7, 3.9)	3.6 (-0.7, 7.9)
Quintile 5	428	2.5 (-2.9, 7.8)	8.9 (4.6, 13.3)***
P-trend	—	0.39	<0.001

DISCUSSION

Main Findings

In this large, population-based, prospective cohort study of pregnant women in Netherlands, we showed that maternal dietary factors during pregnancy were associated with childhood bone mass. Maternal daily intake of proteins, calcium, and phosphorus and concentrations of vitamin B-12 were positively associated with childhood bone mass, whereas carbohydrates intake and homocysteine concentrations were negatively associated with childhood bone mass. Multiple regression analysis, including all dietary factors associated with childhood bone mass, showed that maternal phosphorus intake and homocysteine

Table 6. Associations of simultaneously analyzed dietary factors with childhood bone mass in the Generation R Study Cohort, Rotterdam, Netherlands. Values are based on multiple linear regression models by using backward selection and reflect the difference (95% CI) for each SD increase in daily nutrient intake or ln-transformed blood concentration.

Dietary factors	Bone mineral content (g) (n=2,029)*		Bone mineral density (mg/cm2) (n=1945)**	
	ß (95% CI)	P-trend	ß (95% CI)	P-trend
Protein (SD)	-	-	**2.6 (1.2, 4.0)**	**<0.001**
Carbohydrate (SD)	-	-	-	-
Phosphorus (SD)	**2.8 (1.1, 4.5)**	**0.002**	-	-
Homocysteine – SD	-1.8 (-3.6, 0.1)	0.058	-	-
Vitamin B12 – SD	-	-	**2.3 (0.8, 3.7)**	**0.002**

*Model was adjusted for covariates that remained in the model as follows: maternal parity, marital status, and gestational age at blood collection; and sex, birth weight, breastfeeding duration, participation in sports, and age, weight measured as lean mass plus fat mass, and height at measurement of the child.'

**Model was adjusted for covariates that remained in the model as follows: maternal smoking; birth weight, birth length, participation in sports, and age, weight measured as lean mass plus fat mass, and height at measurement of the child.

concentrations most-strongly affected absolute bone mineral accrual, whereas maternal protein intake and vitamin B-12 concentrations were the strongest predictors of bone mass adjusted for bone size in childhood.

Interpretation of main results

Maternal phosphorus intake showed the strongest association with childhood bone mass when simultaneously analyzed with other significant dietary factors. However, results from the simultaneous analysis must be carefully interpreted because collinearity between some nutrients was present. Although multicollinearity did not affect the validity of the model as a whole, it complicated the estimation of the individual contribution of predictors in the model. Maternal protein, calcium, and phosphorus intake were highly correlated (Online Resource Table S1), which likely reflected their shared dietary sources [33]. As a result, we were unable to indicate which of the independent nutrients may be most essential for childhood bone health.

Accordingly, in previous studies regarding the association of maternal dietary intakes with offspring bone health, discrepant associations have been described; in the study of Jones et al [6] in 330 Tasmanian mothers and their 8-y old children, maternal phosphorus intake most-strongly predicted childhood bone mass, whereas in the study of Ganpule et al [5] in 797 Indian mothers and their 6-y old children, the intake of calcium-rich foods most-strongly predicted childhood bone mass. From animal studies, phosphorus and calcium are known to be codependent for bone health [34]. Likewise, both may be essential to fetal bone mineral accrual.

In a British study in 4,451 mothers and their 9-y-old children, maternal protein intake showed a positive association with childhood bone mass, which attenuated after a simultaneous analysis with mineral and vitamin intakes [7]. In the studies of Jones et al [6] at 8-y-old and Yin et al [8] in the same population at the age of 16 y, of all macronutrients, maternal fat intake was the strongest negative predictor of childhood bone mass. In our study, maternal fat intake did not show an association. However, because intakes of macronutrients are highly correlated (Online Resource Table S1), it may be that, instead of a higher or lower intake of one specific macronutrient, a certain balance between intakes of macronutrients is the most optimal for childhood bone development. The existence of an optimal balance needs to be further studied.

Maternal homocysteine concentrations showed a negative association with childhood BMC, whereas vitamin B-12 concentrations showed a positive association with BMD. Homocysteine and vitamin B-12 concentrations did not show an association in the previous dian study [5]. The population differed from our study population in size, and moreover, it was reported that 60% of the mothers had low vitamin B-12 concentrations (<150 pmol/L) compared with 26% in our population. Furthermore, the children in this study had very low weight and were short and thin according to NHANES references [35].

Potential Mechanism

Diet during pregnancy has been shown to be associated with epigenetic changes altering postnatal transcriptional activity of genes that affect childhood body composition [36] and possibly bone mass [11, 12]. Protein restriction in pregnant rats has been associated with reduced methylation of the glucocorticoid receptor promoter [37], and increased expression of this receptor could sensitize osteoblasts to cortisol and thereby reduce offspring bone mass [12]. Recently, in umbilical cord tissue, the methylation of the *eNOS* gene, which is the key to osteocyte, osteoblast, and osteoclast function, was shown to be associated with childhood bone mass [38]. Additional studies are needed to confirm a direct link between maternal diet, epigenetic changes, and offspring bone development. Alternatively, higher intakes of protein, which is a component of the organic bone matrix, and calcium and phosphorus, which are the main bone-forming minerals, may lead to increased fetal bone mineral accrual through direct effects [10,39]. In adults, higher protein intake has been shown to increase concentrations of insulin-like growth factor I, which is an osteotrophic factor, as well as intestinal calcium uptake [40, 41]. Vitamin B-12 and folate may play a crucial role in fetal bone health by providing methyl donors for DNA methylation according to the mechanisms suggested above, or it may exert a direct effect by acting as a cofactor for osteoblast function [42,43] or for the metabolism of homocysteine [44,45]. Elevated homocysteine concentrations have been shown to be a strong risk factor for osteoporotic fractures in adulthood [17,46]. The deleterious effect of homocysteine may

involve the disturbance of the bone matrix or a shift of bone metabolism toward bone resorption [47].

Methodological Considerations

Strengths of this study lie in the setting and design. We measured maternal dietary intakes and concentrations combined with childhood bone mass in a large population. In addition, we collected detailed information on many potential confounding variables. However, as in any observational study, residual confounding as a result of unmeasured sociodemographic and lifestyle factors should still be considered, especially because a healthy diet is likely to be associated with a health-conscious lifestyle.

In total, 88% of all eligible mothers fully completed the FFQ, and 79% of mothers participated in the blood sample collection. The bone mass in children of mothers who did not have complete data on dietary factors was not different from the bone mass in children of mothers who had. However, missing information may have led to a loss of power. In mothers whose children did not have DXA data available, intakes of protein, calcium, phosphorus, and magnesium and concentrations of folate and vitamin B-12 were lower, whereas homocysteine concentrations were higher ($P < 0.05$). Our results would be biased if effect estimates differed between those included and not included in the analyses. This bias seems unlikely because a loss to follow-up in a cohort study is unlikely to be related to the research question [48] but cannot be excluded.

Although the FFQ yielded valid estimates of nutrient intakes when validated against 3-d 24-h recalls, measurement error may still have occurred. Because measurement error in nutrient intake is probably random, it might have led to an underestimation of associations. The sensitivity analysis did not suggest that the infant diet explained the associations of the maternal diet with childhood bone mass. Nevertheless, we could not fully exclude that the diet of the child in later life, which may be more strongly related to the maternal diet and bone mass in childhood, has influenced the results. Last, as discussed above, collinearity between protein, calcium, and phosphorus intakes complicated the interpretation of specific effects of individual nutrients. As a result, we were unable to disentangle the independent contributions of these nutrients to childhood bone health.

In conclusion, in a Dutch population-based cohort, we observed that the maternal diet during pregnancy is associated with bone mineral accrual in childhood. In addition to the previously identified role of maternal phosphorus and calcium intakes, this study indicates a role for maternal protein intake as well as homocysteine and vitamin B-12 blood concentrations in childhood bone development. Additional studies are needed to further explore the components of the most-optimal nutrition favoring bone development and to explore

whether nutrition during pregnancy may be a potential target for improving bone health during the life course.

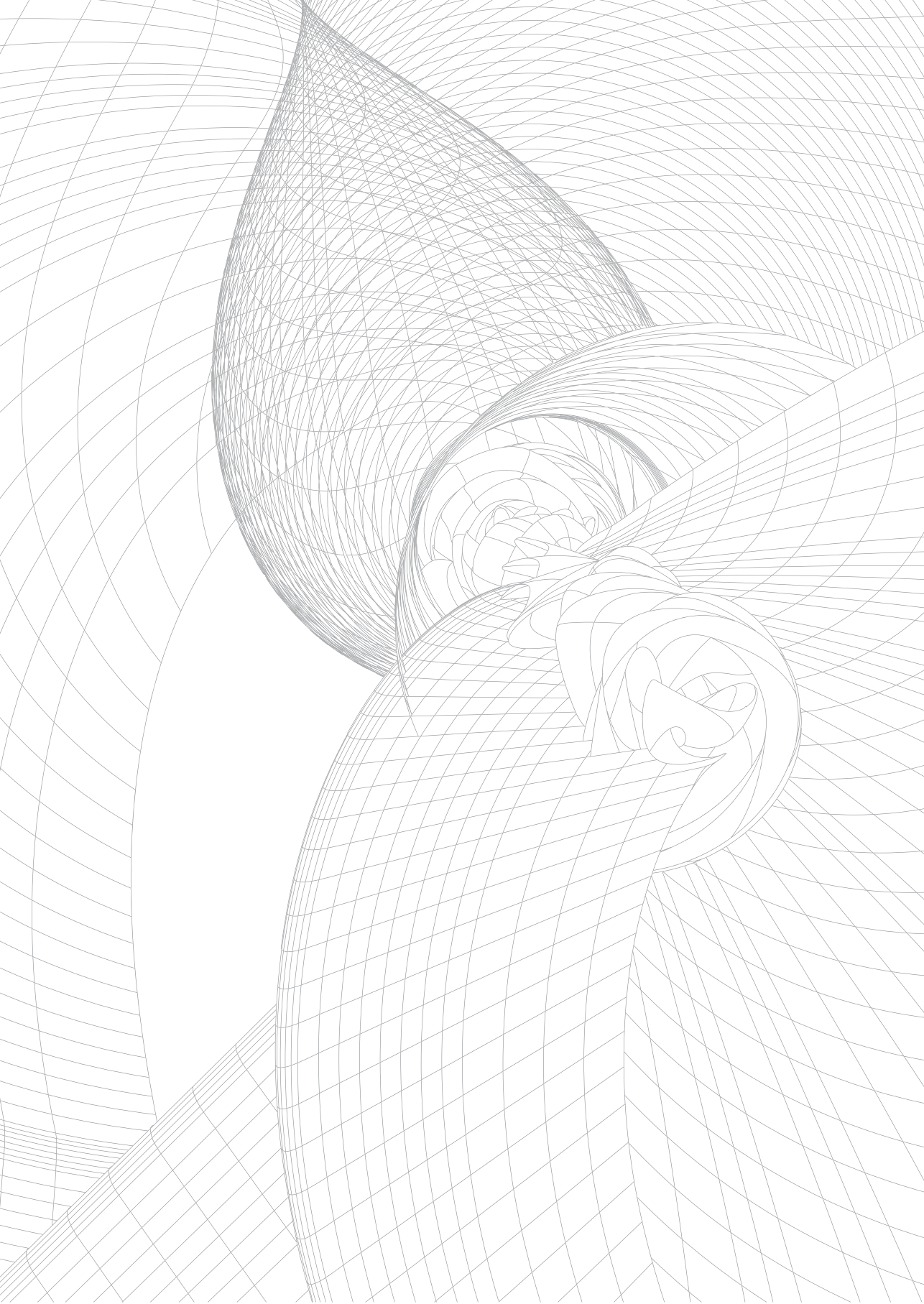
Detailed acknowledgements and online resources can be found in the published article online: http://ajcn.nutrition.org/content/98/1/224.long

REFERENCES

1. Cooper C. Osteoporosis: disease severity and consequent fracture management. **Osteoporos Int** 2010;21 Suppl 2:S425-9.
2. Cooper C, Cole ZA, Holroyd CR, et al. Secular trends in the incidence of hip and other osteoporotic fractures. **Osteoporos Int** 2011;22:1277-88.
3. Harvey N, Dennison E, Cooper C. Osteoporosis: impact on health and economics. **Nat Rev Rheumatol** 2010;6:99-105.
4. Cooper C, Westlake S, Harvey N, Javaid K, Dennison E, Hanson M. Review: developmental origins of osteoporotic fracture**. Osteoporos Int** 2006;17:337-47.
5. Ganpule A, Yajnik CS, Fall CH, et al. Bone mass in Indian children--relationships to maternal nutritional status and diet during pregnancy: the Pune Maternal Nutrition Study. **J Clin Endocrinol Metab** 2006;91:2994-3001.
6. Jones G, Riley MD, Dwyer T. Maternal diet during pregnancy is associated with bone mineral density in children: a longitudinal study. **Eur J Clin Nutr** 2000;54:749-56.
7. Tobias JH, Steer CD, Emmett PM, et al. Bone mass in childhood is related to maternal diet in pregnancy. **Osteoporos Int** 2005;16:1731-41.
8. Yin J, Dwyer T, Riley M, Cochrane J, Jones G. The association between maternal diet during pregnancy and bone mass of the children at age 16. **Eur J Clin Nutr** 2010;64:131-7.
9. Cole ZA, Gale CR, Javaid MK, et al. Maternal dietary patterns during pregnancy and childhood bone mass: a longitudinal study**. J Bone Miner Res** 2009;24:663-8.
10. Prentice A. Diet, nutrition and the prevention of osteoporosis. **Public Health Nutr** 2004;7:227-43.
11. Devlin MJ, Bouxsein ML. Influence of pre- and peri-natal nutrition on skeletal acquisition and maintenance. **Bone** 2012;50:444-51.
12. Goodfellow LR, Cooper C, Harvey NC. Regulation of placental calcium transport and offspring bone health. **Front Endocrinol (Lausanne)** 2011;2:3.
13. Wong IP, Nguyen AD, Khor EC, Enriquez RF, Eisman JA, Sainsbury A, Herzog H, Baldock PA. Neuropeptide Y is a critical modulator of leptin's regulation of cortical bone. **J Bone Miner Res** 2013;28: 886–98.
14. Idelevich A, Sato K, Baron R. What are the effects of leptin on bone and where are they exerted? **J Bone Miner Res** 2013;28:18–21.
15. Holroyd C, Harvey N, Dennison E, Cooper C. Epigenetic influences in the developmental origins of osteoporosis. **Osteoporos Int** 2012;23:401-10.
16. Jacques PF, Selhub J, Bostom AG, Wilson PW, Rosenberg IH. The effect of folic acid fortification on plasma folate and total homocysteine concentrations. **N Engl J Med** 1999;340:1449-54.
17. van Meurs JB, Dhonukshe-Rutten RA, Pluijm SM, et al. Homocysteine levels and the risk of osteoporotic fracture. **N Engl J Med** 2004;350:2033-41.
18. Jaddoe VW, van Duijn CM, van der Heijden AJ, et al. The Generation R Study: design and cohort update 2010. **Eur J Epidemiol** 2010;25:823-41.
19. Klipstein-Grobusch K, den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, Witteman JC. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. **Eur J Clin Nutr** 1998;52:588–96.
20. Statistics Netherlands. Allochtonen in Nederland. [Classification of the population with a foreign background in the Netherlands.] 2001. Available from www.cbs.nl (in Dutch).
21. Statistics Netherlands. Enquêteonderzoek onder allochtonen. [Survey amongst ethnic minorities.] 2005. Available from: www.cbs.nl (in Dutch).
22. Donders-Engelen M vdHL, Hulshof K. Tabel maten en gewichten. 2nd ed. Zeist, Netherlands: University of Wageningen, 2003.
23. Netherlands-Nutrition-Centre. Nevo: Dutch food composition database 2006. The Hague, Netherlands, 2006.

24. Jaddoe VW, Bakker R, van Duijn CM, van der Heijden AJ, Lindemans J, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Uitterlinden AG, et al. The Generation R Study Biobank: a resource for epidemiological studies in children and their parents. **Eur J Epidemiol** 2007;22:917–23.
25. Heaney RP. Bone mineral content, not bone mineral density, is the correct bone measure for growth studies. **Am J Clin Nutr** 2003;78:350-1; author reply 351-2.
26. Prentice A, Parsons TJ, Cole TJ. Uncritical use of bone mineral density in absorptiometry may lead to size-related artifacts in the identification of bone mineral determinants. **Am J Clin Nutr** 1994;60: 837-42.
27. Warner JT, Cowan FJ, Dunstan FD, Evans WD, Webb DK, Gregory JW. Measured and predicted bone mineral content in healthy boys and girls aged 6-18 years: adjustment for body size and puberty. **Acta Paediatr** 1998;87:244-9.
28. Durmus B, van Rossem L, Duijts L, et al. Breast-feeding and growth in children until the age of 3 years: the Generation R Study. **Br J Nutr** 2011;105:1704-11.
29. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. **Am J Clin Nutr** 1993;58:489-96.
30. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies**. Am J Clin Nutr** 1997;65:1220S-1228S; discussion 1229S-1231S.
31. Kleinbaum D, Kupper L, Nizam A, Muller K. Applied regression analysis and other multivariable methods. 4 ed. Belmont, CA: Duxbury Press, 2008.
32. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. **BMJ** 2009;338:b2393.
33. Massey LK. Dietary animal and plant protein and human bone health: a whole foods approach. **J Nutr** 2003;133:862S-865S.
34. Shapiro R, Heaney RP. Co-dependence of calcium and phosphorus for growth and bone development under conditions of varying deficiency. **Bone** 2003;32:532-40.
35. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, et al. CDC growth charts: United States. Adv Data 2000:1-27.
36. Godfrey KM, Sheppard A, Gluckman PD, et al. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. **Diabetes** 2011;60:1528-34.
37. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. **Br J Nutr** 2007;97: 1064-73.
38. Harvey NC, Lillycrop KA, Garratt E, et al. Evaluation of methylation status of the eNOS promoter at birth in relation to childhood bone mineral content. **Calcif Tissue Int** 2012;90:120-7.
39. Palacios C. The role of nutrients in bone health, from A to Z. Crit Rev Food Sci Nutr 2006;46:621-8.
40. Kerstetter JE, O'Brien KO, Caseria DM, Wall DE, Insogna KL. The impact of dietary protein on calcium absorption and kinetic measures of bone turnover in women. **J Clin Endocrinol Metab** 2005;90: 26-31.
41. Dawson-Hughes B, Harris SS, Rasmussen H, Song L, Dallal GE. Effect of dietary protein supplements on calcium excretion in healthy older men and women. **J Clin Endocrinol Metab** 2004;89:1169-73.
42. Carmel R, Lau KH, Baylink DJ, Saxena S, Singer FR. Cobalamin and osteoblast-specific proteins. **N Engl J Med** 1988;319:70-5.
43. Kim GS, Kim CH, Park JY, Lee KU, Park CS. Effects of vitamin B12 on cell proliferation and cellular alkaline phosphatase activity in human bone marrow stromal osteoprogenitor cells and UMR106 osteoblastic cells. **Metabolism** 1996;45:1443-6.
44. Cashman KD. Homocysteine and osteoporotic fracture risk: a potential role for B vitamins. **Nutr Rev** 2005;63:29-36.

45. Ahmadieh H, Arabi A. Vitamins and bone health: beyond calcium and vitamin D. **Nutr Rev** 2011;69: 584-98.
46. McLean RR, Jacques PF, Selhub J, et al. Homocysteine as a predictive factor for hip fracture in older persons. **N Engl J Med** 2004;350:2042-9.
47. Herrmann M, Peter Schmidt J, Umanskaya N, et al. The role of hyperhomocysteinemia as well as folate, vitamin B(6) and B(12) deficiencies in osteoporosis: a systematic review. **Clin Chem Lab Med** 2007;45:1621-32.
48. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? **Epidemiology** 2006;17:413–8.



Chapter 3.2

Fetal and Childhood Growth Patterns Associated with Bone Mass in School-Age Children: The Generation R Study

Denise H.M. Heppe, Carolina Medina-Gomez, Albert Hofman, Fernando Rivadeneira, Vincent W.V. Jaddoe

J Bone Miner Res. 2014 Dec;29(12):2584-93. doi: 10.1002/jbmr.2299.

ABSTRACT

Low birth weight is associated with lower bone accrual in children and peak bone mass in adults. We assessed how different patterns of longitudinal fetal and early childhood growth influence bone properties at school age. In 5431 children participating in a population-based prospective cohort study, we measured fetal growth by ultrasound at 20 and 30 weeks gestation, and childhood growth at birth, 1, 2, 3, and 4 years of age. We analyzed these growth measurements in relation to total body (less head) BMD measured by DXA at age 6. We used conditional growth modeling; a technique which takes into account correlation between repeatedly measured growth measures. Our results showed that estimated fetal weight gain, femur length growth between 20 and 30 weeks of gestation, femur length growth between 30 weeks and birth, as well as all height and weight growth measurements from birth to 4 years of age were all positively associated with BMC, bone area (BA), and BMD (all $P < 0.01$). Fetal femur length growth between 30 weeks and birth was positively associated with BMC and BA (both $P < 0.001$), but not with BMD. Overall, childhood growth measurements exerted a larger influence on bone measures than fetal growth measures. The strongest effect estimate was observed during the first year of life. Children born small (<10th percentile) for gestational age (SGA) had lower BMC and BA, but not BMD, than children born appropriate for gestational age (AGA), whereas children born large (>90th percentile) for gestational age (LGA) had higher BMC and BA (all $P < 0.001$). These differences were no longer present in children showing subsequent accelerated and decelerated infant growth, respectively. We conclude that both fetal and childhood growth patterns are associated with bone mineral accrual, showing the strongest effect estimates in infancy. Compensatory infant growth counteracts the adverse consequences of fetal growth restriction on bone development.

INTRODUCTION

Early life factors influence the development of bone health and osteoporosis during the life-course [1]. Several studies have consistently shown that low birth weight leads to lower bone accrual in children and peak bone mass acquisition in adults [2,3]. However, birth weight is an inappropriate measure of fetal growth, because different adverse fetal growth patterns may still result in the same birth weight [4]. Also, birth weight is strongly correlated with infant growth. A low or high birth weight is frequently compensated for by catch-up growth or catch-down growth during the first 2 years of life [5]. Studies assessing the effects of directly measured fetal growth in different trimesters along with early postnatal growth on bone mineral accrual in later life are scarce. Nevertheless, these studies are important to identify specific early critical periods for bone development. A previous study among 380 children suggested that fetal growth from 19 to 34 weeks of gestation affected childhood bone development at age 4 years [6]. In another study, among the same population, including 628 children, fetal as well as early postnatal growth contributed to bone development at age 4 years [7]. On the other hand, a study among 123 adolescents, found fetal growth and early postnatal growth to be a less crucial determinant of adolescent bone development than prepubertal growth [8]. These findings suggest that bone accrual is influenced by different critical periods, though diverse methodological challenges interfere with identifying effects across time in a conclusive manner; when growth measurements are widely separated in time, pinpointing the most influential period of growth is very difficult [8]. Furthermore, the identification of a critical period of growth on subsequent bone development is challenged by the correlation existing between repeatedly collected growth measures [9], and the unknown influence of growth realignment following an earlier period of growth deviation.

We investigated the independent associations of repeatedly measured fetal and childhood growth characteristics and bone mineral density (BMD) measured by DXA at age 6 years in 5,431 children participating in a population-based birth cohort. We applied conditional growth modeling [10], which enables the simultaneous assessment of correlated growth measures to identify independent critical periods, to further elucidate the independent role of fetal and childhood growth on bone development.

MATERIALS AND METHODS

Study population

This study is embedded in the Generation R Study, a population-based prospective cohort study from fetal life onward in Rotterdam, the Netherlands [11]. All mothers who were resident in the study area and had an expected delivery date between April 2,002 and January

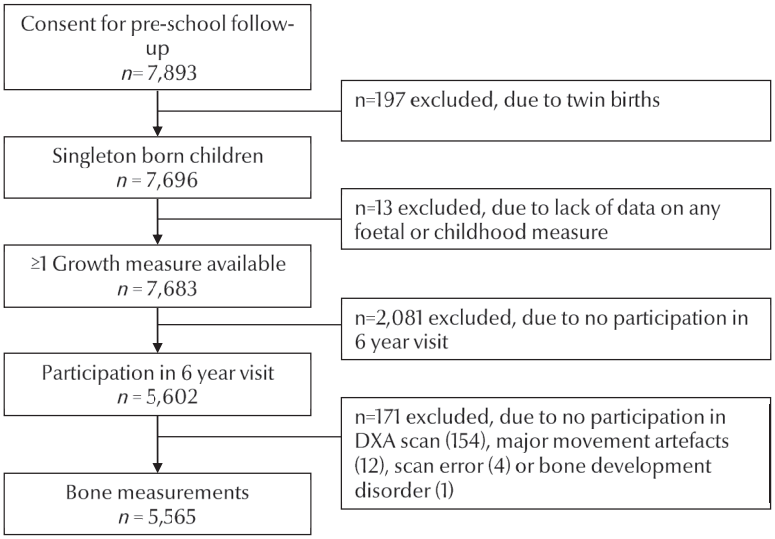


Figure 1. Flowchart of participants included for analysis, from the Generation R Study, Rotterdam, the Netherlands.

2,006 were eligible. The study aimed toward enrollment in the first trimester but allowed enrollment until delivery of the child. In total, 75% of all mothers enrolled before 18 weeks of gestation. Of all eligible children in the study area, 61% participated at birth in the study. The study was approved by the Medical Ethics Committee of the Erasmus Medical Centre, Rotterdam (MEC 198.782/2001/31), and conducted according to the guidelines of the Helsinki Declaration. Written informed consent was obtained from all participants.

In total, 7,893 children of mothers who gave consent for follow-up in the preschool phase (0 to 4 years) were eligible for this study. Of the 7,696 singleton born children, growth was measured at least once in 7,683 children. Of these children, 5,602 visited our research center around the age of 6 years. DXA scanning was successfully performed in 5,431 children (69% of the eligible population). A flowchart of included participants is shown in Figure 1.

FETAL AND CHILDHOOD GROWTH CHARACTERISTICS

Fetal ultrasound examinations were performed in each trimester of pregnancy. Medians (interquartile ranges [IQRs]) of these visits were 13.1 (2.4), 20.5 (1.3), and 30.4 (1.1) weeks of gestation for the first, second, and third trimester, respectively. In total, 88% of the examinations took place at either of the two research centers of the study. The remaining examinations were carried out in one of five hospitals in the vicinity under guidance of our research staff. In order to achieve optimal reproducibility all sonographers were experi-

enced and underwent additional training according to guidelines from The Fetal Medicine Foundation [12]. Gestational age was determined at the first fetal ultrasound examination, because 39% of pregnant women had unknown or irregular last menstrual periods and because using last menstrual period for this purpose has several described limitations [12]. In the second and third trimesters, fetal head circumference (HC), abdominal circumference (AC), and femur length (FL) were measured to the nearest millimeter using standardized ultrasound procedures [13-16]. A brief description of the techniques applied is available in the Supporting Information. Fetal weight (EFW) was estimated using HC, AC, and FL in the formula from Hadlock: log10 EFW = 1.5662 – 0.0108 (HC) + 0.0468 (AC) + 0.171 (FL) + 0.00034 (HC)2 – 0.003685 (AC * FL) [17]. In a previous study within the Generation R Study, reference curves were developed based on fetal growth characteristics of the whole study population [12]. In the current study, we used these reference curves to calculate gestational age–adjusted SD scores.

Information about offspring sex, gestational age, and weight at birth was obtained from medical records and hospital registries. Very preterm birth was defined as birth occurring before 32.0 weeks of gestation, and preterm birth as birth between 32.0 and 37.0 weeks of

Table 1. Parental, fetal and child characteristics, the Generation R study, Rotterdam, the Netherlands.

Characteristic	Boys	Girls	*P*-value[1]
	n=2,718	*n*=2,732	
Maternal characteristics			
Age (years)	30.9 (5.1)	30.8 (5.0)	0.38
Height (cm)	167.9 (7.2)	167.8 (7.5)	0.62
Pre-pregnancy weight (kg)	66.2 (12.1)	66.8 (12.5)	0.19
Pre-pregnancy BMI (kg/m²)	23.4 (4.1)	23.6 (4.2)	0.13
Parity ≥1			
No	1466 (54)	1499 (55)	0.39
Yes	1150 (42)	1129 (42)	
Missing	106 (4)	90 (3)	
Single motherhood			
No	2192 (81)	2203 (81)	0.8
Yes	285 (11)	284 (10)	
Missing	245 (9)	231 (9)	
Educational status			
Primary	199 (7)	214 (8)	0.76
Secondary	1021 (38)	1038 (38)	
Higher	1259 (46)	1234 (45)	
Missing	243 (9)	232 (9)	

Table 1 (continued)

Characteristic	Boys	Girls	*P*-value[1]
	n=2,718	*n*=2,732	
Smoking during pregnancy			
Never	1766 (65)	1797 (66)	0.08
Until pregnancy was known	199 (7)	233 (9)	
Continued	382 (14)	333 (12)	
Missing	375 (14)	355 (13)	
Alcohol use during pregnancy			
No	1200 (44)	1226 (45)	
Yes	930 (34)	952 (35)	0.23
Missing	592 (22)	1045 (36)	
Start folic acid supplement use			
Preconception	812 (30)	886 (33)	0.18
First 10 weeks	588 (22)	571 (21)	
No	439 (16)	418 (15)	
Missing	883 (32)	843 (31)	
Paternal characteristics			
Age (years)	33.4 (5.4)	33.5 (5.5)	0.44
Height (cm)	182.4 (7.9)	182.4 (7.8)	0.39
Weight (kg)	83.8 (12.8)	84.2 (12.9)	0.95
Body mass index (kg/m^2)	25.2 (3.3)	25.3 (3.4)	0.26
Child characteristics			
Gestational age at birth (weeks)	39.9 (1.7)	39.8 (1.7)	0.17
Ethnicity			
European	2026 (74)	2022 (74)	0.89
African	413 (15)	403 (15)	
Asian	146 (5)	144 (5)	
Missing	137 (5)	149 (6)	
Breast feeding			
Never	186 (7)	188 (7)	0.64
>0-3 months	655 (23)	656 (23)	
>3 months	1021 (38)	1055 (39)	
Missing	819 (30)	777 (29)	
Participation in sports at age 6			
Never	1291 (47)	1226 (45)	<0.001
1/week	739 (27)	976 (36)	
≥2/week	289 (11)	88 (3)	
Missing	403 (15)	428 (16)	

[1]*P*-values obtained by Students *t*-tests for continuous variables and chi-square tests for categorical variables.

gestation. Small for gestational age (SGA) and large for gestational age (LGA) was defined as sex- and gestational age–adjusted birth weight below the 10th percentile and above the 90th percentile, respectively. Childhood growth was routinely measured at the Community Health Centres at the median ages of 6.2 (IQR 0.4), 11.1 (IQR 0.7), 24.8 (IQR 1.6), 36.7 (IQR 1.4), and 45.8 (IQR 1.3) months following standardized protocols. Sex- and age-adjusted SD scores were calculated using Growth Analyser 3.5 (Dutch Growth Research Foundation, Rotterdam, the Netherlands; http://www.growthanalyser.org) [18,19]. In accordance with earlier studies, we defined an increase or decrease in weight greater than 0.67 SD from birth to the age of 24 months as accelerated or decelerated growth, respectively[20,21] At the age of 6 years, we measured weight in our research center using an electronic personal scale (Seca, Almere, The Netherlands) and height using a Harpenden stadiometer (Holtain Limited, Dyfed, UK) following standardized protocols.

BMD measurements

Total body BMD (g/cm^2), bone mineral content (BMC; g), and bone area (BA; cm^2) were measured at a median age of 6.0 (IQR 0.37) years using a DXA scanner (iDXA; GE Lunar, Madison, WI, USA). As described in detail earlier [22], well-trained research assistants obtained the DXA scans using the same device and software (enCORE) following standard manufacturer protocols. In our analyses, we used areal total body less head (TBLH) BMD, BMC, and BA as recommended by the International Society for Clinical Densitometry for pediatric evaluations of bone health [23] All measures were adjusted for skeletal size by using body height or weight as covariate in the models to correct for artifacts arising from periods of rapid growth [24]; this is needed because areal BMD measured on larger bones overestimates true (volumetric) BMD, whereas on smaller bones it can underestimate BMD across individuals [25, 26]. In subsequent comparative analyses, we, in addition to the other covariates, corrected BMC for BA to further adjust for size effects [25].

Covariates

We registered maternal age at enrollment and collected information about maternal education, marital status, parity, and country of birth, and country of birth of the father and grandparents by questionnaire at enrollment in the study. Maternal smoking and alcohol habits were assessed in each trimester. We measured parental height and weight at the research center and obtained information about maternal weight before pregnancy by questionnaire. Because the enrollment in our study was during pregnancy, we were not able to measure maternal weight before pregnancy; however, correlation of pre-pregnancy weight obtained by questionnaire and weight measured at enrollment was high ($P = 0.95$; $P < 0.01$). We categorized ethnicity into three main groups: Western (Dutch, Turkish, other European, American, and Oceanic), African (Moroccan, other African, Antillean, Surinamese-Creole, and Cape Verdean), and Asian (Indonesian, other Asian, and Surinamese-Hindu) descent

according to the three largest transcontinental ancestral groups. Information about breast feeding [27] and participation in sports was obtained from postnatal questionnaires.

Statistical analysis

We used t tests and chi-square tests to compare differences in subject characteristics between boys and girls. We calculated age-adjusted SD scores for all bone measures based on their distribution in the whole study population and analyzed them following four strategies. First, we performed multiple linear regression analyses to assess the individual associations of fetal and childhood growth measures with bone measures at the age of 6 years. Second, we assessed the associations of these growth measures with bone measures using conditional change modeling [10]. In conditional growth modeling, a growth measure at a specific time point is adjusted for growth predicted by prior growth measures. Accordingly, we calculated standardized residuals by regression of the growth measure of interest on prior growth measures [10], obtaining growth measures independent of prior growth measures and statistically independent of each other across time. This approach enabled a simultaneous analysis of all growth measures with bone measures in order to identify the period of growth most critical to bone development. In an attempt to eliminate potential artifacts caused by bone size and to further distinguish potential effects on bone size from bone mineral accrual, we additionally corrected BMC for BA in a sensitivity analysis. Third, we assessed associations of birth outcomes (gestational age, birth weight, gestational age–adjusted birth weight) with bone measures at the age of 6 years. Fourth, we explored the associations of gestational age–adjusted birth weight with bone measures stratified for the postnatal growth pattern. Based on previous literature, all models were adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, smoking, use of folic acid supplements, paternal weight and height, and child's sex, ethnicity, breastfeeding duration, and participation in sports [28-31]. Models concerning weight measures were additionally adjusted for current height, whereas models concerning height were adjusted for current soft tissue weight calculated as "lean + fat mass" (thereby excluding the contribution of bone mass to the child's weight). Because missing values add up in conditional modeling and to prevent bias associated with missing data, we used multiple imputations (five imputations) to impute missing values in growth measures and covariates. Missing values for growth measures and covariates were imputed based on the correlation of the missing variables with other participant characteristics and other available growth measures, according to the Markov Chain Monte Carlo method [32]. The percentage of missing values for any fetal growth measure was lower than 16%, and for any childhood growth measure it was lower than 38%. Of all children, 6% did not have any data on growth from 1 to 4 years of age. Results from the complete case analyses were similar to results from the imputed analyses. We only present results for the imputed analyses. Comparing infants born SGA and not born SGA, we would be able to detect statistically significant difference in childhood BMD

of 0.13 SD (type I error of 5% and a type II error of 20% [power 80%]) (33). Analyses were performed using the SPSS Predictive Analytic Software version 17.0 for Windows (PASW Inc, Chicago, IL, USA).

Fetal growth, childhood growth, and bone measures

The associations of individually modeled growth measures with bone parameters at age 6 are shown in Online Resource Table S1. In short, all fetal length and weight measures were positively associated with BMC and BA (all $P < 0.05$), whereas all childhood growth height and weight measures were positively associated with BMC, BA, and BMD (all $P < 0.01$). The magnitude of the effect estimates increased with advancing age. When applying a conditional model, estimated fetal weight gain between 20 and 30 weeks gestation, 30 weeks and birth, as well as childhood weight gain from birth to 4 years of age were all positively associated with BMC, BA, and BMD (all $P < 0.001$) (Figure 2). Fetal femur length growth between 20 and 30 weeks gestation and height growth from birth to 4 years were all positively associated with BMC, BA, and BMD (all $P < 0.01$). Fetal femur length growth between 30 weeks gestation and birth was positively associated with BMC and BA (both $P < 0.001$), but not with BMD. Effect estimates (in SD) for childhood height and weight growth measures were larger than for fetal growth. The largest effect estimates were found for the associations of height growth during the first year with BMC and BA, and weight gain during the first year with BMC, BA, and BMD. The size of the effect estimates decreased after the first year of age, except for the association of height growth with BMD, which peaked at 2 to 3 years of age. The corresponding effect estimates are shown in Online Resource Table S2.

We further explored whether the associations of height and weight growth during the first year with bone measures were driven by growth during the first 6 months by replacing growth from birth to age 1 year by two separate measures for growth from birth to 6 months and from 6 to 12 months in our models. Height and weight growth during the first 6 months showed stronger associations with bone measures than growth from 6 to 12 months, yet effect estimates were not larger than those observed for growth during the first year as a whole (data not shown).

Size is a major determinant of bone mass. To demonstrate its impact on the conditional growth analysis, Online Resource Table S3 shows the results from the analyses unadjusted for size. Size adjustment reduced effect sizes approximately by one third. In a second sensitivity analysis, to further distinguish an increase in bone mineral accrual from bone size, we additionally adjusted BMC models for BA. As a result, fetal femur growth measures were no longer associated with BMC. However, estimated fetal weight measures, and postnatal height and weight growth measures remained positively associated with BMC, although effect estimates were less than half the size of the effect estimates for BMC not adjusted for BA (shown in Online Resource Table S4).

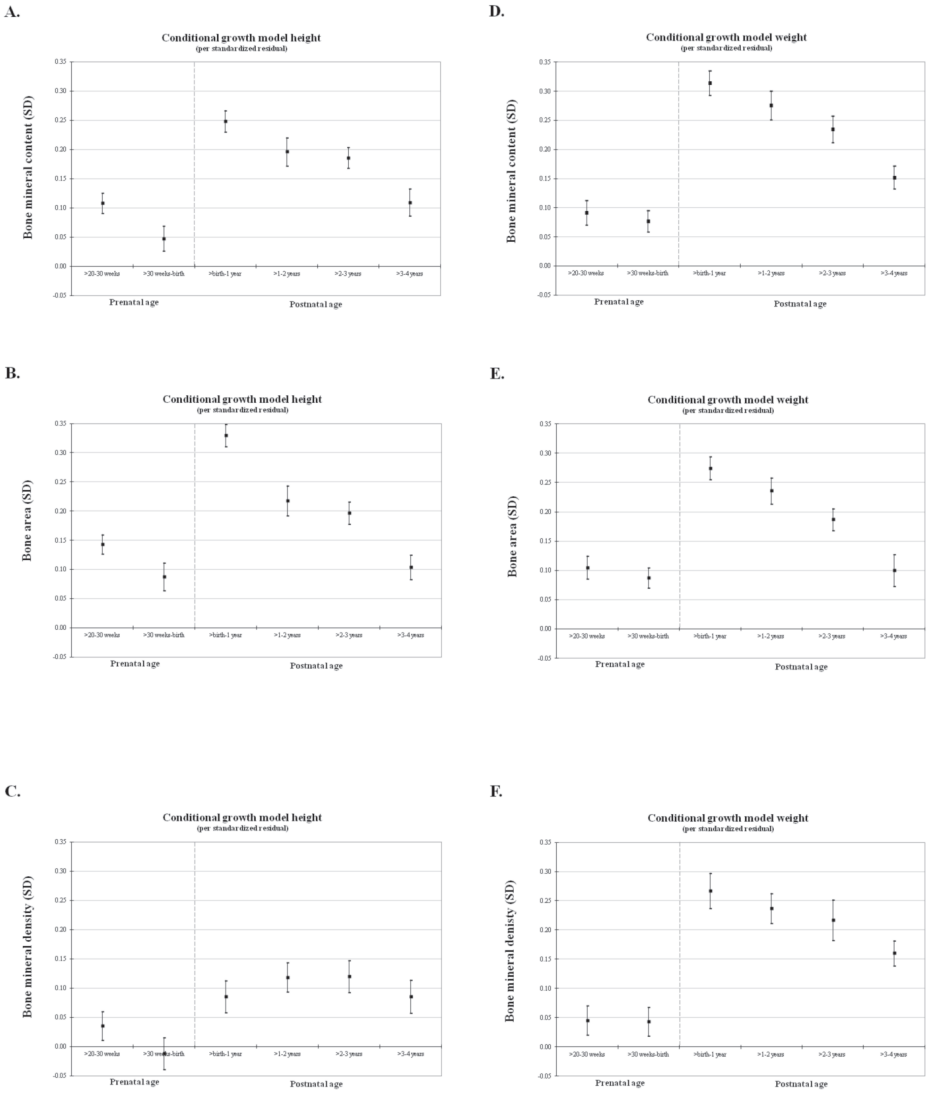


Figure 2. (A–F) Associations of conditional fetal and childhood growth with bone measures at age 6 years, from the Generation R Study Cohort, Rotterdam, The Netherlands. Values are based on multiple linear regression models and reflect the coefficients and 95% CI per standardized residual of conditionally modeled growth. Conditional growth variables are independent of prior growth. Models are adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, smoking, daily protein intake, use of folic acid supplements, paternal weight and height, and sex, ethnicity, breastfeeding duration, participation in sports, and for current height (weight models) or weight measured as "lean+fat mass" (height models) of the child and mutually for the other growth measures.

Table 2. Average fetal growth, childhood growth and bone mass measures at age 6, the Generation R study, Rotterdam, the Netherlands.

Period	(Gestational) age median (IQR)	Growth character-istic	Unit	Boys		Girls		*P*-value[1]
				n	mean (SD)	*n*	mean (SD)	
Trimes-ter 2	20.5 (1.3) weeks	Femur length	mm	2292	33.4 (3.5)	2316	33.4 (3.4)	0.72
		Estimated fetal weight	g	2281	385 (94)	2309	376 (87)	<0.001
Trimes-ter 3	30.4 (1.1) weeks	Femur length	mm	2368	57.3 (3.0)	2394	57.6 (3.0)	0.007
		Estimated fetal weight	g	2364	1635 (258)	2383	1619 (260)	0.03
Birth	40.1 (1.9) weeks	Birth length	cm	1698	50.6 (2.4)	1703	49.9 (2.2)	<0.001
		Birth weight	g	2719	3503 (562)	2716	3372 (527)	<0.001
1 year	11.1 (0.7) months	Height	cm	2141	75.1 (2.5)	2116	73.5 (2.5)	<0.001
		Weight	kg	2143	10.0 (1.1)	2123	9.3 (1.0)	<0.001
2 years	24.8 (1.6) months	Height	cm	1987	88.9 (3.4)	1984	87.7 (3.4)	<0.001
		Weight	kg	2023	13.2 (1.5)	2006	12.7 (1.5)	<0.001
3 years	36.7 (1.4) months	Height	cm	1884	97.9 (3.8)	1911	96.8 (3.8)	<0.001
		Weight	kg	1908	15.5 (1.8)	1929	15.0 (1.9)	<0.001
4 years	45.8 (1.3) months	Height	cm	1687	103.7 (4.1)	1666	102.8 (4.2)	<0.001
		Weight	kg	1697	17.1 (2.2)	1669	16.8 (2.3)	<0.001
6 years	72.2 (4.1) months	Height	cm	2719	119.5 (5.7)	2713	118.7 (5.7)	<0.001
		Weight	kg	2719	23.2 (3.8)	2713	22.9 (4.1)	0.008
		BMC	g	2722	523 (100)	2718	519 (98)	0.01
		Bone area	cm^2	2722	942 (115)	2718	941 (109)	0.17
		BMD	g/cm^2	2722	0.552 (0.051)	2718	0.549 (0.052)	0.66

[1]*P*-values obtained by Students *t*-tests for continuous variables and chi-square tests for categorical variables. *SD* denotes standard deviation. BMC= Bone Mineral Content, BMD= Bone Mineral Density

Birth outcomes and bone measures

Gestational age at birth showed a weak positive association with BMC (P for trend 0.05) and BA (P for trend 0.02) at 6 years of age, not with BMD (Table 3). Children born preterm had a −0.09 SD lower BMC (95% CI, −0.19, 0.00) and a −0.08 SD lower BA (95% CI, −0.18, 0.01) at school age. Birth weight showed a stronger positive association with both BMC and BA (both P for trend <0.001) and a weak positive association with BMD (P for trend 0.06). When birth weight was adjusted for gestational age at birth, it was still positively associated with BMC and BA, but not with BMD. As compared to children born AGA, children born SGA had a −0.07 SD (95% CI, −0.14, 0.00) lower BMC and a −0.11 SD (95% CI, −0.18, −0.05) lower BA, whereas children born LGA had a 0.12 SD (95% CI, 0.06, 0.18) higher BMC and a 0.16 SD (95% CI, 0.10, 0.23) higher BA.

Table 3. Associations of birth outcomes with bone mass at age 6, the Generation R Study Cohort, Rotterdam, The Netherlands. Values are based on multiple linear regression models and reflect the coefficients and 95% CI for each category. Models are adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, smoking, use of folic acid supplements, paternal weight and height, and gender, ethnicity, breastfeeding duration, participation in sports and current height of the child. SGA denotes small for gestational age, AGA denotes appropriate for gestational age, LGA denotes large for gestational age.

Fetal growth	*n*	Bone mineral content (SD)		Bone area (SD)		Bone mineral density (SD)	
		β	95% CI	β	95% CI	β	95% CI
Gestational age adjusted for birth weight (weeks)							
<32	28	−0.09	−0.35, 0.17	−0.06	−0.31, 0.19	−0.08	−0.40, 0.23
≥32–37	225	−0.09	−0.19, 0.00	−0.08	−0.18, 0.01	−0.06	−0.18, 0.05
≥37–42	**4761**	**0**		**0**		**0**	
≥42	380	0.01	−0.06, 0.09	0.01	−0.06, 0.08	0.01	−0.08, 0.10
P for trend		0.05		0.03		0.35	
Birth weight (g)							
<2000	64	−0.11	−0.29, 0.07	−0.06	−0.23, 0.11	−0.12	−0.34, 0.09
≥2000–2500	170	−0.10	−0.21, 0.01	−0.15	−0.26, −0.05[a]	−0.01	−0.15, 0.12
≥2500–3000	808	−0.08	−0.14, −0.02[b]	−0.07	−0.13, −0.01[b]	−0.07	−0.15, 0.00[a]
≥3000–3500	**1912**	**0**		**0**		**0**	
≥3500–4000	1755	0.04	−0.01, 0.09	0.1	0.05, 0.14[c]	−0.03	−0.09, 0.02
≥4000–4500	602	0.11	0.04, 0.18b	0.15	0.08, 0.21[c]	0.05	−0.04, 0.13
≥4500	124	0.19	0.06, 0.32b	0.29	0.16, 0.41c	−0.02	−0.18, 0.14
P for trend		<0.001		<0.001		0.06	
Birth weight adjusted for gestational age (SD)							
SGA	501	−0.07	−0.14, 0.00[a]	−0.11	−0.18, −0.05[c]	−0.01	−0.09, 0.07
AGA	**4311**	**0**		**0**		**0**	
LGA	576	0.12	0.06, 0.18[c]	0.16	0.10, 0.23[c]	0.02	−0.05, 0.10

[a]*P* value <0.05 [b]*P* value <0.01 [c]*P* value <0.001

Birth weight, infant growth, and bone measures

As compared to children born AGA with normal infant growth, children born SGA without growth realignment between 0 and 2 years of age had a −0.30 SD (95% CI, −0.42, −0.18) lower BMC, a −0.35 SD (95% CI, −0.47, −0.24) lower BA, and a −0.21 SD (95% CI, −0.36, −0.06) lower BMD at age 6 years (Fig. 3A–C). Children born LGA without growth realignment during infancy had a 0.44 SD (95% CI, 0.33, 0.55) higher BMC, a 0.44 SD (95% CI, 0.34, 0.55) higher BA, and a 0.28 SD (95% CI, 0.14, 0.41) higher BMD at age 6 years than children born AGA with normal infant growth. Children born SGA and LGA who did show growth realignment during infancy had a similar BMD, BMC, and BA to that of children born AGA with normal growth. The corresponding effect estimates are shown in Online Resource Table S5.

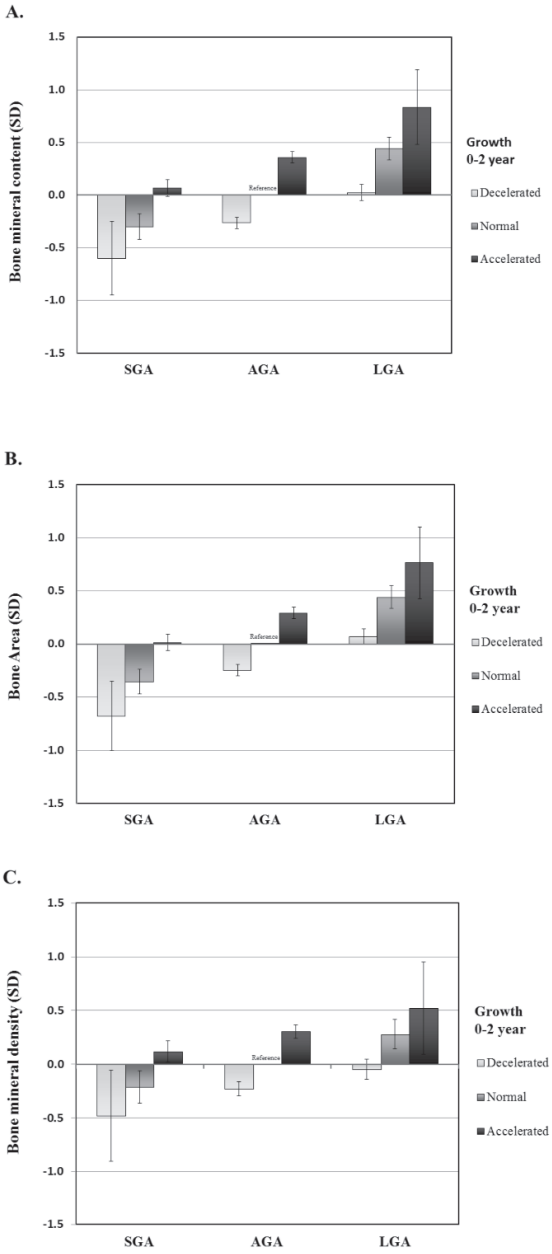


Figure 3. (A–C) Associations of birth weight with bone measures at age 6 years, stratified for postnatal growth patterns, from the Generation R Study Cohort, Rotterdam, The Netherlands. Values are based on multiple linear regression models and the bars and lines reflect the coefficients and 95% CI for each category of birth weight and postnatal growth pattern. Models are adjusted for maternal age, weight, height, parity, educational level, marital status, alcohol use, smoking, use of folic acid supplements, paternal weight and height, and sex, ethnicity, breastfeeding duration, participation in sports, and current height of the child. SGA = small for gestational age; AGA = appropriate for gestational age; LGA = large for gestational age.

DISCUSSION

Main findings

In this large population-based prospective cohort study of pregnant women and their children in the Netherlands, we found that both fetal and childhood growth, as reflected by height and weight gain, were positively associated with bone accrual at school age. Childhood growth showed larger effect estimates than fetal growth, whereas growth occurring in the first year of life showed the strongest positive association with bone mass accrual in later childhood. Gestational duration and birth weight were positively associated with bone parameters at 6 years of age. However, growth realignment between birth and 2 years of age in children born SGA or LGA led to similar bone measures at the age of 6 years to children born AGA who showed normal postnatal growth.

Methodological considerations

One of the major strengths of the study is that it is a large-scale, population-based, prospective cohort study rich in assessments of prenatal and early childhood growth. This unique setting of repeated measures of fetal growth, childhood anthropometrics, and bone measures enabled us to evaluate the independent associations of early growth with bone health. We used DXA, a well-validated technique to assess bone mass accrual in children. Nevertheless, our study is not free of limitations. Of the eligible children, 69% participated in the 6-year visit at our research center. Children who did not participate grew slower during fetal life and, accordingly, had lower birth weight and length than those participating in the study. They more often showed accelerated growth in the first 2 years of life, were of non-European descent, and had mothers with a lower educational level (all $P < 0.05$). Further, because in the Netherlands it is not obligatory to attend the routine Community Health Centre visits, only 40% of the children had complete data on growth from 1 to 4 years of age and 6% did not have data from these routine visits at all. Children with incomplete data on average had a higher weight and BMC, and larger BA at the age of 6 years (all $P < 0.05$), but overall similar BMD levels. Although we used multiple imputation, we cannot exclude that missing information may still have led to loss of power or biased estimates. Nevertheless, this will only be the case if effect estimates would have differed systematically between those children included and not included in the analyses. This is unlikely, because loss to follow-up in our cohort is not expected to be related to the studied research question [34], but cannot be fully excluded.

Establishing gestational age by ultrasound is considered superior to the use of the last menstrual period (LMP)[35] because almost 40% of pregnant women have unknown or irregular last menstrual periods [12]. However, use of first trimester ultrasounds assumes the variation in fetal growth before that ultrasound to be zero, possibly leading to underestimated effect

estimates in early pregnancy. We minimized this unwanted side effect by using crown-rump length and biparietal diameter for pregnancy dating [36,37], but not for assessing fetal growth. Yet as a result of the correlation between fetal growth measures, underestimation of effect estimates may still have occurred. Nonetheless, because we studied relative change in size within time periods by conditional modeling, we do not expect our pregnancy dating strategy to have substantially influenced our results. The validity of ultrasound estimation of fetal weight has often been debated. A systematic review assessed the validity of estimated fetal weight measurements by reviewing measurement errors across 42 studies [38]. In this review, the authors found that all methods used to estimate fetal weight, including the method of Hadlock, have insignificant systematic error. Random error, on the other hand, averaged 10%. Another limitation possibly leading to random error may be the fact that growth measures from age 1 to 4 years were acquired from routine Community Health Centres. Nevertheless, measurements within these clinics were performed using standardized protocols. Even though routinely collected measurements of this type have previously been shown to have good accuracy by lacking systematic error [39], random error may still have been introduced. Random error may reduce power and lead to underestimation of effects. Furthermore, the 0.67 SD cut-off that we used to define "accelerated growth" has been internationally recognized to represent clinically significant catch-up growth, but does not necessarily represent, and should not be interpreted as a biological phenomenon. Last, although we collected detailed information on many potential confounding variables, residual confounding due to unmeasured sociodemographic and lifestyle factors could still be influencing the results.

Interpretation of main results

Our results confirmed that early growth is associated with both bone size (BA) and bone mineral (BMC) accrual at school age. Overall, effect estimates for models including BMD were consistent in direction, yet smaller in magnitude. This attenuation of the effect on BMD is likely consequence of the influence of changes in bone area on BMD (eg, larger bone areas result in comparatively lower BMD). Further, as DXA is a two-dimensional assessment of a three-dimensional structure, the overestimation of areal BMD in larger bones compared to volumetric BMD may also, although to a lesser extent, reduce the effect estimates. However, the prominent, yet incomplete, attenuation of effect estimates by additional correction of BMC for BA supports the idea that early growth is associated with both increased bone size as well as mineral accrual. The attenuation of effect estimates by size correction seemed to be larger for height than for weight measures. Possibly, height growth, and in particular increase in fetal femur length, are more closely related to the actual skeletal frame size, whereas weight gain is only indicative of the loading effects on the skeleton during postnatal life.

Evidence supporting an influence of early growth on adult peak bone mass acquisition is increasing [2]. Recently, fast weight and height gain during childhood and adolescence were positively associated with bone strength among 1,658 adults 60 to 64 years old [40]. Only a few studies have assessed the association of fetal growth with bone development by actually measuring fetal growth instead of using birth weight as a proxy for fetal growth [6-8, 41]. In line with our results, Beltrand and colleagues [41] found that fetal growth restriction (≥20 percentile reduction in estimated fetal weight between 22 weeks of gestation and birth) led to lower BMC in 185 newborns, independent of birth weight. Among 380 British children, fetal femur length and abdominal circumference growth during 19 to 34 weeks of gestation were positively associated with BMC and BMD at age 6 years. The effect of fetal abdominal growth was independent of current height, weight, or bone size [6]. Among 119 adolescents living in Denmark, third trimester fetal growth velocity, birth weight, and growth in the first year were positively associated with BMC [8]. However, these associations fully disappeared when adjusted for current height and weight. For that reason, the authors concluded that growth in later life, rather than early growth, may be crucial to bone health in adolescence. However, pinpointing the most influential period of growth is less precise when assessments are so widely separated in time.

Assessment of repeatedly measured growth is challenged by some methodological issues. In fact, "early size" adjusted for "later size" in regression analysis is a measure of change in size between the earlier and later measurement, rather than a measure of absolute growth [9]. To overcome these issues, only one previous study used conditional modeling [10] to study the associations of linear and abdominal growth measured at 11, 19, and 34 weeks of gestation, birth, and 1, 2, 3, and 4 years with bone measures among 628 4-year-olds [7]. The results were consistent with our observations, showing that both fetal and childhood growth are positively associated with bone development at age 4 years, whereas growth in the first 2 postnatal years contributed most strongly. These results are also in line with our results indicating that children born with a low birth weight who showed growth realignment in the first 2 years had similar bone mass to children with a normal birth weight and normal postnatal growth. Our findings highlight the importance of early growth patterns determining bone health later in childhood.

Potential underlying mechanisms

Bone growth during fetal development and postnatal life involves complex regulatory processes mediated by growth factors, cytokines, hormones, mechanical stimuli, and diverse environmental influences. These processes are largely controlled by genetics, epigenetic regulation, and availability of nutrients and diverse exposures during fetal life, childhood, and adolescence [42]. By adaptation to environmental cues, early growth even from fetal life may already program later bone development [43]. Among others, hormones like leptin,

growth hormone (GH), and cortisol have been suggested to play a prominent role in this "programming" of bone mineral accrual [44,45]. However, the exact mechanisms underlying the process remain unclear. Altered leptin levels, resulting from low or high nutrient availability, are proposed to program bone development by stimulating differentiation of mesenchymal stem cells into bone (osteoblasts) precursors, over adipogenic lineages, as well as stimulating cortical bone over trabecular bone formation [44,46]. Further, leptin levels are negatively correlated with fetal growth retardation [47] and positively with postnatal catch-up growth and neonatal BA and BMC [48,49]. Similarly, the GH/IGF-I axis has long been considered a major determinant of bone mass acquisition. The axis is negatively affected by fetal growth restriction [42] and essential to achieve catch-up growth in fetal growth–retarded infants [50]. IGF-I levels in neonates as well as in children are positively correlated with bone mass [51,52]. On the other hand, endogenous cortisol inhibits osteoblast function [53]. Serum cortisol levels are higher in infants born SGA [54], especially in those who do not achieve catch-up growth [55]. High normal endogenous cortisol levels have been negatively associated with bone mass, predominantly in boys [56,57]. Nutritional aspects such as breastfeeding, calcium and vitamin D intake, and environmental exposures such as sunlight and physical activity may also exert an effect on these associations. Nevertheless, the fact that effect estimates remained essentially unchanged upon correction for a large number of nutritional and environmental factors does not seem to corroborate this contention.

Conclusion

Both fetal and childhood growth predict bone development at 6 years of age. Weight and height growth in the first year of life appeared to have the largest impact on bone mineral accrual. Compensatory growth in the first 2 postnatal years reduced the adverse consequences of slower growth velocity in fetal life on childhood bone mass. Because childhood bone mass tends to track into adulthood, fetal life and infancy may be critical periods to attain optimal bone health and possibly reduce the risk of osteoporosis in later life. The mechanisms underlying these findings are largely unknown and warrant further study.

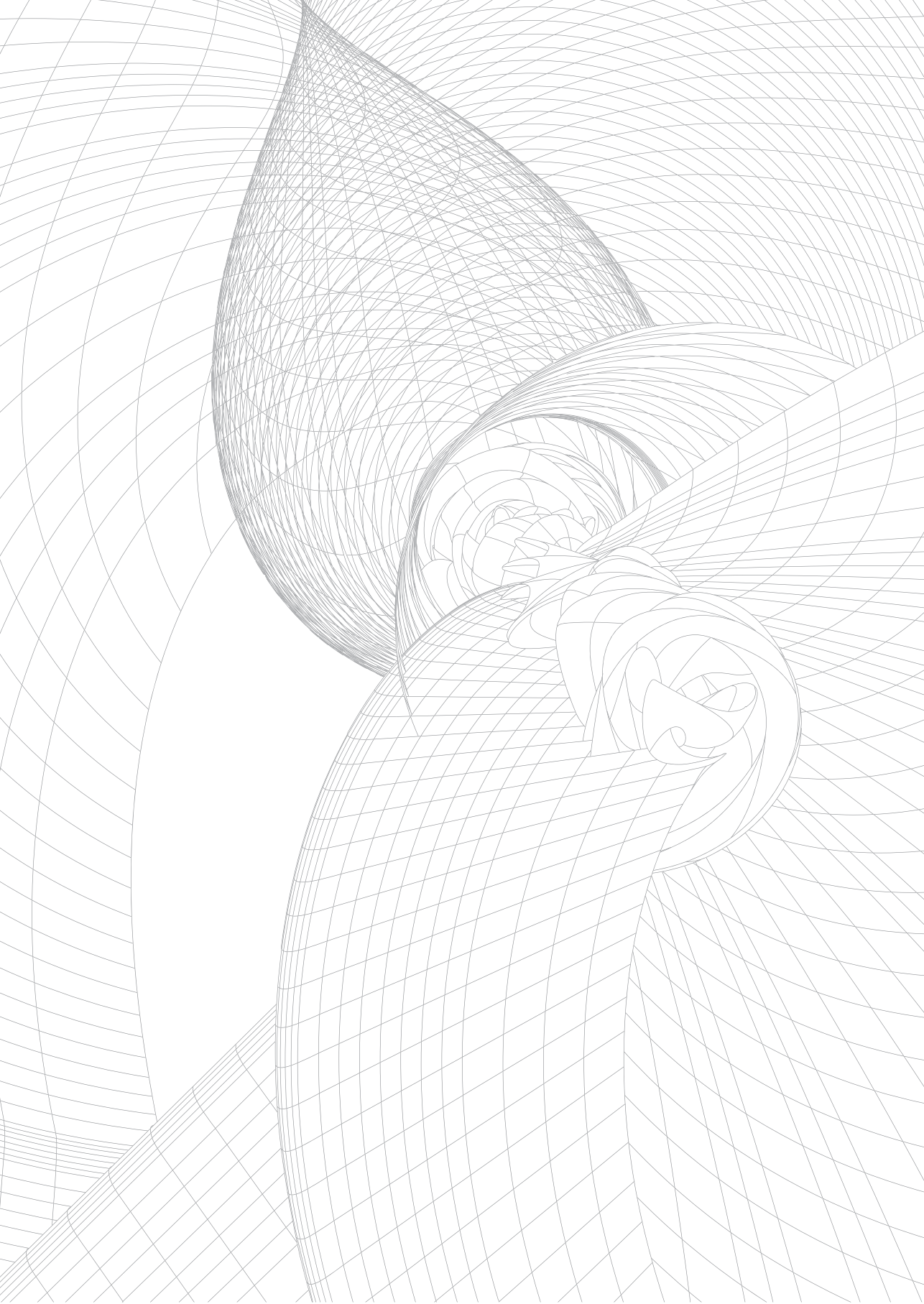
Detailed acknowledgements and online resources can be found in the published article online: http://onlinelibrary.wiley.com/doi/10.1002/jbmr.2299/full

REFERENCES

1. Cooper C, Westlake S, Harvey N, Javaid K, Dennison E, Hanson M. Review: developmental origins of osteoporotic fracture. **Osteoporos Int**. 2006; 17(3):337–47.
2. Baird J, Kurshid MA, Kim M, Harvey N, Dennison E, Cooper C. Does birthweight predict bone mass in adulthood? A systematic review and meta-analysis. **Osteoporos Int**. 2011; 22(5):1323–34.
3. Martínez-Mesa J, Restrepo-Méndez MC, González DA, et al. Life-course evidence of birth weight effects on bone mass: systematic review and meta-analysis. **Osteoporos Int**. 2013 Jan; 24(1):7–18.
4. Wilcox AJ. Intrauterine growth retardation: beyond birthweight criteria. **Early Hum Dev**. 1983; 8(3–4): 189–93.
5. Taal HR, Vd Heijden AJ, Steegers EA, Hofman A, Jaddoe VW. Small and large size for gestational age at birth, infant growth, and childhood overweight. **Obesity** (Silver Spring). 2013; 21(6):1261–8.
6. Harvey N, Mahon P, Robinson S, et al. SWS Study Group. Different indices of fetal growth predict bone size and volumetric density at 4 years old. **J Bone Miner Res**. 2010 Apr; 25(4):920–7.
7. Harvey NC, Mahon PA, Kim M, et al. Intrauterine growth and postnatal skeletal development: findings from the Southampton Women's Survey. **Paediatr Perinat Epidemiol.** 2012; 26(1):34–44.
8. Jensen RB, Vielwerth S, Frystyk J, et al. Fetal growth velocity, size in early life and adolescence, and prediction of bone mass: association to the GH-IGF axis. **J Bone Miner Res.** 2008; 23(3):439–46.
9. Lucas A, Fewtrell MS, Cole TJ. Fetal origins of adult disease-the hypothesis revisited. **BMJ.** 1999; 319(7204):245–9.
10. Keijzer-Veen MG, Euser AM, van Montfoort N, Dekker FW, Vandenbroucke JP, Van Houwelingen HC. A regression model with unexplained residuals was preferred in the analysis of the fetal origins of adult diseases hypothesis. **J Clin Epidemiol.** 2005; 58(12):1320–4.
11. Jaddoe VW, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. **Eur J Epidemiol**. 2012; 27(9):739–56.
12. Verburg BO, Steegers EA, De Ridder M, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. **Ultrasound Obstet Gynecol.** 2008; 31(4):388–96.
13. Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. **Br J Obstet Gynaecol.** 1975; 82(9):702–10.
14. Hadlock FP, Harrist RB, Deter RL, Park SK. Fetal femur length as a predictor of menstrual age: sonographically measured. **AJR Am J Roentgenol**. 1982; 138(5):875–8.
15. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal abdominal circumference as a predictor of menstrual age. **AJR Am J Roentgenol**. 1982; 139(2):367–70.
16. Shepard M, Filly RA. A standardized plane for biparietal diameter measurement. **J Ultrasound Med**. 1982; 1(4):145–50.
17. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. **Am J Obstet Gynecol**. 1985; 151(3):333–7.
18. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977–1981). **Acta Paediatr Scand.** 1991; 80(8–9):756–62.
19. Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in The Netherlands 1955–1997. **Pediatr Res.** 2000; 47(3):316–23.
20. Heppe DH, Kiefte-de Jong JC, Durmus B, et al. Parental, fetal, and infant risk factors for preschool overweight: the Generation R Study. **Pediatr Res.** 2013; 73(1):120–7.
21. Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. **BMJ**. 2000; 320(7240):967–71.
22. Heppe DH, Medina-Gomez C, Hofman A, Franco OH, Rivadeneira F, Jaddoe VW. Maternal first-trimester diet and childhood bone mass: the Generation R Study. **Am J Clin Nutr**. 2013; 98(1):224–32.

23. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008; 19(10):1369–78.
24. Heaney RP. Bone mineral content, not bone mineral density, is the correct bone measure for growth studies. **Am J Clin Nutr**. 2003; 78(2):350–1; author reply 351–2.
25. Prentice A, Parsons TJ, Cole TJ. Uncritical use of bone mineral density in absorptiometry may lead to size-related artifacts in the identification of bone mineral determinants. **Am J Clin Nutr.** 1994; 60(6): 837–42.
26. Warner JT, Cowan FJ, Dunstan FD, Evans WD, Webb DK, Gregory JW. Measured and predicted bone mineral content in healthy boys and girls aged 6–18 years: adjustment for body size and puberty. **Acta Paediatr**. 1998; 87(3):244–9.
27. Durmus B, van Rossem L, Duijts L, et al. Breast-feeding and growth in children until the age of 3 years: the Generation R Study. **Br J Nutr.** 2011; 105(11):1704–11.
28. Cooper C, Harvey N, Cole Z, Hanson M, Dennison E. Developmental origins of osteoporosis: the role of maternal nutrition. **Adv Exp Med Biol**. 2009; 646:31–9.
29. Cooper C, Cawley M, Bhalla A, et al. Childhood growth, physical activity, and peak bone mass in women. **J Bone Miner Res.** 1995; 10(6):940–7.
30. Rudang R, Mellstrom D, Clark E, Ohlsson C, Lorentzon M. Advancing maternal age is associated with lower bone mineral density in young adult male offspring. Osteoporos Int. 2012; 23(2):475–82.
31. Harvey NC, Javaid MK, Poole JR, et al. Paternal skeletal size predicts intrauterine bone mineral accrual**. J Clin Endocrinol Metab**. 2008; 93(5):1676–81.
32. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological, clinical research: potential, pitfalls. **BMJ.** 2009; 338:b2393.
33. Lenth RV. Java Applets for Power and Sample Size [Computer software]. Retrieved 2014 May 15. Available from: http://www.stat.uiowa.edu/~rlenth/Power.
34. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? **Epidemiology**. 2006; 17(4):413–8.
35. Tunon K, Eik-Nes SH, Grottum P. A comparison between ultrasound and a reliable last menstrual period as predictors of the day of delivery in 15,000 examinations. **Ultrasound Obstet Gynecol**. 1996; 8(3):178–85.
36. Altman DG, Chitty LS. New charts for ultrasound dating of pregnancy. Ultrasound **Obstet Gynecol**. 1997; 10(3):174–91.
37. Robinson HP, Sweet EM, Adam AH. The accuracy of radiological estimates of gestational age using early fetal crown-rump length measurements by ultrasound as a basis for comparison. **Br J Obstet Gynaecol**. 1979; 86(7):525–8.
38. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. Ultrasound **Obstet Gynecol.** 2005; 25(1):80–9.
39. Howe LD, Tilling K, Lawlor DA. Accuracy of height and weight data from child health records. **Arch Dis Child.** 2009; 94(12):950–4.
40. Kuh D, Wills AK, Shah I, et al. Growth from birth to adulthood and bone phenotype in early old age: a British birth cohort study. **J Bone Miner Res.** 2014; 29(1):123–33.
41. Beltrand J, Alison M, Nicolescu R, et al. Bone mineral content at birth is determined both by birth weight and fetal growth pattern. **Pediatr Res.** 2008; 64(1):86–90.
42. Setia S, Sridhar MG. Changes in GH/IGF-1 axis in intrauterine growth retardation: consequences of fetal programming? **Horm Metab Res.** 2009; 41(11):791–8.
43. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. **N Engl J Med**. 2008; 359(1):61–73.
44. Devlin MJ, Bouxsein ML. Influence of pre- and peri-natal nutrition on skeletal acquisition and maintenance. **Bone**. 2012; 50(2):444–51.

45. Goodfellow LR, Cooper C, Harvey NC. Regulation of placental calcium transport and offspring bone health. **Front Endocrinol (Lausanne).** 2011; 2:3.
46. Wong IP, Nguyen AD, Khor EC, et al. Neuropeptide Y is a critical modulator of leptin's regulation of cortical bone. **J Bone Miner Res**. 2013 Apr; 28(4):886–98.
47. Jaquet D, Leger J, Levy-Marchal C, Oury JF, Czernichow P. Ontogeny of leptin in human fetuses and newborns: effect of intrauterine growth retardation on serum leptin concentrations. **J Clin Endocrinol Metab**. 1998; 83(4):1243–6.
48. Javaid MK, Godfrey KM, Taylor P, et al. Umbilical cord leptin predicts neonatal bone mass. **Calcif Tissue Int**. 2005; 76(5):341–7.
49. Jaquet D, Leger J, Tabone MD, Czernichow P, Levy-Marchal C. High serum leptin concentrations during catch-up growth of children born with intrauterine growth retardation. **J Clin Endocrinol Metab**. 1999; 84(6):1949–53.
50. Cianfarani S, Ladaki C, Geremia C. Hormonal regulation of postnatal growth in children born small for gestational age. **Hormone Res**. 2006; 65(Suppl 3):70–4.
51. Javaid MK, Godfrey KM, Taylor P, et al. Umbilical venous IGF-1 concentration, neonatal bone mass, and body composition. **J Bone Miner Res**. 2004; 19(1):56–63.
52. van Coeverden SC, Netelenbos JC, de Ridder CM, Roos JC, Popp-Snijders C, Delemarre-van de Waal HA. Bone metabolism markers and bone mass in healthy pubertal boys and girls. **Clin Endocrinol (Oxf)**. 2002; 57(1):107–16.
53. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. **Br J Nutr**. 2007; 97(6): 1064–73.
54. Verkauskiene R, Jaquet D, Deghmoun S, Chevenne D, Czernichow P, Levy-Marchal C. Smallness for gestational age is associated with persistent change in insulin-like growth factor I (IGF-I) and the ratio of IGF-I/IGF-binding protein-3 in adulthood. **J Clin Endocrinol Metab**. 2005; 90(10):5672–6.
55. Cianfarani S, Geremia C, Scott CD, Germani D. Growth, IGF system, and cortisol in children with intrauterine growth retardation: is catch-up growth affected by reprogramming of the hypothalamic-pituitary-adrenal axis? **Pediatr Res**. 2002; 51(1):94–9.
56. Remer T, Boye KR, Hartmann MF, et al. Adrenal steroid hormones and metaphyseal bone in children. **Hormone Res**. 2004; 62(5):221–6.
57. Remer T, Boye KR, Hartmann M, et al. Adrenarche and bone modeling and remodeling at the proximal radius: weak androgens make stronger cortical bone in healthy children. **J Bone Miner Res**. 2003; 18(8):1539–46.



Chapter 3.3

Bone mass and strength in school age children exhibit sexual dimorphism related to differences in lean mass: the Generation R Study

Carolina Medina-Gomez, Denise HM Heppe, Jia-Lian Yin, Katerina Trajanoska, André G Uitterlinden, Thomas J Beck, Vincent WV Jaddoe, and Fernando Rivadeneira

J Bone Miner Res. 2015 Nov 24. doi: 10.1002/jbmr.2755.

ABSTRACT

Bone strength, a key determinant of fracture risk, has been shown to display clear sexual dimorphism after puberty. We sought to determine whether sex differences in bone mass and hip bone geometry as an index of strength, exist in school age pre-pubertal children and the degree to which the differences are independent of body size and lean mass. We studied 3,514 children whose whole body and hip scans were measured using the same densitometer (GE-Lunar iDXA) at a mean age of 6.2 years. Hip DXA scans underwent hip structural analysis (HSA) with derivation of bone strength indices. Sex-differences in these parameters were assessed by regression models adjusted for age, height, ethnicity, weight and lean mass fraction (LMF).Whole body BMD and BMC levels were 1.3% and 4.3% higher in girls after adjustment by LMF. Independent of LMF, boys had 1.5% shorter femurs, 1.9% and 2.2% narrower shaft and femoral neck with 1.6-3.4% thicker cortices than girls. Consequent with this geometry configuration, girls observed 6.6% higher stresses in the medial femoral neck than boys. When considering LMF, the sexual differences on the derived bone strength indices were attenuated, suggesting that differences in muscle loads may reflect an innate disadvantage in bone strength in girls, as consequence of their lower muscular acquisition. In summary, we show that bone sexual dimorphism is already present at 6 years of age, with boys having stronger bones than girls, relation which is influenced by body composition and likely due to differential adaption to mechanical loading. Our results support the view that early life interventions (i.e., increased physical activity) targeted during the pre- and peri-pubertal stages may be of high importance particularly in girls as before puberty onset, muscle mass is strongly associated with bone density and geometry in children.

INTRODUCTION

Bone strength is determined by bone material properties and by how this material is distributed to resist applied mechanical forces (structural geometry). These factors are adaptive in nature and believed to evolve differently between sexes through adolescence and adulthood [1-3]. Such sexual dimorphism of bone and its susceptibility to fracture has been attributed to the distinctive hormonal changes pinnacling in adolescence, as well as those occurring in women during menopause. An important aspect of skeletal growth is that the structural geometry of bone continually adapts to muscle forces applied during physical activity. The theoretical background for such adaptation is expressed in the Frost Mechanostat, which basically states that throughout life and especially during skeletal growth, bone tissue is placed where it is needed and removed from where it is not, always in response to strain stimuli imposed by muscle forces and detected by cells within and upon bone surfaces[4]. Previous work has shown that during puberty, girls accrue more bone mass relative to muscle mass than boys, presumably due to the higher release of estrogen [5,6]. Before puberty, the presence and origin of sex differences in bone accrual are less clear [1,5,7].

Bone development processes alter the amount of bone tissue within bones as well as its spatial placement, with implications evident in bone density and bone structural aspects (geometry); different bone configurations with the same BMD can have very different mechanical strengths [8]. Forces acting along the bone axis are supported by its cross-sectional area (CSA), while bending is resisted by the distribution of bone from its center of mass and quantified by the section modulus (Z) [9]. Presumably at any point in skeletal growth, bone cross-sections will have adapted to prevalent loading conditions so that stresses would be expected to be equal regardless of body size and other factors [10].

Previous research suggests that the skeleton is most sensitive to mechanical loading during childhood and adolescence[11,12]. The mechanical stimuli applied to bone are induced by muscle contractions that move the body in opposition to gravitational forces. Yet, the relative contribution of each of these sources (including ground, substrate and joint reaction forces), to the adaptive response of bone remains controversial given their interdependence[13,14]. These osteogenic stimuli may differ between boys and girls, but are difficult to quantify as they depend on factors like: body size and weight, physical activity intensity and type of activity among others. In the present study, we aimed to examine pre-pubertal sex differences in bone accrual and hip structure in children of 6 years of age. Specifically, we investigated if sex differences in children of school age observed across bone mass, density and geometry parameters are explained by distinct patterns of scaling to loading and differences in bone stresses.

MATERIALS AND METHODS

Study Population

The Generation R Study is a prospective prenatal cohort study in which 9,778 pregnant women living in Rotterdam and with delivery date between April 2002 and January 2006 were enrolled. Details of study design and data collection can be found elsewhere [15]. Ethnic background of the children was assessed by country of birth of her/his parents using a questionnaire. The participating child was deemed to be of non-Dutch ethnic origin if one parent was born abroad (definition Statistics Netherlands [16,17]). If both parents were born abroad, the country of birth of the participant's mother defined the ethnic background. Fifteen different ethnicities based on parental country of birth, were narrowed to three main ancestry groups including: Europeans: all European countries, together with Americans, Oceanics and North Africans (with parents from: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia, and Western Sahara) [18]; Africans: including Sub Saharan Africans, Dutch Antilleans and Surinamese Creoles; and Asians: including all Asian countries and Surinamese Hindustanis. The study was approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam (MEC 198.782/2001/31), and conducted according to the guidelines of the Helsinki Declaration with written informed consent obtained from all participants.

Bone mineral density and anthropometric measurements

Participants underwent dual-energy X-ray absorptiometry (DXA), following standard manufacturer protocols, using the same GE-Lunar iDXA device (GE Healthcare Lunar, Madison, WI) during their visit to the research centre at a mean age of 6 years. The coefficient of variation of whole-body scans was determined earlier to be 0.23% [19]. Of the 6,508 children with whole body scans, 3,976 (61.1%) also had total hip scans measured. All scans were analysed by well-trained research assistants using the enCORE software by the same manufacturer. After quality control, 107 hip and 62 whole body scans were discarded due to incorrect positioning, movement during the scan or device errors. Weight, whole body lean and fat mass were derived from the DXA scans. Lean mass fraction (LMF) was calculated as whole body lean mass divided by weight. Femur length (FL) was measured from whole body DXA scans using the ruler tool of the enCORE software along a straight line extending from the highest point of the greater trochanter to the distal midpoint between the femoral condyles. Participant height was measured to the nearest 0.1 cm using a Harpenden stadiometer (Holtain Limited, Dyfed, U.K.).

Hip Structural Analysis

Structural parameters were derived from the total hip scans using hip structural analysis (HSA) software developed by Beck and colleagues [20]. Precision studies done in adults (postmenopausal women) using older generation scanners yielded coefficients of variation

between 1- 5% [21]. Parameters are calculated from the proximal femur in three different regions: Narrow Neck (NN), across the narrowest diameter of the femoral neck, Intertrochanteric (IT), along the bisector of the neck-shaft angle, and the femur shaft (FS), at a point 1.5 times minimum femoral neck width distal to the intersection of the neck and shaft axes. The latter region is considered to be 100% cortical bone (requiring no assumptions on trabecular/cortical bone content) and is relatively axially symmetric.

In this study we focused on the cross-sectional bone surface area (CSA), section modulus (Z) as well as BMD, estimated cortical thickness (CT), the outer diameter (W) at the shaft and narrow neck regions. We also compared Neck- Shaft Angle, and neck length which influence the bending moments at the proximal end of the femur.

Derived strength indices

We used two types of measurements to assess strength, namely: *strength-to-load index* at the shaft and *net stress* at the femoral neck using a mechanical model [22]. As bending bone strength scales to Z, we calculated a composite index of bone strength relative to skeletal loading, comparing the effect of weight and LMF in both boys and girls, using the following equations.

Strength-to-load Index (SLI):

$$FS\_SLI_{weight} = \frac{FS\_Z/FL}{Weight}$$

And additionally adjusting for LMF,

$$FS\_SLI_{LMF} = \frac{FS\_Z/FL}{Weight*LMF}$$

whereas, the stress analysis incorporates the forces acting on the femur at its narrowest point using engineering beam theory. As,

Net stress (NS) on the medial surface of the femoral neck from both bending and axial forces is given by

$$\delta_{Medial} = \frac{M_S * y_{Medial}}{I} + \frac{P_S}{A}$$

where M_s and P_s are the bending moment and axial load on the neck cross-section. I and A are the moment of inertia and cross-sectional area of the narrow neck, and y_{medial} is the perpendicular distance between the medial surface and the neutral axis. Correction for lean mass was done by multiplying the resulting stress by LMF. A detailed description of the derivation of medial femoral neck stance stress has been described elsewhere [22].

Statistical analysis

The analyses reported here are based on a sample of 3,514 children within an age range of 4.9 – 9.1 years (where <5% of the population was over 8 years old), who had good quality DXA scans, underwent HSA analysis and had valid information on height and ethnicity.

Whole body and Femoral Neck DXA parameters

As recommended for paediatric populations by the International Society for Clinical Densitometry whole body less head (WBLH) was used to assess bone density and content in the children [23]. Femoral Neck (FN) BMD was measured on the right leg for all participants. Multivariate regression analyses were performed followed by the calculation of least square (LS)-means of WBLH-BMC and WBLH- and FN-BMD for each sex. We determined sex differences after accounting for the effect of age, height, weight and ethnic background (first model). Next, we adjusted further for LMF (second model), and compared again the sex derived LS-means.

Hip Structural Analysis

We compared BMD, CT, W, CSA, and Z -both at the NN and FS regions-between boys and girls by LS-means using the basic (first) model and further adjusting for LMF, as explained above. In addition, we corrected the models using femur length rather than height. Sex comparisons for the derived strength indices were adjusted only by age and ethnicity, as in their formulation they include a size parameter. Pearson correlations of these indices corrected by LMF, and DXA bone parameters adjusted by age, height, gender and ethnicity were also calculated.

Based on previous studies, we performed all described analyses adjusting further for gestational age, birthweight, breastfeeding history TV watching, participation in sports and hours playing outside, as well as maternal height and weight before pregnancy. To avoid loss of power or potential biases associated with missing data we used multivariate imputation by chained equations (MICE) procedure [24]. The amount of missing values ranged from 1% to 40% across these variables. We pooled results of the analyses performed in each of 10 imputed datasets. All statistical analyses were checked for multicollinearity using variation inflation factors (VIF) and performed using R statistical software [25].

RESULTS

Descriptive characteristics of the 3,514 children under study are shown in **Table 1**. Within the three defined ethnicities there were approximately the same number of boys and girls. Between boys and girls there were no differences in age, weight, WBLH-BMD, or WBLH-

Table 1. Descriptive Characteristics of Study Participants

	Boys (N=1,732)	**Girls (N=1,782)**	**Δ %[a]**	P
Ethnicity				
European Ancestry	1432	1483		0.165
Asian Ancestry	108	101		0.493
African Ancestry	192	198		0.667
Anthopometrics				
Age (years)	6.175 ± 0.633	6.147 ± 0.587	0.45%	0.167
Height (m)	**1.197 ± 0.628**	1.189 ±0.618	0.67%	5.37E-05
Weight (Kg)	23.21 ± 4.36	22.99 ± 4.62	0.95%	0.15
Lean mass (Kg)	**16.95 ± 2.44**	15.75 ± 2.29	7.08%	<2.20E-16
Lean mass fraction	**0.736 ± 0.048**	0.692 ± 0.053	5.98%	<2.20E-16
Fat mass (Kg)	5.42 ± 2.28	**6.42 ± 2.67**	-18.45%	<2.20E-16
Fat mass fraction	0.227 ± 0.050	**0.272 ± 0.055**	-19.82%	<2.20E-16
Femur length (cm)	27.28 ± 1.98	**27.68 ± 2.02**	-1.47%	3.25E-09
DXA parameters				
WBLH-BMC (g)	533.14 ± 114.50	528.71 ± 110.94	0.83%	0.240
WBLH-BMD (g/cm2)	0.55 ± 0.056	0.549 ± 0.053	0.18%	0.510
FN-BMD (g/cm2)	**0.727 ± 0.095**	0.681 ± 0.090	6.33%	<2.20E-16
HSA parameters				
FS_BMD (g/cm2)	**0.419 ± 0.059**	0.396 ±0.058	5.49%	<2.20E-16
NN_BMD (g/cm2)	**0.391 ± 0.056**	0.362 ± 0.052	7.42%	<2.20E-16
FS_CSA (cm2)	**0.747 ± 0.128**	0.707 ± 0.123	5.35%	<2.20E-16
NN_CSA (cm2)	**0.751 ± 0.143**	0.706 ± 0.133	5.99%	<2.20E-16
FS_width (cm)	1.872 ± 0.163	1.876 ± 0.160	-0.21%	0.452
NN_width (cm)	2.019 ± 0.228	**2.048 ± 0.234**	-1.44%	2.00E-04
FS_Z (cm3)	**0.252 ± 0.060**	0.240 ± 0.057	4.76%	1.43E-10
NN_Z (cm3)	**0.254 ± 0.072**	0.236 ± 0.066	7.09%	1.49E-14
FS_CT (mm)	**1.377 ± 0.211**	1.295 ± 0.205	5.95%	<2.20E-16
NN_CT (mm)	**0.742 ± 0.011**	0.686 ± 0.105	7.55%	<2.20E-16
Neck-shaft angle	**143.34 ± 6.18**	140.12 ± 6.14	2.25%	<2.20E-16
Necklength	**3.50 ± 0.57**	3.30 ± 0.52	5.71%	<2.20E-16

[a] $\Delta\ \% = \frac{(P_{boys} - P_{girls})}{P_{boys}} * 100$, where P represent the parameter of interest

WBLH= whole body less head; BMC=bone mineral content; BMD=bone mineral density; FN=femoral neck; FS=femur shaft; NN=narrow neck; CSA=cross-sectional area; Z=section modulus; CT=cortical thickness; LMF=lean mass fraction.

Table 2. Comparison of marginal means for regression models of BMD DXA and HSA parameters

	First model[a]				Second model[b]			
	Boys	Girls	Difference[c]	P	Boys	Girls	Difference[c]	P
WBLH-BMD (g/cm^2)	0.551 [0.549, 0.554]	0.553 [0.551, 0.555]	-0.0017 [-0.004, 0.0005]	0.135	0.549 [0.547, 0.551]	**0.556 [0.554, 0.559]**	-0.0074 [-0.0099, -0.0048]	<0.0001
WBLH-BMC (g)	531.50 [528.79, 534.21]	**536.58 [533.87, 539.30]**	-5.083 [-7.794, -2.373]	0.0002	525.59 [522.89, 528.29]	**545.17 [542.37 , 547.98]**	-19.580 [-22.704, -16.455]	<0.0001
FN-BMD (g/cm^2)	**0.722 [0.716, 0.727]**	0.679 [0.673, 0.684]	0.0429 [0.0374, 0.0485]	<0.0001	**0.712 [0.707, 0.718]**	0.693 [0.687, 0.698]	0.0192 [0.0128, 0.0257]	<0.0001
FS_BMD (g/cm^2)	**0.415 [0.412, 0.419]**	0.394 [0.391, 0.398]	0.0208 [0.0175, 0.0242]	<0.0001	**0.409 [0.406, 0.412]**	0.404 [0.400, 0.407]	0.0053 [0.0014, 0.0092]	0.0077
NN_BMD (g/cm^2)	**0.388 [0.385, 0.391]**	0.361 [0.358, 0.364]	0.0267 [0.0235, 0.0300]	<0.0001	**0.383 [0.379, 0.386]**	0.369 [0.366, 0.373]	0.0130 [0.0092, 0.0168]	<0.0001
FS_CSA (cm^2)	**0.739 [0.734, 0.745]**	0.708 [0.702, 0.714]	0.0311 [0.0255, 0.0367]	<0.0001	0.724 [0.719, 0.729]	0.729 [0.724, 0.735]	-0.0058 [-0.0121, 0.0006]	0.084
NN_CSA (cm^2)	**0.743 [0.736, 0.749]**	0.706 [0.699, 0.713]	0.0369 [0.0300, 0.0437]	<0.0001	**0.731 [0.724, 0.738]**	0.723 [0.716, 0.730]	0.0081 [0.00003, 0.01608]	0.045
FS_width (cm)	1.868 [1.860, 1.876]	**1.884 [1.876, 1.892]**	-0.0166 [-0.0247, -0.0084]	0.0001	1.860 [1.851, 1.868]	**1.896 [1.887, 1.904]**	-0.0359 [-0.0456, -0.0263]	<0.0001
NN_width (cm)	2.01 [1.995, 2.021]	**2.050 [2.037, 2.063]**	-0.0415 [-0.0545, -0.0285]	<0.0001	2.007 [1.994, 2.021]	**2.052 [2.038, 2.066]**	-0.0446 [-0.0291, -0.0602]	<0.0001
FS_Z (cm^3)	**0.249 [0.247, 0.251]**	0.240 [0.238, 0.243]	0.0084[0.0061, 0.0108]	<0.0001	0.243 [0.240, 0.245]	**0.250 [0.248, 0.252]**	-0.0074 [-0.0099, -0.0047]	<0.0001
NN_Z (cm^3)	**0.250 [0.247, 0.253]**	0.237 [0.233, 0.240]	0.0134 [0.0101, 0.0168]	<0.0001	0.245 [0.242, 0.249]	0.244 [0.242, 0.248]	0.0011 [-0.0028, 0.0050]	0.546
FS_CT (mm)	**1.367 [1.355, 1.379]**	1.291 [1.278, 1.303]	0.0076 [0.0064, 0.0088]	<0.0001	**1.345 [1.332, 1.357]**	1.323 [1.31, 1.336]	0.0022 [0.0007, 0.0035]	0.003
NN_CT (mm)	**0.738 [0.731, 0.744]**	0.685 [0.678, 0.691]	0.0053 [0.0046, 0.0059]	<0.0001	**0.726 [0.719, 0.733]**	0.701 [0.694, 0.708]	0.0025 [0.0018, 0.0033]	<0.0001

[a]First model adjusted by: ethnicity, age, height, weight

[b]Second model adjusted by: ethnicity, age, height, weight and LMF

[c]Difference is the regression coefficient for sex taking boys as reference [95% Confidence interval]

All results adjusted means [95% confidence intervals]. WBLH= whole body less head; BMC=bone mineral content; BMD=bone mineral density; FN=femoral neck; FS=femur shaft; NN=narrow neck; CSA=cross-sectional area; Z=section modulus; CT=cortical thickness.

BMC. Boys had greater height, lean mass, LMF, and FN-BMD, but shorter femur lengths and lower body fat as compared to girls. With regard to the HSA variables, boys had greater NSA and longer neck lengths. In both the narrow neck and shaft regions, BMD was higher in boys consistent with their thicker cortices and greater bone cross sectional area; nonetheless, girls showed wider narrow necks than boys. Even though wider bones have greater bending resistance, the index of bending strength was still greater in boys. No significant differences between sexes were observed in shaft width or centroid position.

Results of the applied models for all studied variables are shown in detail in **Table 2**. In general, after adjustment by age, height, weight and ethnicity no differences in WBLH-BMD were observed across sexes, although boys had 0.96% lower BMC. Additional adjustment for LMF resulted in boys having 1.28% and 4.30% lower WBLH-BMD and WBLH-BMC than girls (**Figure 2**). Moreover in analyses not adjusted for LMF (first model), boys had 5.96% higher FN-BMD, 5.06% higher shaft and 6.96% higher narrow neck BMD than girls. Likewise, boys had thicker cortices (5.56% in the shaft and 7.22% in the narrow neck), larger CSA (4.19% in the shaft and 4.98% in the narrow neck), and higher section modulus (3.61% in the shaft and 5.20% in the narrow neck) than girls. In contrast, girls had wider shaft (0.86%) and narrow neck (1.99%) diameters. After correction for LMF, the differences in BMD were less pronounced, with boys still having 2.67% higher FN-BMD, 3.66% higher shaft BMD and 1.22% higher narrow neck BMD than girls. This was still the consequence of thicker cortices in the boys (1.64% greater in the shaft and 3.44% in the narrow neck). The

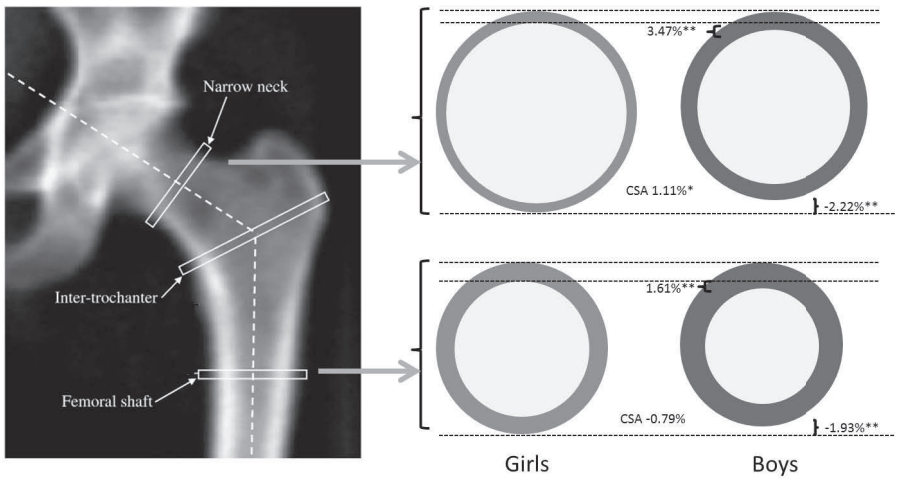


Figure 1. Sex differences in femur geometry at the narrow neck and femoral shaft after adjustment by age, height, weight, lean mass fraction and ethnicity. Reported percentages correspond to the sex difference in the parameters taking the boys as the reference group. Asterisks represent level of significance of the displayed differences (*P<0.05) and (**P<0.01). Boys have thicker cortices and narrower femur width than girls. Please note the drawing is for illustration purposes only and is not to scale.

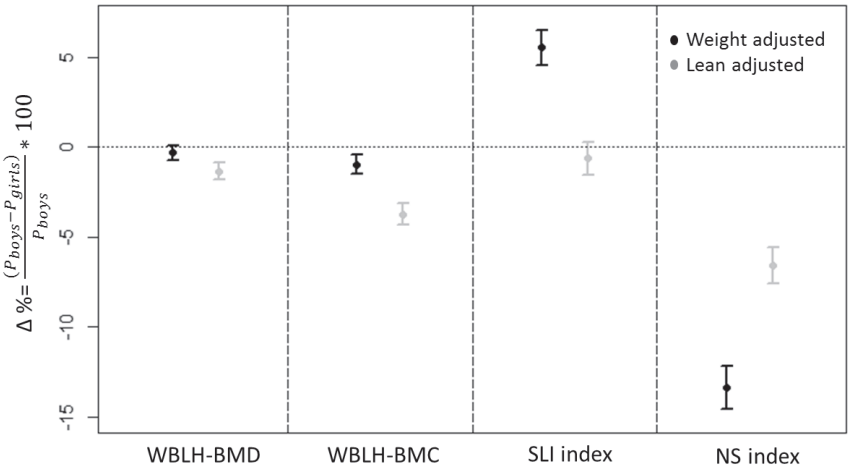


Figure 2. Sex differences in bone mass and strength parameters depending on the load applied. The strength-to-load index (SLI) was calculated at the shaft evaluating bone strength relative to skeletal loading. The stress (NS) index was calculated at the narrowest point of the femur neck, incorporating both bending and axial forces. Differences between boys and girls are presented as percentages in order to make them comparable across parameters

differences in bone dimensions were accentuated after LMF correction (**Figure 1**), with girls having 1.93% higher shaft and 2.22% narrow neck width. Similarly, after LMF correction the indices of axial strength (CSA) were now similar at the shaft (-0.79%, not significant) but still remained higher in boys at the narrow neck (1.11%); the index of bending strength (Z) became lower in boys than in girls (-2.88%) at the shaft while at the narrow neck was attenuated yet higher on boys (0.41%). As it is plausible that differences in temporality in the development and growth between the peripheral and axial skeleton exist and influence the results, we used femur length rather than height in our models as sensitivity analysis (Online Resource Table S1). Nonetheless, results and derived conclusions did basically not change.

Before adjustment for LMF the strength-to-load index calculated at the shaft based on weight only was 5.54% greater in boys, while stress of the femoral neck was 13.37% lower, as compared to girls. These differences were largely attenuated when LMF was considered in the derivation of both parameters. After LMF correction, the strength-to-load index was 0.64% higher in girls, yet no longer significant. The calculated stress was still higher in girls but with the difference reduced to 6.58% (**Figure 2** and **Table 3**). Correlation between these two strength parameters including LMF was moderate (ρ= -0.42). Adjusted WBLH-BMD had a correlation of 0.38 with strength-to-load index of the shaft and of -0.28 with the axial stress index at the narrow neck. Similarly, WBLH-BMC had a correlation of 0.40 with strength-to-load index of the shaft and of -0.25 with the axial stress at the narrow neck. After

Table 3. Strength indices and stress sex comparisons for the Generation R Study

	Boys	Girls	Difference[a]	P
FS_SLI_Weight	**3.96E-4 [3.92E-4, 4.00E-4]**	3.74E-4 [3.70E-4, 3.78E-4]	2.19E-5 [1.81E-5, 2.58E-5]	<0.0001
FS_SLI_WeightLMF	5.41E-4 [5.36E-4, 5.46E-4]	5.44E-4 [5.39E-4, 5.49E-4]	3.44E-6 [-8.26E-6, 1.38E-6]	0.181
FN_NS_Weight (Mpa)	8.758 [8.653, 8.864]	**9.928 [9.823,10.033]**	-1.17 [-1.066,-1.276]	<0.0001
FN_NS_WeightLMF (Mpa)	6.361 [6.296, 6.427]	**6.779 [6.713, 6.844]**	-0.418 [-0.353,-0.484]	<0.0001

[a]Difference is the regression coefficient for sex taking boys as reference [95% Confidence interval]
All results adjusted means [95% confidence intervals]. FS=Femur Shaft; SLI=strength-to-load index; LMF=lean mass fraction, NS= net stress
All models adjusted by: ethnicity and age

correction for LMF, FN-BMD showed correlations of 0.45 and of -0.53 with these indices, respectively. All reported correlations had P<0.0001.

Adjustment for further potential cofounders essentially did not modify these observations and all effect estimates did not statistically differ from those reported in tables 2 and 3. All reported models had a VIF under 10.

DISCUSSION

This study assessed sex differences in hip structure and femur strength in 3,514 pre-pubertal children with a mean age of 6.2 years. No differences in whole-body BMD or BMC were found between sexes. Nevertheless, after correction for LMF, it became apparent that girls have higher whole-body BMD and BMC per unit of lean mass. At the hip, BMD was higher in boys regardless of any applied adjustment. Additionally, boys had shorter femur length than girls. Hip bone geometry parameters showed that boys had thicker cortices and narrower femur widths at both the narrow neck and shaft regions, underlying their higher indices of axial (CSA) and bending (Z) strength. Likewise, girls observe greater stresses and lower or similar strength per load unit at the analysed sites when considering body weight. These results indicate that on average the femur of boys have better resistance to axial and bending forces at both the narrow-neck and proximal shaft. Yet, these sex differences are largely attenuated by lean mass adjustment, suggesting adaptation to differences in the amount of muscle mass and presumably to the forces it generates, while other mechanisms (i.e., genetic, hormonal and environmental determinants) could still play a role.

In agreement with previous reports we observed that pre-pubertal girls of our study had significantly higher whole body fat mass and lower lean mass than boys [5,26,27]. We found that sex differences in bone mass greatly depended on lean mass. After adjustment for LMF both WBLH-BMD and WBLH-BMC were higher in girls than in boys; while differences in BMC were more evident than differences in BMD (**Figure 2**). Higher BMC accrual per lean

mass unit in girls during puberty has been reported earlier [1,3,28], and here we show this difference appears to be already evident at six years of age.

In our study we describe strong sexual dimorphism at children of pre-pubertal school age. Sexual dimorphism of skeletal geometry is known to accentuate during puberty, where periosteal diameter expands to a greater extent in boys as compared with girls [29-32]. Among these studies, Sayers *et al*. reported the largest one performed in 3,914 English children of peri-pubertal age (mean age 13.8 years), also using geometrical and strength parameters derived from DXA showing that boys have thicker cortices and higher hip-bending strength [3]. Nonetheless, in that study they observed wider femurs in boys than girls, which is in contradiction to our findings. This may well reflect the different ages between ours and their study population at the time of measurement. Similarly, a Canadian study conducted in 138 adolescents also showed that boys develop greater periosteal apposition, resulting in greater femoral neck width and higher bending strength than girls. Nevertheless, in line with the findings of our study, measurements performed before the start of the growth spurt (peak height velocity), depicted narrower femoral width values in boys [29]. Another, interesting yet intriguing finding is that, we found girls to have longer femurs than boys despite having no differences in overall height. To the extent of our knowledge there are no previous studies describing differences in femur length between sexes during childhood. Replication of this finding in other populations should be attempted. Altogether, pre-pubertal girls have on average wider and longer bones, which may represent an adaption to the constraint on periosteal apposition to occur after onset of puberty.

While hormones play a key role on the bone sexual dimorphism arising during puberty their influence at earlier ages is less clear. Garnett et al. showed that sex differences in IGF-1, estradiol, testosterone, and leptin levels were already present in children of 7-8 years of age and were associated with LS-BMD [5]. Since we did not measure hormone levels, we cannot exclude that relevant hormonal influences (i.e., androgens) could also modulate bone metabolism directly and/or indirectly through bone composition and underlie the differences observed in our study. On the other hand, site specific regulation has also been described for both periosteal and endosteal apposition [3,33], even at the genetic level [34]. Therefore, it is possible that local regulation such as differential mechanical loading between bones, predominate over systemic effects like the influence of hormonal and nutritional factors in definition of bone characteristics [33].

Our study showed that the sexual dimorphism of the skeleton observed during the pre-pubertal period already reflects differences in bone strength, including higher stresses of the femoral neck and lower strength per unit load at the shaft in girls as compared to boys (**Table 3**). These two indices are moderately correlated hence providing complementary

information across femoral sites of measurement (shaft vs narrow-neck). In addition, they are modestly correlated with BMD indicating that they reflect mechanical properties which are not fully captured by areal BMD measurements. The higher stresses observed in the femoral neck of girls as compared to boys, are undoubtedly complex but may be due in part to thinner cortices in wider bones (**Figure 1**). The implications of these findings for the risk of osteoporosis and fracture in adult life remain to be determined. Nevertheless, the magnitude of these sex differences are not negligible (e.g., 5% lower femoral neck BMD in girls); considering the 1% yearly average decrease in BMD in the elderly and that a BMD decrease of 1 standard deviation represents between 1.5 to 3 times increased risk of fracture (all types and hip, respectively) [35].

As postulated earlier by other studies, we also observe that the described sexual dimorphism of bone is attenuated when considering differences in lean mass. This is expected as muscle-induced mechanical loads have an important role determining the three dimensional distribution of tissue within bones, as strains generated by muscle contraction have been reported to be more osteogenic than static gravitational loading, exerted by body weight alone [36]. Nonetheless, there is a high interdependence among both stimuli and their role in bone functional adaptation is by no means mutually exclusive [37]. In our analyses however, the magnitude of these forces, are only crudely represented by the indices we used, which consider body weight, lean body mass and fraction. Previous work has shown that in young women, bending strength of the hip (Z) scales with the proportion of lean mass in their bodies [38]. In our study, we also show this is the case and that in fact this relation is already evident in pre-pubertal children of school age. Differences in the strength-to-load index are no longer significant while the differences in stress reduce in more than 50% when including LMF, therefore lean mass has a great impact on bone mechanical resistance (**Figure 2**). The similar strength reached at the shaft after lean mass adjustment is not reflected in similar FS-BMD between boys and girls suggesting an stronger association of muscular stimuli with bone geometry than with bone density at this site. These results indicate that a program of physical activity at early ages could have positive effects on bone stability, as has been demonstrated by several studies [39-44]. Girls in particular would benefit the most from these interventions as in the general population they have lower levels of physical activity than boys [40] and their bone formation on the periosteal surface greatly decreases during puberty [45].

Our study arises in a powerful setting where research subjects are relatively standardized as they are all residents of the city of Rotterdam, The Netherlands, and were measured using the same dual-energy X-ray absorptiometry (DXA) device at a similar age. This study design reduces the impact of inter-machine variance, as well as local environmental and lifestyle exposures influencing bone health variation such as, major differences in access

to medical care, sunlight exposure, diet or physical activity. Eventually this also explains why correction for a large set of these confounders did not essentially modify the observed associations. Further, despite the multi-ethnic background of our cohort, ethnicity is not expected to play a role in these relationships, as we observed a very similar distribution of ethnic groups across sexes.

This study has some limitations. The three dimensional model makes assumptions as to the distribution of bone mineral at the narrow neck. The estimation of cortical thickness at the narrow neck (NN_CT) is the only parameter assuming a fixed proportion of cortical vs trabecular bone (which has not been validated in children). All other HSA parameters are derived at both sites (NN and shaft) directly from measured mineral mass (g/cm^2) and do not involve the estimate of cortical thickness (cm) or its distribution assumption at the femoral neck. Further, the shaft site is purely cortical so no assumption is required. In addition, an intrinsic error of the HSA method may arise when applied to children as their mineralization density is lower than that of adults, thus resulting in underestimation of Z and CSA [29] and potentially less precise measurements than in adults. Nonetheless, as the same error is introduced to the measurements regardless of sex of the participant, this or any other systematic errors are not expected to affect the absolute sex differences we observed. Further, given the consistency of the observed differences between sexes it is unlikely they can be explained by random error alone derived from less precise HSA assessments. It is unlikely that studies using three dimensional imaging techniques with greater radiation exposure (i.e., quantitative computerized tomography) can assess the hip structure parameters in a similar large set of healthy children to verify our findings. Nevertheless, studies (including our own) incorporating pQCT measurements are increasingly being performed in children, using cross-sections of the tibia and/or radius that could be used to verify and extrapolate our findings. Further, although we ran sensitivity analyses including other possible confounders found in literature, there is still the possibility of residual confounding due to measurement error in explanatory variables influencing our results. In addition, stimuli from muscle contractions were not directly assessed by our approach, under the assumption these influences are partially captured by lean mass. Moreover, given the observational nature of our study causal relationships between the observed differences in body composition (lean mass) and BMD or strength parameters cannot be unequivocally assessed.

In summary, we have shown here that sexual dimorphism in both bone mass and density measured at the whole body and femur, together with bone distribution at the femoral shaft and narrow neck, is already present at a pre-pubertal age; and that these sex differences are closely related to lean (muscle) mass. Given the relative "weaker" bone configuration of girls already present at an early age and the influence of lean mass fraction on these

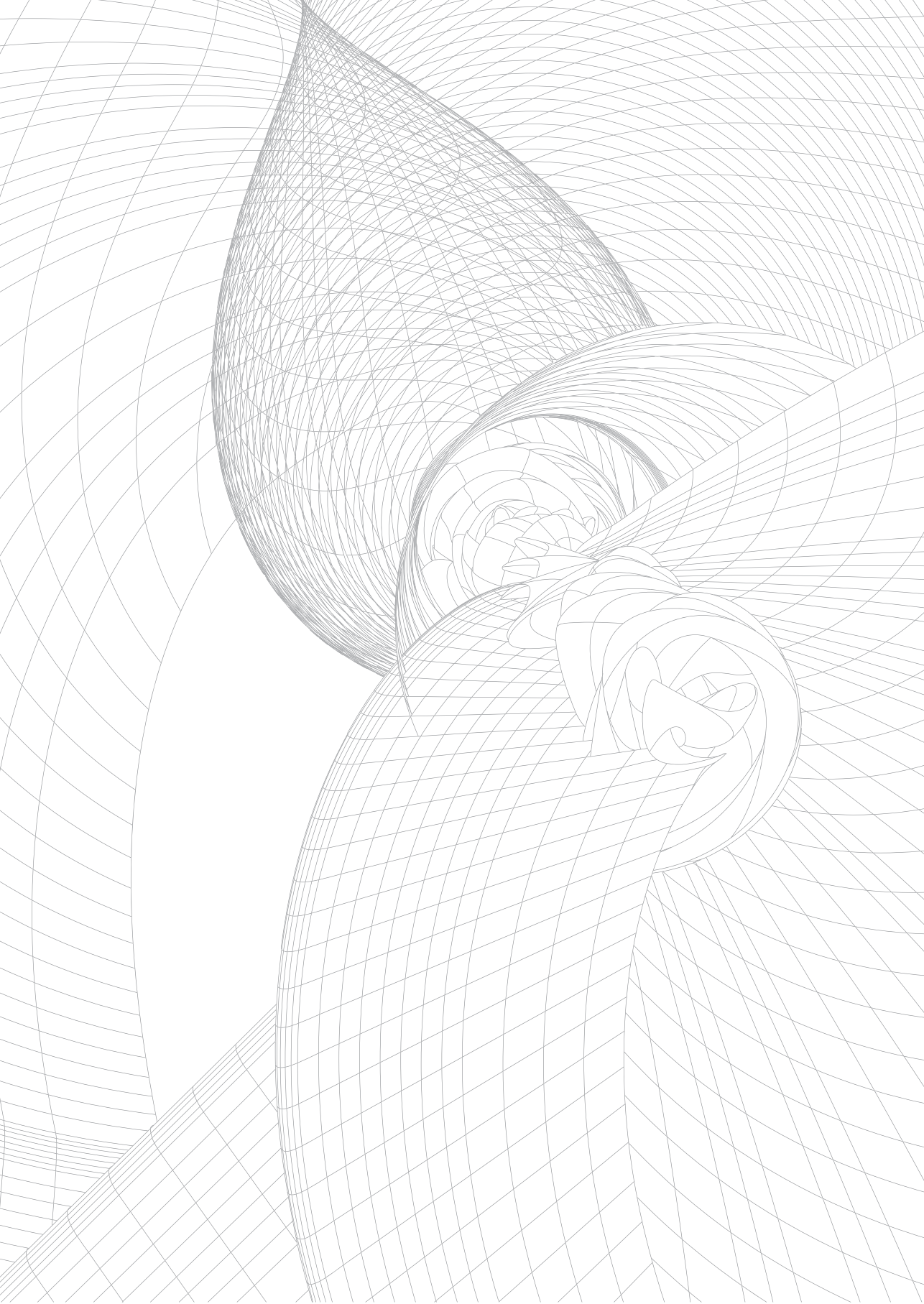
relationship, establishment of exercise programs during the pre- and peri-pubertal years may be an effective way to decrease fracture risk later in life.

Detailed acknowledgements and online resources can be found in the published article online: http://onlinelibrary.wiley.com/doi/10.1002/jbmr.2755/full

REFERENCES

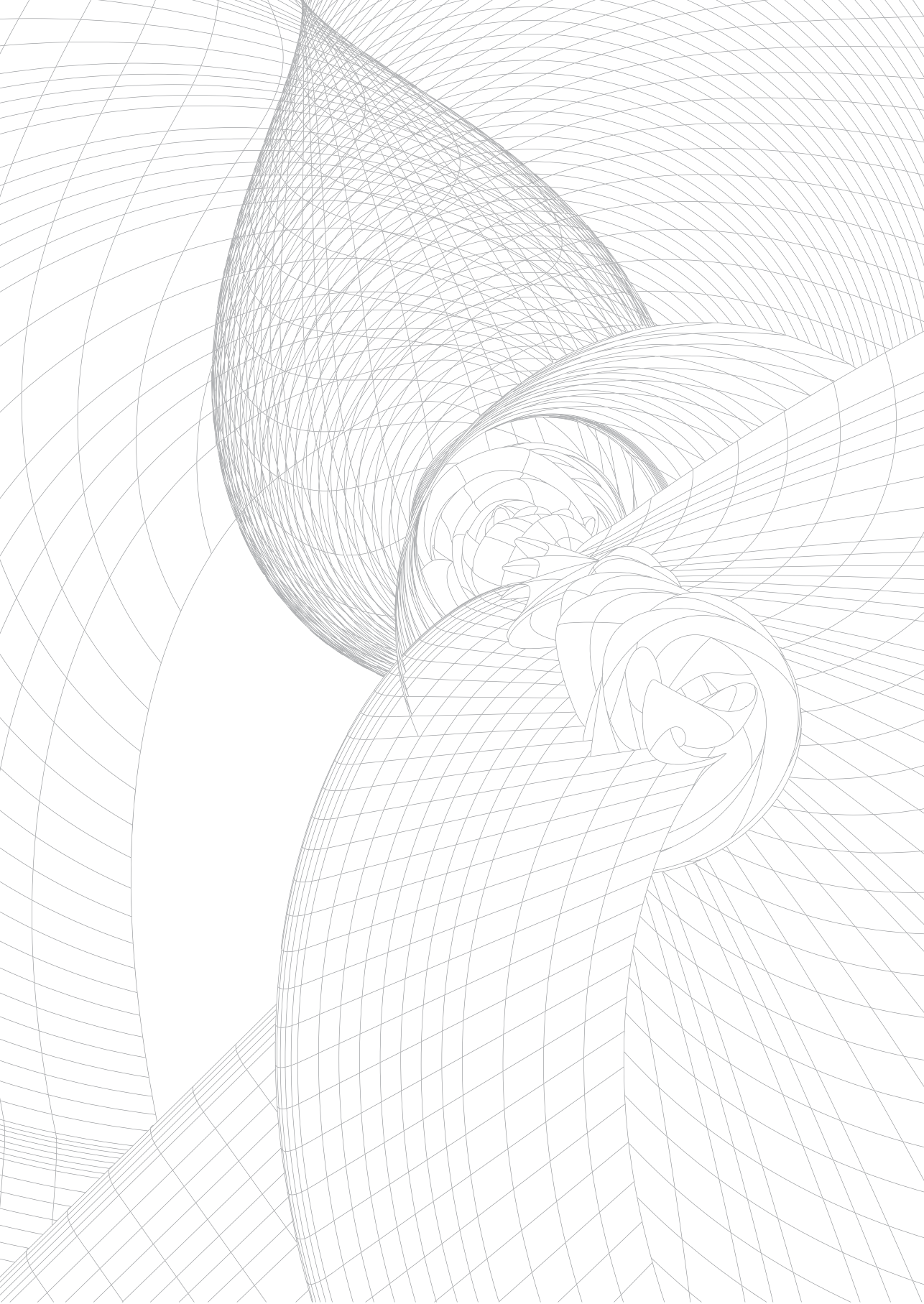
1. Ashby RL, Adams JE, Roberts SA, Mughal MZ, Ward KA. The muscle-bone unit of peripheral and central skeletal sites in children and young adults. **Osteoporos Int**. 2011;22(1):121-32.
2. Lang TF. The bone-muscle relationship in men and women. **J Osteoporos**. 2011;2011:702735.
3. Sayers A, Marcus M, Rubin C, McGeehin MA, Tobias JH. Investigation of sex differences in hip structure in peripubertal children. **J Clin Endocrinol Metab**. 2010;95(8):3876-83.
4. Frost HM. Bone's mechanostat: a 2003 update. **Anat Rec A Discov Mol Cell Evol Biol**. 2003;275(2): 1081-101.
5. Garnett SP, Hogler W, Blades B, et al. Relation between hormones and body composition, including bone, in prepubertal children. **Am J Clin Nutr**. 2004;80(4):966-72.
6. Schiessl H, Frost HM, Jee WS. Estrogen and bone-muscle strength and mass relationships. **Bone**. 1998; 22(1):1-6.
7. Horlick M, Thornton J, Wang J, Levine LS, Fedun B, Pierson RN, Jr. Bone mineral in prepubertal children: gender and ethnicity. **J Bone Miner Res**. 2000;15(7):1393-7.
8. Kaptoge S, Dalzell N, Loveridge N, Beck TJ, Khaw KT, Reeve J. Effects of gender, anthropometric variables, and aging on the evolution of hip strength in men and women aged over 65. **Bone**. 2003; 32(5):561-70.
9. Beck TJ, Looker AC, Mourtada F, Daphtary MM, Ruff CB. Age trends in femur stresses from a simulated fall on the hip among men and women: evidence of homeostatic adaptation underlying the decline in hip BMD. **J Bone Miner Res**. 2006;21(9):1425-32.
10. Ruff CB. Body size, body shape, and long bone strength in modern humans. **J Hum Evol**. 2000;38(2): 269-90.
11. Farr JN, Laddu DR, Going SB. Exercise, hormones and skeletal adaptations during childhood and adolescence. **Pediatr Exerc Sci**. 2014;26(4):384-91.
12. Janz KF, Letuchy EM, Eichenberger Gilmore JM, et al. Early physical activity provides sustained bone health benefits later in childhood. **Med Sci Sports Exerc**. 2010;42(6):1072-8.
13. Beck BR. Muscle forces or gravity--what predominates mechanical loading on bone? Introduction. **Med Sci Sports Exerc**. 2009;41(11):2033-6.
14. Judex S, Carlson KJ. Is bone's response to mechanical signals dominated by gravitational loading? **Med Sci Sports Exerc**. 2009;41(11):2037-43.
15. Jaddoe VW, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. **Eur J Epidemiol**. 2012;27(9):739-56.
16. Keij I. Standaarddefinitie allochtonen. Centraal Bureau voor de Statistiek Nederland; 2000.
17. Alders M. Classification of the population with a foreign background in the Netherlands. The Measure and Mismeasure of Populations The statistical use of ethnic and racial categories in multicultural societies. Paris: Centre d'Etudes et de Recherches Internationales (CERI) and the Institut National d'Etudes Démographiques (INED). 2001.
18. Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. United Nations Statistics Division; 2012.
19. Heppe DH, Medina-Gomez C, Hofman A, Franco OH, Rivadeneira F, Jaddoe VW. Maternal first-trimester diet and childhood bone mass: the Generation R Study. **Am J Clin Nutr**. 2013;98(1):224-32.
20. Beck TJ. Extending DXA beyond bone mineral density: understanding hip structure analysis. **Curr Osteoporos Rep**. 2007;5(2):49-55.
21. Khoo BCC, Beck TJ, Qiao QH, et al. In vivo short-term precision of hip comparison with bone mineral density structure analysis variables in using paired dual-energy X-ray absorptiometry scans from multi-center clinical trials. **Bone**. 2005;37(1):112-21.
22. Hamilton CJ, Jamal SA, Beck TJ, et al. Heterogeneity in Skeletal Load Adaptation Points to a Role for Modeling in the Pathogenesis of Osteoporotic Fracture. **Journal of Clinical Densitometry**. 2014;17(1): 170-6.

23. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
24. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. **J Stat Softw**. 2011;45(3):1-67.
25. Team" RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
26. Taylor RW, Gold E, Manning P, Goulding A. Gender differences in body fat content are present well before puberty. **Int J Obes Relat Metab Disord**. 1997;21(11):1082-4.
27. Mast M, Kortzinger I, Konig E, Muller MJ. Gender differences in fat mass of 5-7-year old children. **Int J Obes Relat Metab Disord**. 1998;22(9):878-84.
28. Hogler W, Briody J, Woodhead HJ, Chan A, Cowell CT. Importance of lean mass in the interpretation of total body densitometry in children and adolescents. **J Pediatr**. 2003;143(1):81-8.
29. Forwood MR, Bailey DA, Beck TJ, Mirwald RL, Baxter-Jones AD, Uusi-Rasi K. Sexual dimorphism of the femoral neck during the adolescent growth spurt: a structural analysis. **Bone**. 2004;35(4):973-81.
30. Seeman E. Pathogenesis of bone fragility in women and men. **Lancet**. 2002;359(9320):1841-50.
31. Kontulainen SA, Macdonald HM, Khan KM, McKay HA. Examining bone surfaces across puberty: a 20-month pQCT trial. **J Bone Miner Res**. 2005;20(7):1202-7.
32. Macdonald H, Kontulainen S, Petit M, Janssen P, McKay H. Bone strength and its determinants in pre- and early pubertal boys and girls. **Bone**. 2006;39(3):598-608.
33. Rauch F. Bone accrual in children: adding substance to surfaces. **Pediatrics**. 2007;119 Suppl 2:S137-40.
34. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **PLoS Genet**. 2014;10(6):e1004423.
35. Johnell O, Kanis JA, Oden A, et al. Predictive value of BMD for hip and other fractures. **J Bone Miner Res**. 2005;20(7):1185-94.
36. Petit MA, Beck TJ, Hughes JM, Lin HM, Bentley C, Lloyd T. Proximal femur mechanical adaptation to weight gain in late adolescence: a six-year longitudinal study. **J Bone Miner Res**. 2008;23(2):180-8.
37. Robling AG. Is bone's response to mechanical signals dominated by muscle forces? **Med Sci Sports Exerc**. 2009;41(11):2044-9.
38. Lloyd T, Beck TJ, Lin HM, et al. Modifiable determinants of bone status in young women. **Bone**. 2002; 30(2):416-21.
39. Warden SJ, Mantila Roosa SM, Kersh ME, et al. Physical activity when young provides lifelong benefits to cortical bone size and strength in men. **Proc Natl Acad Sci U S A**. 2014;111(14):5337-42.
40. Detter F, Rosengren BE, Dencker M, Lorentzon M, Nilsson JA, Karlsson MK. A 6-year exercise program improves skeletal traits without affecting fracture risk: a prospective controlled study in 2621 children. **J Bone Miner Res**. 2014;29(6):1325-36.
41. Tan VP, Macdonald HM, Kim S, et al. Influence of physical activity on bone strength in children and adolescents: a systematic review and narrative synthesis. **J Bone Miner Res**. 2014;29(10):2161-81.
42. MacKelvie KJ, Petit MA, Khan KM, Beck TJ, McKay HA. Bone mass and structure are enhanced following a 2-year randomized controlled trial of exercise in prepubertal boys. **Bone**. 2004;34(4):755-64.
43. Petit MA, McKay HA, MacKelvie KJ, Heinonen A, Khan KM, Beck TJ. A randomized school-based jumping intervention confers site and maturity-specific benefits on bone structural properties in girls: a hip structural analysis study. **J Bone Miner Res**. 2002;17(3):363-72.
44. Macdonald HM, Kontulainen SA, Khan KM, McKay HA. Is a school-based physical activity intervention effective for increasing tibial bone strength in boys and girls? **J Bone Miner Res**. 2007;22(3): 434-46.
45. Martin RB. Size, structure and gender: lessons about fracture risk. **J Musculoskelet Neuronal Interact**. 2002;2(3):209-11.



Chapter 4

Genetic studies of pediatric BMD



Chapter 4.1

Meta-analysis of Genome-Wide Scans for Total Body BMD in Children and Adults Reveals Allelic Heterogeneity and Age-specific Effects at the *WNT16* Locus

Carolina Medina-Gomez* , John P Kemp*, Karol Estrada, Joel Eriksson, Jeff Liu, Sjur Reppe, David M Evans, Denise HM Heppe, Liesbeth Vandenput, Lizbeth Herrera, Susan M Ring, Claudia J Kruithof, Nicholas J Timpson, M Carola Zillikens, Ole K Olstad, Hou-Feng Zheng, Brent Richards, Beate St Pourcain, Albert Hofman, Vincent WV Jaddoe, George Davey Smith, Mattias Lorentzon, Kaare M Gautvik, André G Uitterlinden, Robert Brommage, Claes Ohlsson, Jonathan H Tobias, and Fernando Rivadeneira

*These authors equally contributed to this manuscript

PLoS Genet. 2012 July; 8(7): e1002718. Published online 2012 July 5. doi: 10.1371/journal.pgen.1002718

ABSTRACT

To identify genetic loci influencing bone accrual we performed a genome-wide association scan for total-body bone mineral density (TB-BMD) variation in 2,660 children of different ethnicities. We discovered variants in 7q31.31 associated with BMD measurements with the lowest $P=4.1x10^{-11}$ observed for rs917727 with minor allele frequency of 0.37. We sought replication for all SNPs located +/-500 kb from rs917727 in 11,052 additional individuals from five independent studies including children and adults; together with *de-novo* genotyping of rs3801387 (in perfect linkage disequilibrium (LD) with rs917727) in 1,014 mothers of children from the discovery cohort. The top signal mapping in the surroundings of *WNT16* was replicated across studies with a meta-analysis $P=2.6x10^{-31}$ and an effect size explaining between 0.6-1.8% of TB-BMD variance. Conditional analyses on this signal revealed a secondary signal for total body BMD ($P=1.42x10^{-10}$) for rs4609139 and mapping to *C7orf58*. We also examined the genomic region for association with skull BMD to test if the associations were independent of skeletal loading. We identified two signals influencing skull BMD variation including rs917727 ($P=1.9x10^{-16}$) and rs7801723 ($P=8.9x10^{-28}$) also mapping to *C7orf58* (r^2=0.50 with rs4609139). *Wnt16* knockout (KO) mice with reduced total body BMD and gene expression profiles in human bone biopsies support a role of *C7orf58* and *WNT16* on the BMD phenotypes observed at the human population level. In summary, we detected two independent signals influencing total body and skull BMD variation in children and adults, thus demonstrating the presence of allelic heterogeneity at the *WNT16* locus. One of the skull BMD signals mapping to *C7orf58* is mostly driven by children suggesting temporal determination on peak bone mass acquisition. Our life-course approach postulates that these genetic effects influencing peak bone mass accrual may impact the risk of osteoporosis later in life.

AUTHOR SUMMARY

Genetic investigations on bone mineral density (BMD) variation in children allows the identification of factors determining peak bone mass and their influence on developing osteoporosis later in life. We ran a genome-wide association study (GWAS) for total body BMD based on 2,660 children of different ethnic background, followed by replication in an additional 12,066 individuals comprising children, young adults and elderly populations. Our GWAS meta-analysis identified two independent signals in the 7q31.31 locus, arising from SNPs in the vicinity of *WNT16*, *FAM3C* and *C7orf58*. These variants were also associated with skull BMD, a skeletal trait with much less environmental influence for which one of the signals displayed age-specific effects. Integration of functional studies in a *Wnt16* knockout mouse model and gene expression profiles in human bone tissue provided

additional evidence for *WNT16* and *C7orf58* to be underlying the described associations. All together our findings demonstrate the relevance of these factors for bone biology, the attainment of peak bone mass and their likely impact on bone fragility later in life.

INTRODUCTION

Roughly 30 to 50% of women and 15 to 30% of men experience an osteoporosis-related fracture during their lifetime [1]. In adults, bone mineral density (BMD) measured at skeletal sites where osteoporotic fractures occur more frequently (i.e., lumbar spine, hip and forearm) is used for the diagnosis of osteoporosis and assessment of fracture risk. BMD measured at a given point in time is the result of peak-bone mass acquisition and subsequent bone loss in later life.

Due to the rapid changes in bone area in early life, the total body measurement (less head) is the preferred measurement to evaluate bone health in children [2]. The total body BMD measurement (in both children and adults) incorporates components of both cortical (~80%) and to a lesser extent trabecular (~20%) bone [3]. Moreover, it is likely that the genes underlying skeletal growth and bone loss differ in importance across the lifespan and can act in a site specific manner [4-6]. Peak bone mass is an important determinant of the risk of osteoporosis later in life [7,8]. Early identification of individuals prone to low peak BMD may allow implementing strategies (interventions) which can delay the onset of osteoporosis.

From a genetic perspective, the discovery of loci influencing peak bone mass should be based on younger populations to avoid the noise introduced by bone loss later in life. A relatively recent Genome Wide Association Study (GWAS) in native British children successfully identified an association between total body derived BMD and variants in the osteoblast transcription factor gene *Osterix* [9], an early acting developmental gene shown to influence peak bone mass accrual but also BMD in adults [10,11] .

The purpose of this study was to identify genetic variants associated with total body BMD (TB-BMD) in children, thus targeting variants involved in bone accrual. We ran a GWAS on children from the multiethnic Generation R Study and then replicated our findings in five additional cohorts including Northern European individuals covering different age groups ranging from children to elderly adults, allowing any life-course effect of the discovered variants to be evaluated

RESULTS

Association with Total Body BMD in the Discovery Cohort

To search for loci influencing total body BMD variation we performed genome-wide association analysis in a subset of 2,660 children from the Generation R Study with DXA scans and GWAS data. The Generation R Study is a population-based multiethnic birth cohort currently assessing children at an average age of 6.1 (SD 0.28) years. Table S1 shows population characteristics of these children overall and stratified by ethnicity. To increase the genome coverage of common variants we imputed genotypes for 3,021,329 SNPs in reference to the combined CEU, CHB/JPT and YRI HapMap Phase II panels using MACH/minimac software taking into account the admixed nature of the Generation R population [12]. The GWAS for TB-BMD in these individuals adjusted for age, gender, weight and 20 principal components, showed appropriate control for population structure with genomic inflation factors (λ) approaching unity (Figure 1A), and revealed a genome-wide significant association (lowest $P=4.1x10^{-11}$ for rs917727) mapping to the 7q.31 locus (Figure 1B).

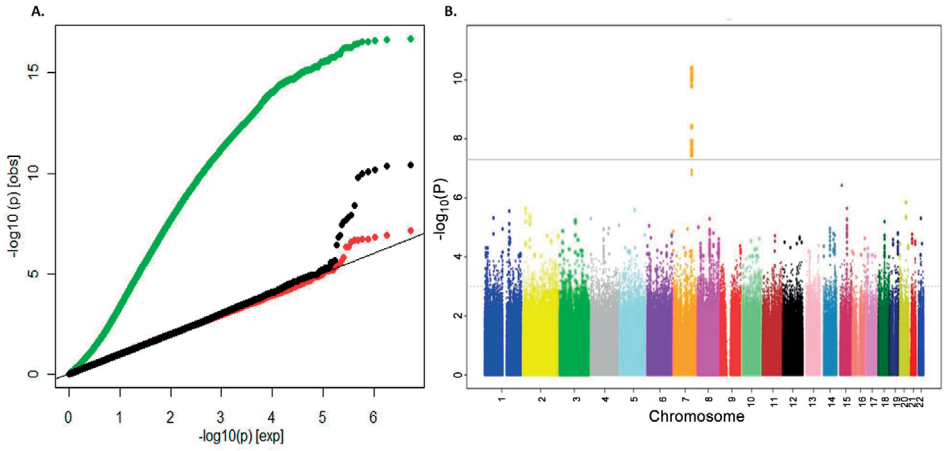


Figure 1. Genome-wide association of TB-BMD in the discovery cohort. A. Q-Q plot showing the inflation of the test statistics when correction for data structure is not applied (green dots) and the loss of power when no weight correction is applied (red dots) in comparison with the applied model (black dots) and B. Manhattan Plot of the genome wide association analysis of TB-BMD in the Generation R (discovery) cohort of model correcting by age, gender and body weight.

Replication of the 7q31.31 association signal

We sought replication of 721 SNPs spanning the region comprised by +/- 500 kb from the top associated SNP (rs917727). We did this across five populations with total body DXA scans and GWAS data including: the *Avon Longitudinal Study of Parents and their Children* (ALSPAC), the *Gothenburg Osteoporosis and Obesity Determinants* (GOOD) and the Rotterdam Study (RS-I, RS-II and RS-III) cohorts. These replication cohorts were selected to

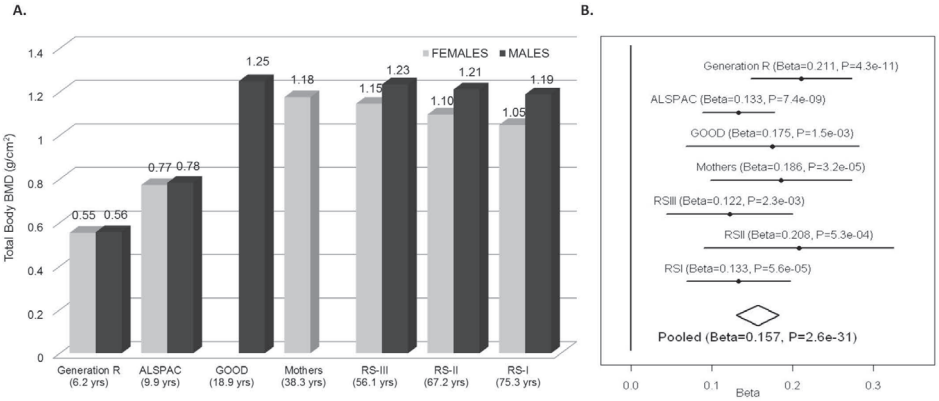


Figure 2. TB-BMD across cohorts and meta-analysis. A. Mean Total Body BMD in each cohort by gender showing the highest BMD levels in young adults and overall higher levels in male than in female participants. B. Forest plot of the association of TB-BMD and rs917727. Results are reported per copy of the G-allele (MAF=0.27).

cover a wide spectrum of age groups to assess the genetic association with total body BMD variation throughout different life periods (Figure 2A) comprising: ALSPAC children (n=5,334; mean age 9.9 years), GOOD young adults (n=938; mean age 18.9 years) and a set of individuals over age 45 years RS-III (n=1,594; mean age 56.1 years), RS-II (n=750; mean 67.2 years); and RS-I, (n=2,436, mean age 75.3 years). In addition, a sample of young women of Northern European descent (mothers of the Generation R participants) lacking GWAS scans, MoGENR, (n=1,014; mean age 38 years) were *de-novo* genotyped for rs3801387, a perfect proxy (r^2=1, based on the Hapmap phase II-CEU panel) of the top associated SNP. Detailed population characteristics of these cohorts can be found in Table S2. From the 721 SNPs used in the meta-analysis 20 surpassed the genome-wide significant threshold (Table 1) whilst a further 22 were suggestive of association ($P<1x10^{-5}$). These SNPs had a minor allele frequency (MAF) ranging between 0.23-0.30 across studies (Table 1). The top associated SNP was rs917727 ($P=1.28x10^{-27}$ and $P=2.6x10^{-31}$ when including Generation R mothers), which had a combined effect of 0.16 SD increment per copy of the minor allele (Figure 2B). The effect of the rs917727 explained on average 0.9% of the phenotypic variance in standardized BMD and had no significant evidence for statistical heterogeneity across the meta-analyzed cohorts (I^2=17%, P=0.306). The GWAS signal mapped to a 66.3 Kb region of high LD (r^2>0.8) between the *FAM3C* and *WNT16* genes (Figure 3A).

Conditional analyses for secondary signals

To assess the presence of allelic heterogeneity at the locus we carried out a conditional analysis conditioning on the top signal. After meta-analysis we identified a secondary independent signal mapping to *C7orf58* (Figure 3B) including 67 SNPs surpassing ($P<5x10^{-8}$) genome-wide significant level (Table S3). In general, most of these 67 SNPs were in high

Table 1. Genome-wide significant markers of the total body BMD GWAS meta-analysis

			DISCOVERY			REPLICATION																		
			Generation R			ALSPAC			GOOD			RS-III			RS-II			RS-I			COMBINED				
VARIANT			n= 2,660			n= 5,334			n= 938			n= 1,594			n= 750			n= 2,436			n=13,712***				
SNP	A1	R2*	Freq.	BETA**	P	Freq	BETA**	P	Freq	BETA**	P	Freq	BETA**	P	Freq	BETA**	P	Freq	BETA**	P	BETA**	P	I2	HetP
rs917727	T	1	0.296	0.21	**4.31E-11**	0.273	0.133	**7.43E-09**	0.238	0.175	0.002	0.274	0.122	0.002	0.263	0.208	5.34E-04	0.269	0.133	5.6E-05	0.154	**1.28E-27**	17	0.306
rs2908004	A	0.55	0.501	0.16	**1.23E-08**	0.442	0.122	**1.32E-09**	0.431	0.14	0.004	0.456	0.111	0.002	0.437	0.233	1.01E-05	0.449	0.124	5.01E-05	0.136	**1.66E-27**	0	0.416
rs917726	T	1	0.282	0.208	**6.42E-11**	0.273	0.134	**3.53E-09**	0.238	0.175	0.002	0.274	0.122	0.003	0.263	0.208	5.35E-04	0.269	0.133	6.39E-05	0.154	**1.80E-27**	11	0.347
rs718766	C	1	0.272	0.208	**8.63E-11**	0.273	0.134	**3.55E-09**	0.238	0.177	0.002	0.274	0.122	0.003	0.263	0.209	5.40E-04	0.269	0.134	6.60E-05	0.154	**2.17E-27**	11	0.347
rs3801382	G	1	0.275	0.199	**1.04E-10**	0.273	0.129	**3.37E-09**	0.238	0.171	0.002	0.274	0.12	0.003	0.263	0.206	5.22E-04	0.269	0.132	6.28E-05	0.149	**2.24E-27**	6	0.378
rs7776725	*C*	*1*	*0.264*	*0.214*	***4.58E-11***	*0.273*	*0.136*	***3.65E-09***	*0.239*	*0.187*	*0.001*	*0.274*	*0.123*	*0.003*	*0.263*	*0.209*	*5.49E-04*	*0.269*	*0.134*	*6.71E-05*	*0.156*	***2.36E-27***	*16*	*0.312*
rs2536189	*G*	*0.55*	*0.498*	*0.153*	***2.23E-08***	*0.442*	*0.122*	***1.32E-09***	*0.431*	*0.14*	*0.004*	*0.456*	*0.111*	*0.002*	*0.437*	*0.232*	*1.02E-05*	*0.449*	*0.124*	*5.04E-05*	*0.135*	***3.06E-27***	*0*	*0.444*
rs3801387	G	1	0.274	0.197	**1.61E-10**	0.272	0.129	**3.34E-09**	0.238	0.166	0.002	0.275	0.121	0.002	0.263	0.205	5.71E-04	0.270	0.131	7.09E-05	0.148	**4.32E-27**	0	0.415
rs4727924	T	0.51	0.467	0.172	**3.88E-09**	0.459	0.118	**7.05E-09**	0.454	0.144	0.003	0.477	0.112	0.002	0.463	0.218	2.57E-05	0.476	0.103	4.65E-04	0.132	**5.25E-26**	22	0.267
rs2536182	G	0.53	0.471	0.156	**2.69E-08**	0.454	0.116	**3.05E-09**	0.439	0.151	0.002	0.466	0.117	0.001	0.452	0.223	1.43E-05	0.463	0.098	8.54E-04	0.129	**1.28E-25**	20	0.286
rs2536180	C	0.51	0.495	0.143	1.57E-07	0.463	0.113	**5.66E-09**	0.456	0.137	0.003	0.478	0.11	0.002	0.465	0.215	2.30E-05	0.477	0.102	4.57E-04	0.124	**4.21E-25**	0	0.438
rs2707466	*T*	*0.51*	*0.485*	*0.152*	***3.55E-08***	*0.425*	*0.118*	***1.11E-08***	*0.418*	*0.142*	*0.004*	*0.440*	*0.111*	*0.003*	*0.424*	*0.237*	*1.03E-05*	*0.435*	*0.119*	*1.33E-04*	*0.133*	***5.13E-25***	*5*	*0.386*
rs2254595	C	0.51	0.504	0.142	1.20E-07	0.463	0.113	**5.92E-09**	0.456	0.137	0.003	0.478	0.11	0.002	0.464	0.214	2.45E-05	0.477	0.102	4.48E-04	0.124	**5.24E-25**	0	0.453
rs3779381	G	0.87	0.264	0.18	**1.57E-08**	0.255	0.123	9.95E-08	0.227	0.155	0.006	0.254	0.123	0.003	0.241	0.209	8.52E-04	0.250	0.137	7.81E-05	0.143	**2.00E-23**	0	0.623
rs2536150	C	0.07	0.215	-0.136	6.22E-05	0.176	-0.096	1.57E-04	0.168	-0.13	0.037	0.180	-0.011	0.812	0.174	-0.206	0.003	0.182	-0.071	0.059	-0.099	**3.17E-10**	37	0.162
rs2952559	C	0.06	0.265	-0.133	6.34E-05	0.183	-0.105	4.24E-05	0.165	-0.114	0.077	0.180	-0.034	0.475	0.184	-0.198	0.005	0.184	-0.050	0.180	-0.099	**3.94E-10**	26	0.240
rs13247600	C	0.04	0.064	-0.219	7.07E-04	0.076	-0.117	7.09E-03	0.061	-0.18	0.061	0.082	-0.085	0.244	0.084	-0.266	0.022	0.079	-0.219	3.17E-04	-0.163	**1.31E-09**	0	0.462
rs2707520	C	0.25	0.444	-0.091	0.001	0.502	-0.06	0.003	0.544	-0.04	0.416	0.475	-0.066	0.070	0.532	-0.098	0.065	0.517	-0.090	0.003	-0.073	**6.90E-09**	0	0.865
rs17509082	T	0.39	0.173	0.119	0.001	0.198	0.065	0.007	0.182	0.041	0.492	0.202	0.067	0.133	0.187	0.170	0.008	0.201	0.090	0.010	0.084	**3.21E-08**	0	0.536
rs2908007	A	0.37	0.492	-0.1	4.12E-04	0.603	-0.046	0.027	0.628	-0.062	0.216	0.613	-0.067	0.070	0.610	-0.030	0.569	0.608	-0.100	0.001	-0.070	**4.78E-08**	0	0.528

Bolded rs917727 top-hit, rs3801382 used for conditioning and rs3801387 genotyped in Generation R Mothers * Correlation coefficients with rs917727 based on HapMap release22 CEU population. **Effect estimates expressed as standardized adjusted SD per copy of allele (A1) ***Results do not include Generation R mothers only genotyped for rs3801387. Underline: rs7776725 Top-hit wrist fracture GWAS, rs2536189 Top-hit forearm BMD GWAS, rs2707466 Top-hit for Cortical thickness in Zheng et al (19).

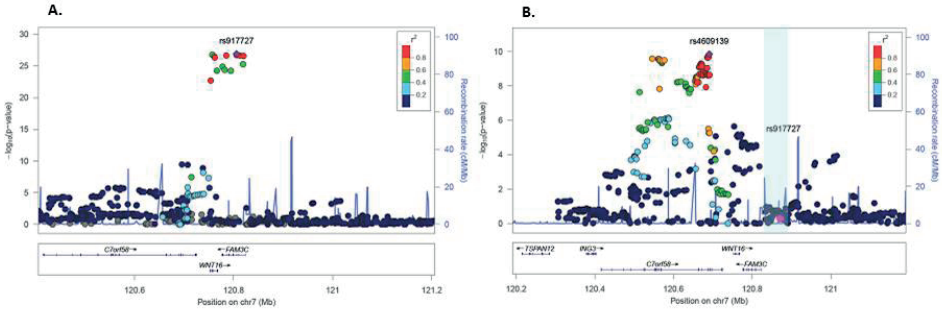


Figure 3. Association plots for TB-BMD. A. SNP association plot for TB-BMD-associated region of Chromosome 7q31.31. B. SNP association plot for TB-BMD-associated region of Chromosome 7q31.31 after conditioning on rs3801382. Genetic coordinates are as per HapMap phase II-CEU. *Data from the mothers of Generation R is not included.

LD with each other (r^2: 0.80-1.00, based on the Hapmap phase II-CEU panel), displayed moderate heterogeneity across cohorts (I^2 between 11 and 47%) and had smaller effect sizes (standardized SD -0.067 to -0.081) as compared to those observed for the main signal. The lead SNP rs4609139 associated at $P=1.42x10^{-10}$ (Figure S1) had an average MAF of 0.35 across studies and a combined effect of -0.08 BMD standard deviations (SE:0.0126) per copy of the minor allele, explaining on average 0.2% of the phenotypic variance in standardized BMD.

Evaluation of covariates

In the discovery cohort we observed prominent effects of the covariates on the SNP-phenotype relationships. To illustrate this we present beta estimates as standardized coefficients from null-intercept centered models in Table S4 (T-S4). As expected, lack of correction for principal components of the sex- and age- adjusted model (Model 0 in T-S4 and Figure 1A) generated important inflation of the test statistic, which is severely reduced by inclusion of twenty principal components (Model 1 in T-S4 and Figure 1A). Inclusion of weight in the sex-, age- and PCs- corrected model (Final Model in T-S4 and Figure 1A) increased the significance of the putative signal by reducing the standard error of the SNP effect estimate (from 0.030 to 0.022). Such reduction of the error variance led to genome-wide significance after weight correction. Weight is an important determinant of peak bone mass accrual related to both loading and size effects as illustrated by the positive relationship with total body BMD also evident across our SNP-Phenotype models (Table S4).

Association with skull BMD

The impact of weight correction on the standard errors of the association led us to hypothesize that, at least in children, the effect of variants in the *WNT16* region on total-body BMD was independent of skeletal loading (of which body weight is a proxy). For this reason,

we analyzed the 721 SNPs of the same genomic region in relation to skull BMD across all five cohorts with GWAS data and total body BMD (without Generation R mothers). Skull BMD measured by the total body DXA scan constitutes an independent measurement in children since the head region is excluded from the total body BMD assessment. This is done given the large variation in density and area inflicted by the skull on the head region, which is particularly evident in paediatric populations [13]. The head DXA region is suitable to evaluate the relationships with skeletal loading considering that its direct influence on the mineralization process of the skull is negligible. Despite being a skeletal region composed of laminar bones, the proportion of mostly cortical (95% for the inner and outer layers) but also trabecular (inner lamella) bone in the skull is similarly high to that of the overall skeleton (80% cortical) as compared to other skeletal sites. In addition, the BMD of the skull is subject to the same patterns of peak bone mass accrual and decrease with aging observed across the lifetime (Figure 4A). Since, as expected, weight was not a significant covariate in the analysis of skull BMD (P=0.09) we performed the meta-analysis using a sex-, age-, height- and PCs-corrected model across the five cohorts with GWAS data. The strongest association signal with skull BMD mapped to *C7orf58* (the gene underlying the secondary signal of TB-BMD). The most significantly associated SNP was rs7801723 with MAF between 0.33 and 0.39 across cohorts and in moderate LD (r^2=0.56) with rs4609139 (secondary signal in TB-BMD). The combined effect of rs7801723 was -0.14 BMD standard deviations (SE:0.012) per copy of the minor allele (P=8.9x10^{-28}) showing a high heterogeneity I^2 of 60.7% and P_{het}=0.03 (Figure 4B). Moreover, we identified 147 variants in the 7q31.31 locus achieving genome-wide significance (Table S5) and suggesting the existence of two independent signals (Figure 5) in the regional meta-analysis of skull BMD. The rs917727

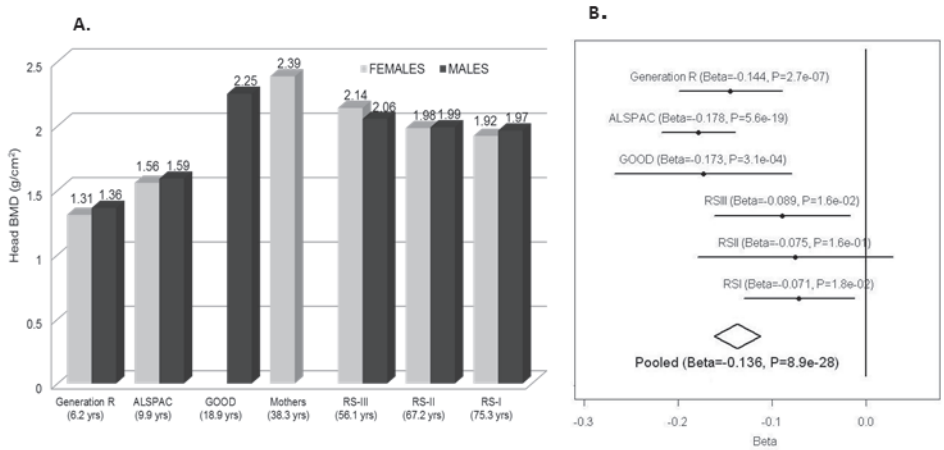


Figure 4. Skull-BMD across cohorts and meta-analysis A. Gender-specific mean Skull BMD for each cohort. B. Forrest plot of the association of skull BMD with rs7801723. The results are reported per copy of the T-allele (MAF=0.37).

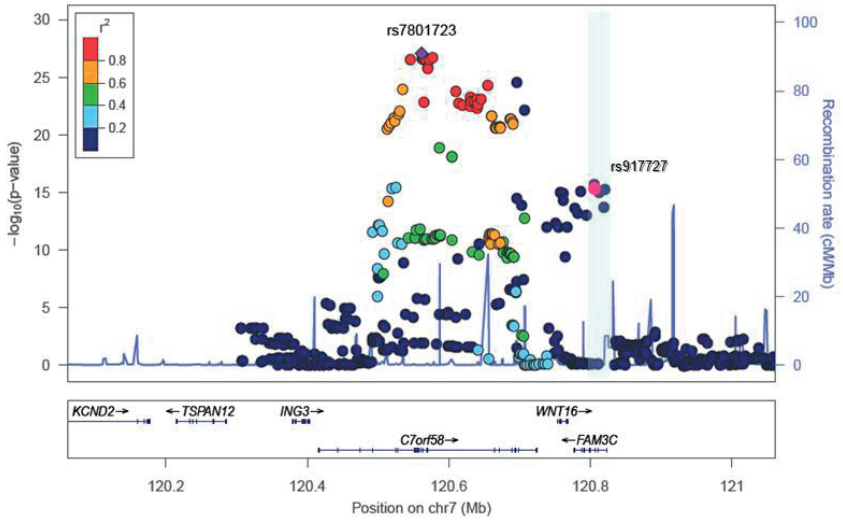


Figure 5. Association plots for skull-BMD. SNP association plot for the skull BMD-associated region in chromosome 7q31.31, based on 13,712 individuals from the five different cohorts with GWAS information.

SNP (primary signal of total body BMD) was also associated at genome-wide significant level ($P=1.9x10^{-16}$) with skull BMD with an effect estimate of 0.12 BMD standard deviations (SE:0.014) per copy of the minor allele and no evidence for significant heterogeneity ($I^2=0\%$).

The heterogeneity at rs7801723 appeared to be driven by different effects in younger and older populations (Figure 4B). For this reason, we stratified the analysis according to whether individuals within the cohorts had achieved total skeletal maturation (RS-I, RS-II, RS-III) or were still in the process of peak bone accrual (GOOD, ALSPAC, GEN-R). The rs7801723 signal seemed to be strongest in the younger populations ($B=-0.16$; $P=2.06x10^{-27}$) since the effect was considerably weaker in older populations ($B=-0.08$; $P=2.7x10^{-4}$), though differences in power due to lower sample size may also play a role (Figure S2). Meta-regression across studies showed a significant relation between mean age and absolute effect sizes observed for rs7801723 (Figure S3) on skull BMD ($Beta_{age}$ meta-regression= 0.0015; $P=0.006$) but not on total BMD ($Beta_{age}$ meta-regression= 0.0008; $P=0.12$). In contrast, the effect of rs917727 on skull BMD seems to be uniform across older ($B=0.14$; $P=4.16x10^{-9}$) and younger ($B=0.10$; $P=4.28x10^{-9}$) populations, displaying no evidence of effect heterogeneity nor a significant relation with the mean age across studies ($Beta_{age}$ meta-regression= -0.0006; $P=0.19$). Similarly, the effect size of rs917727 on total body BMD was not related to the mean age of the studies ($Beta_{age}$ meta-regression= -0.0004; $P=0.57$).

Functional evaluation of the 7q31.31 locus

The effect of variants from the 7q31.31 locus on both total body and skull BMD cannot be unequivocally attributed to any of the closest three genes in the GWAS signal region

(*WNT16*, *FAM3C* and *C7orf58*). This is also complicated by the high LD across the region. Based on current knowledge, *WNT16* is the best candidate at the locus considering that it belongs to the Wnt family of proteins. The Wnt signaling pathway plays ubiquitous key roles in fundamental biological processes, including those critical for bone biology and specifically for bone formation [14,15]. *FAM3C* is a widely expressed gene (including in osteoblasts) which belongs to a cytokine-like gene family without homology to any known cytokines [16]. Minimal information exits about the functional aspects of *C7orf58*. Considering the hypothesis-free nature of our GWAS approach we cannot exclude the possibility that any of these genes may code for proteins involved in BMD regulation.

Wnt16 and Fam3c KO mouse models

Further evidence implicating *WNT16* as the gene underlying these associations with total body BMD at the population level is provided by functional studies on *Wnt16* knockout (KO) mice generated at Lexicon Pharmaceuticals (Table 2). These KO mice have reduced total body BMD at 24 weeks of age, resulting from both reduced total body bone mineral content (BMC) and bone area. BMC and aBMD measured at the spine (a skeletal site more strongly influenced by trabecular bone than total body or femur measurements) were slightly reduced in KO mice but this reduction did not achieve statistical significance. Male and female knockout mice appeared healthy with no discernible morphological or growth defects, and normal femur length, body weight and body composition. We also examined mice from three *Fam3c* KO models testing for differences across DXA phenotypes with

Table 2. Summary statistics for densitometric properties of control (+/+) and *Wnt16* deficient (-/-) mice

Parameter	Male WT Mice	Male *Wnt16* KO Mice	Statistics	Female WT Mice	Female *Wnt16* KO Mice	Statistics
Number of mice	9	12		24	16	
Body Weight (grs)	36.4 ± 1.8	38.5 ± 1.2	Δ = ↑5%, P = 0.42	28.2 ± 0.8	26.8 ± 1.2	Δ = ↓5%, P = 0.30
Lean Body Mass (grs)	27.0± 0.9	27.9 ± 0.7	Δ = ↑4%, P = 0.42	20.5 ± 0.4	19.6 ± 0.6	Δ = ↓4%, P = 0.20
Body Fat (percent)	23.1 ± 2.1	24.9 ± 1.4	Δ =↑8%, P = 0.48	24.2 ± 1.2	23.2 ± 1.6	Δ = ↓4%, P = 0.63
Femur Length (mm)	16.3 ± 0.2	16.2 ± 0.2	Δ = ↑1%, P = 0.58	16.2 ± 0.1	16.1 ± 0.1	Δ = 0%, P = 0.86
Body aBMD (mg/cm^2)	56.9 ± 1.1	54.8 ± 0.7	Δ = ↓4%, P = 0.11	53.9 ± 0.6	48.7 ± 0.6	**Δ = ↓10%, P < 0.001**
Body Bone Area (cm^2)	9.4 ± 0.2	8.6 ± 0.3	**Δ = ↓9%, P = 0.03**	8.9 ± 0.2	8.1 ± 0.1	**Δ = ↓8%, P = 0.002**
Body BMC (mg)	532 ± 16	470 ± 18	**Δ = ↓12%, P = 0.02**	479 ± 11	396 ± 10	**Δ = ↓17%, P < 0.001**
Femur aBMD (mg/cm^2)	89.4 ± 3.3	83.6 ± 1.5	Δ = ↓7%, P = 0.10	78.5 ± 1.3	62.8 ± 1.2	**Δ = ↓20%, P < 0.001**
Femur Bone Area (cm^2)	0.36 ± 0.01	0.34 ± 0.01	Δ =↓7%, P = 0.13	0.35 ± 0.01	0.31 ± 0.01	**Δ = ↓13%, P < 0.001**
Femur BMC (mg)	32.6 ± 1.8	28.0 ± 1.0	**Δ = ↓17%, P = 0.03**	27.8 ± 0.6	19.5 ± 0.7	**Δ = ↓30%, P < 0.001**
Spine aBMD (mg/cm^2)	61.9 ± 1.5	58.8 ± 3.7	Δ = ↓5%, P = 0.50	62.8 ± 1.8	58.2 ± 1.6	Δ = ↓7%, P = 0.08
Spine BMC (mg)	25.8 ± 0.9	23.2 ± 2.0	Δ =↓10%, P = 0.29	26.3 ± 0.9	24.3 ± 0.9	Δ = ↓8%, P = 0.16

Results provided as [mean +/- SEM]

wild type animals. We failed to observe any significant differences across the skeletal phenotypes in each independent *Fam3c* KO strategy (Table S6). Even though these data suggest *Fam3c* does not influence bone mass in mice, the possibility of a false negative due to power limitations cannot be excluded.

Gene transcript - phenotype correlations

We examined the correlation of gene expression transcript levels derived from iliac bone crest biopsies in relation to BMD levels in a distinct cohort of 78 unrelated-Norwegian women with total body scans (of which 51% have osteoporosis) and who are part of a set (n=84) described in detail previously [17]. The investigated region comprised +/-500 Kb of rs917727 and contained seven different genes including *TSPAN12* (3 transcripts), *ING3* (4 transcripts), *C7orf58* (2 transcripts), *WNT16* (2 transcripts), *FAM3C* (3 transcripts), *PTPRZ1* (1 transcript) and one represented by the Affymetrix probe with ID *217206_at* lacking annotation. We only identified significant correlations with BMD measurements of the donors in transcripts from *C7orf58* and W*NT16* (Table 3). Expression levels in one of the transcripts in *WNT16* (224022_x_at) was significantly associated with BMD measured at several skeletal sites including the total body, skull, legs, total hip and lumbar spine (L1-L4 vertebrae). The correlation was positive and of similar magnitude across sites, ranging between 0.25 and 0.31, indicating that higher expression of this gene is correlated with higher BMD, results which are in line with the *Wnt16* KO mice data. Significant correlation with total body lean mass was also observed (r^2=0.31) for this transcript. This suggests a pleiotropic effect on muscle considering that correction for BMI did not importantly influence the correlation of the transcript with total body BMD. There is even a stronger (inverse) correlation of total body and skull BMD with expression levels of one of the *C7orf58* transcripts (228728_at), even approaching a correlation of -0.50 with total body BMD. This suggests high *C7orf58* expression levels of this transcript are related to lower BMD. This transcript was inversely correlated with body weight as well, but still maintained a strong (inverse) correlation (r^2=-0.45) with skull BMD (which as discussed above, is not readily affected by body weight) and also with BMI-adjusted total body BMD (r^2=-0.42). In addition, levels of both *C7orf58* transcripts were significantly correlated with age, which further supports the age-specific effects seen for the GWAS variants mapping to *C7orf58*.

DISCUSSION

In this study we carried out a genome-wide association scan of total body BMD in 2,660 children with regional replication of the top hit in 12,066 children and adults from six additional studies, including *de novo* genotyping of a sample of 1,014 Northwestern European pre-menopausal women, mothers of the children from the discovery cohort. We identified

Table 3. Expression analysis of *C7orf58, WNT16* and *FAM3C* transcripts in hip bone biopsies from 78 Norwegian women

Gene symbol Affymetrix transcript ID Exons covered by probeset	Mean (SD)	*C7orf58* 220032_at 16-19		*C7orf58* 228728_at 24		*WNT16* 221113_s_at 4-5		*WNT16* 224022_x_at 5		*FAM3C* 240062_at 1a		*FAM3C* 236316_at 6-7		*FAM3C* 201889_at 10	
		Pearson r	P	Pearson r	P	Pearson r	P	Pearson r	P	Pearson r	P	Pearson r	P	Pearson r	P
Age (years)	64.2 (9.6)	**0.24**	**0.032**	**0.24**	**0.031**	0.04	0.720	-0.05	0.680	0.14	0.230	0.04	0.710	0.16	0.170
Weight (kg)	66.0 (12.0)	**-0.26**	**0.022**	0.02	0.860	-0.13	0.250	0.19	0.089	-0.05	0.640	-0.07	0.530	-0.05	0.650
BMI (kg/cm2)	23.9 (3.5)	**-0.26**	**0.021**	0.04	0.760	-0.16	0.170	0.13	0.280	-0.02	0.880	0.00	0.980	-0.03	0.760
Height (cm)	165.9 (7.6)	-0.12	0.320	-0.04	0.740	-0.01	0.940	0.18	0.110	-0.09	0.450	-0.14	0.220	-0.08	0.460
Total body BMD (g/cm^2)	1.06 (0.15)	**-0.49**	**5.00E-06**	-0.16	0.150	-0.20	0.085	**0.29**	**0.011**	-0.03	0.810	-0.02	0.860	-0.20	0.083
Total body T-score	-0.81 (1.83)	**-0.49**	**5.20E-06**	-0.16	0.150	-0.19	0.090	**0.29**	**0.010**	-0.03	0.780	-0.02	0.850	-0.19	0.091
Total body Z-score	0.19 (1.52)	**-0.44**	**6.20E-05**	-0.13	0.270	-0.17	0.150	**0.26**	**0.020**	0.02	0.870	0.00	0.990	-0.17	0.140
Total body Z-score, BMI adj	-0.19 (1.51)	**-0.42**	**1.20E-04**	-0.13	0.250	-0.16	0.170	**0.26**	**0.024**	0.02	0.870	0.00	1.000	-0.17	0.150
Head BMD (g/cm^2)	2.07 (0.42)	**-0.46**	**2.90E-05**	-0.09	0.440	**-0.24**	**0.037**	**0.28**	**0.012**	-0.07	0.560	-0.11	0.320	-0.21	0.071
Arms BMD (g/cm^2)	0.80 (0.12)	**-0.37**	**7.90E-04**	-0.18	0.110	-0.21	0.070	0.19	0.089	-0.06	0.630	0.02	0.880	-0.17	0.140
Legs BMD (g/cm^2)	1.13 (0.18)	**-0.47**	**1.70E-05**	-0.19	0.087	-0.14	0.210	**0.27**	**0.016**	-0.03	0.820	0.01	0.910	-0.17	0.140
Total Hip BMD (g/cm^2)	0.88 (0.19)	**-0.42**	**1.90E-04**	-0.18	0.110	-0.16	0.160	**0.29**	**0.012**	-0.01	0.940	0.06	0.630	-0.11	0.350
Total Hip T-score	-0.99 (1.61)	**-0.43**	**9.70E-05**	-0.18	0.120	-0.14	0.210	**0.28**	**0.014**	0.00	0.990	0.07	0.550	-0.12	0.310
Total Hip Z-score	0.03 (1.39)	**-0.36**	**1.10E-03**	-0.13	0.270	-0.12	0.300	**0.25**	**0.025**	0.05	0.660	0.10	0.370	-0.08	0.490
Total Hip Z-score, BMI adj	-0.19 (1.35)	**-0.31**	**5.50E-03**	-0.14	0.230	-0.09	0.460	**0.23**	**0.040**	0.06	0.630	0.11	0.360	-0.07	0.520
L1-L4 BMD (g/cm^2)	0.97 (0.25)	**-0.45**	**3.20E-05**	-0.06	0.630	-0.22	0.054	**0.31**	**0.006**	0.00	0.970	-0.05	0.660	-0.21	0.068
L1-L4 T-score	-1.68 (2.01)	**-0.44**	**7.00E-05**	-0.05	0.680	-0.20	0.082	**0.30**	**0.007**	0.03	0.820	-0.03	0.770	-0.20	0.074
L1-L4 Z-score	-0.47 (1.81)	**-0.38**	**7.20E-04**	0.01	0.950	-0.18	0.110	**0.29**	**0.011**	0.06	0.600	-0.02	0.860	-0.18	0.110
L1-L4 Z-score, BMI adj	-0.29 (1.78)	**-0.34**	**2.60E-03**	0.00	0.990	-0.16	0.170	**0.27**	**0.017**	0.06	0.580	-0.02	0.850	-0.18	0.120
Total body Lean mass (kg)	40.4 (4.8)	-0.22	0.055	0.04	0.720	0.04	0.710	**0.31**	**0.005**	-0.03	0.780	-0.13	0.270	-0.02	0.830
Total body Fat mass (kg)	23.0 (8.4)	-0.22	0.058	0.00	0.970	-0.20	0.075	0.08	0.510	-0.07	0.560	-0.03	0.800	-0.07	0.560

at least two independent signals (primary rs917727 and secondary rs4609139) associated with total body BMD mapping to the 7q31.31 locus harboring (among others) two genes and an open reading frame sequence including *WNT16, FAM3C* and *C7orf58*. To examine whether the observed associations were independent of skeletal loading we tested variants in the 7q31.31 region for association with skull BMD, a non-weight bearing skeletal site. We found the rs917727 influencing skull BMD and an even stronger signal arising from rs7801723 (in partial LD with rs4609139) which unlike the total body signals was largely driven by the younger (children) populations. A DXA-based assessment of the *Wnt16* KO mice showed a lower total body BMD phenotype which is compatible with an effect of *WNT16* variants playing a role on total body BMD at the human population level. Moreover, the analysis of gene transcript expression profiles supports the involvement of variants from both *WNT16* and *C7orf58*. Together, these findings postulate that the *WNT16/C7orf58* locus contains complex patterns of genetic variation, which play an important role in peak bone mass accrual and may likely impact BMD determination at later life.

Genetic variants in the region mapping to *FAM3C* as the closest gene, have been previously reported in the literature as associated with speed of sound as analyzed by quantitative ultrasound in radius and calcaneous in a Korean population [18] and in a recent follow-up study the same SNPs were genotyped *de novo* and were found to be associated with bone mineral density at different sites in individuals of European descent [19]. Nonetheless, to our knowledge this is the first time this locus is associated with total body BMD in children. Total body BMD measured in children corresponds only to the amount of bone accrued up to that point in time; consequently the variants described here have a definite role in the genetic determination of bone acquisition. Furthermore, considering the previously reported association in East Asians, the multiethnic background of our discovery population and the lack of heterogeneity in the total-body signal, one can conclude the effect of these variants is present across populations of different genetic background. To assess this in more detail we examined in the multiethnic Generation R discovery cohort the top associated SNPs of the meta-analysis across clusters of different ethnic backgrounds (Table S7). We confirmed that the effect of the top associated variants was not confined to individuals of European descent. Effect directions and magnitude of the markers were largely similar across groups, despite some evident differences in linkage disequilibrium patterns between markers.

Without additional functional evaluation it is not feasible to undisputedly distinguish which genes in the 7q31.31 region could be underlying the observed GWAS signals. The analysis of gene expression profiles provides supporting evidence for the involvement of variants from *WNT16* and *C7orf58,* while not for *FAM3C* in relation to total body BMD and skull BMD. Similarly, the lower likelihood of an effect arising from *FAM3C* is supported by the absence of an abnormal skeletal phenotype in the *Fam3c* KO mice, although we cannot

exclude the possibility that variants resulting in gain of function of *Fam3c* affect bone. In contrast, functional evidence of the involvement of *WNT16* is well supported by the analysis of expression profiles and the observation of reduced BMD in the *Wnt16* KO mice. The *WNT16* human/mouse sequence alignment shows 93% identity favoring the plausibility of similarity in phenotypic effects across both species. Even though the *Wnt16* KO mice had reductions of both BMC (strongest) and bone area, in humans we found no indication of an effect on bone size (area) in either children or adults. This is consistent with the lack of any observed association between common genetic variation in this region and adult body height as assessed in a sufficiently powered GWAS [20]. We can infer a prominent effect of *WNT16* on cortical bone in humans considering the close resemblance of the cortical phenotype observed in the *Wnt16* KO mice to that reported in humans by Zheng et al. based on pQCT (accompanying submission). This prominent genetic effect on cortical bone is also manifested by the associations of *WNT16* variants with BMD traits measured at skeletal sites rich in cortical bone including the total body, the skull and the forearm as also shown by Zheng et al. (accompanying submission). In addition, Zheng et al. show that this effect on cortical bone influences bone strength in mice and fracture risk in humans. Nevertheless, the KO mouse data should be interpreted with caution since our functional validation sought the confirmation of the KO strategy but did not assess the integrity of the surrounding genomic region (i.e. intact *FAM3C* and/or *C7orf58* function). Further, total body BMD also involves components of trabecular bone which, together with the associations previously reported with ultrasound of the heel [18] and lumbar spine [19] (sites of rich trabecular content), do not fully exclude an effect of *WNT16* (or *FAM3C*) variants on this type of bone.

Both the *secondary signal* unveiled by our total body BMD GWAS and the *strongest signal* in the skull BMD GWAS analyses map to an open reading frame sequence (*C7orf58*) in the region. The effect size for rs4609139 on total body BMD (mapping to *C7orf58*) was up to 50% lower in magnitude than that of rs917727 (mapping in the vicinity of *WNT16*). In contrast, the associations with skull BMD were stronger for variants from the *C7orf58* signal than for those arising from *WNT16*. Further study of the mechanisms by which *C7orf58* exerts its role either independently or together with *WNT16* is warranted.

Several aspects derived from the analysis of skull BMD merit further discussion. Skull BMD is minimally influenced by loading, muscular activity, and in general less masked by environmental influences. From this perspective, fine-tuned mechanosensing mechanisms involved in the regulation of bone metabolism can be better dissected examining skull BMD. The skull BMD associations arising from the *C7orf58* signal are substantially more prominent in the younger populations. Such age dependency was not seen for the signals arising from *WNT16,* which also shows effects already evident at young age, but that do

persist through adulthood until very old age, as corroborated by the replication studies in the older population cohorts. Also, from the strong associations with skull BMD we can infer that these effects likely to be arising from *WNT16* and *C7orf58* are not mediated by a mechanosensing response to skeletal loading. A recent study examining expression patterns at different skeletal sites in rats has shown differential gene expression patterns between the skull, the limbs and the total body, which likely reflects different responses to loading and mechanosensing between skeletal sites [21]. In the latter study, *WNT16* showed opposite expression patterns in arms than in skull. While in rats only one isoform is active, in humans different WNT16 isoforms could play distinct regulatory roles on skeletal development and/or metabolism [22]. Alternatively, modulation by other signaling factors could also modify the (up-/down-) streaming effects of *WNT16*. Recently, it has been shown that WNT16 influences hematopoietic stem cell differentiation via non-canonical Wnt signaling in zebra fish [23]. Whether this is also the case in bone biology remains to be confirmed, since its effect through canonical (beta-catenin mediated) activation has already been established in cartilage [24].

The allelic heterogeneity demonstrated by the conditional analysis on total body BMD is indicative of multiple (at least 2) causal variants in the region within and across phenotypes. The associations with TB-BMD observed for the top hits reported in the accompanying submission by Zheng et al. (namely rs2707466 for pQCT, rs2536189 for forearm BMD and rs7776725 for wrist fracture) are affected differently by conditional analysis on the top signal. As predicted by the complete linkage disequilibrium between rs917727 (top total body BMD SNP) and rs777625 (top wrist fracture SNP associated with TB-BMD with B=0.156; P=2.36×10^{-27}), after conditioning, the effect of rs777625 on TB-BMD is largely gone (B=-0.0013; P=0.93). In contrast, the effects on TB-BMD of rs2536189 (forearm BMD from B=0.135 P=3.06×10^{-27} to B=0.042; P=6.7×10^{-4}) and rs2707466 (pQCT from B=0.133 P=5.13×10^{-25} to B=0.043 P=6.8×10^{-4}) are not completely explained by their moderate LD (with r^2 values between 0.51 and 0.55) with the top associated SNP of the TB-BMD GWAS signal.

In addition to the top associated pQCT SNP (rs2707466), there is yet another non-synonymous variant (rs2908004) annotated within the coding region of *WNT16* which is genome wide significant in our meta-analysis of total body BMD. After conditional analysis the effect sizes of these non-synonymous SNPs show considerable reduction suggesting there are not likely causal to the stronger top GWAS signal. Nevertheless, the residual effect is still significant after conditioning (indicating independence and) implying a weaker effect on TB-BMD or (more likely) partial linkage disequilibrium with yet other causal variants. Such relationships between genetic variants can follow diverse types of complex relationships as recently described for loci displaying allelic heterogeneity [25] for which follow-up investigations are warranted to elucidate the definitive involvement of the genes in this 7q31.31 region.

Finally, this region harbors one or more genes revealing critical effects on bone biology. While we have shown that genetic variants in this locus influence total body BMD variation in children of multiple ethnic background, the relevance of our findings are manifested by the persistence and consistency of the associations observed later in life. Further, the prominent effect on total body BMD we describe agrees with associations observed in adults across diverse skeletal traits [18,19] (see also Zheng et al. accompanying submission), but most importantly, by their effect on risk of fracture (see also Zheng et al. accompanying submission), the most deleterious consequence of osteoporosis.

In summary, this study detected at least two independent GWAS signals influencing total body and skull BMD variation in children, thus confirming the presence of allelic heterogeneity in this *WNT16* locus. In addition, we showed how the effects observed in children are consistently replicated in adults. Specific genetic determination of peak bone mass (rather than bone loss later in life) is suggested by more prominent effects of some markers in children than in adults. These genetic effects likely influence the attainment of peak bone mass accrual and impact the risk of osteoporosis and fracture later in life.

METHODS

Ethics Statement

All research aims and the specific measurements in the participating studies involving human beings have been approved by the correspondent Medical Ethical Committee. Written informed consent was provided by all subjects or their parents in the case of children. Mouse studies were performed in accordance with institutional and regulatory guidelines for animal care and use at Lexicon Pharmaceuticals.

Subjects

Generation R Study

The Generation R Study is a prospective cohort study in which 9,778 pregnant women living in Rotterdam and with delivery date from April 2002 until January 2006 were enrolled. Details of study design and data collection can be found elsewhere [26]. The current study comprised 2,660 children (mean age 6.16, SD=0.39 years), of which 1,511 are of Dutch Northern European origin, who had both GWAS and DXA-based BMD measurements. DXA measurements were recorded on children visiting a unique research centre at around 5 years old accompanied by their mothers. All research aims and the specific measurements in the Generation R Study have been approved by the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam and written informed consent was provided by all parents.

Avon Longitudinal Study of Parents and their Children (ALSPAC)

In-silico replication of the GWAS signals was initially pursued in The Avon Longitudinal Study of Parents and their Children (ALSPAC). This is a longitudinal population-based birth cohort that recruited pregnant women residing in Avon, UK, with an expected delivery date between 1st April 1991 and 31st December 1992. This cohort is described in detail on the website (http://www.alspac.bris.ac.uk) and elsewhere [27,28]. Total body BMD and genome-wide SNP data were available for 5,334 unrelated children (mean age=9.9, SD=0.32 years) all of Northern-European descent. Ethical approval was obtained from the ALSPAC Law and Ethics committee and relevant local ethics committees, and written informed consent was provided by all parents.

The Gothenburg Osteoporosis and Obesity Determinants (GOOD)

The GOOD Study is a population-based cohort in which male subjects from between 18 and 20 years of age in the Gothenburg area in Sweden were randomly selected using national population registers and invited to participate in this initiative by phone. From the selected candidates 1,068 agreed to participate providing oral and written informed consent [29,30]. The GOOD Study was approved by the local ethics committee at Gothenburg University. A subset of 938 individuals from this study with DXA measurements and GWAS data were included in this analysis.

Rotterdam Study (RS I, II & III)

Additional in-silico replication in elderly adults was pursued in participants of the Rotterdam Study, a large prospective population-based cohort study of white subjects aged 45 years and older living in the Ommoord District of Rotterdam, The Netherlands who are studied for the occurrence of chronic diseases and disability [31]. Subjects were derived from the three different cohorts of the Rotterdam Study including RS-I (n=2,436), RS-II (n=750) and RS-III (n=1,594) comprising individuals of Northwestern European Ancestry with available BMD-DXA measurements and GWAS data. Approval of the Medical Ethics Committee of the Erasmus University Rotterdam was obtained for the three cohorts of the Rotterdam Study. From all participants written informed consent was acquired.

Bone mineral density measurements

Total body and Head BMD were measured in all participants using dual-energy X-ray absorptiometry (DXA) following standard manufacturer protocols. GE-Lunar iDXA was the devise used in the Generation R Study while the other cohorts employed GE Lunar (GE-Lunar Prodigy; GE Healthcare, Chalfont St Giles, UK). Bone mineral content (BMC) was derived from the projected bone area (BA) as BMC (mg) = BMD g/cm^2 X BA (cm^2). As recommended by the International Society for Clinical Densitometry total body less head (TBLH) was the measurement used in the Generation R Study and ALSPAC instead of total body BMD [2].

Genotype assessment

Genotyping was performed using the Illumina HumanHap 610 QUAD microarray in The Generation R, GOOD and RS-III cohorts while Illumina HumanHap 550 was the platform used for ALSPAC, RS-I and RS-II cohorts. Stringent quality control of the genotype and imputation process was performed in each study (Table S8). Samples with gender discrepancy, excess of heterozygosity or duplicates were excluded from analysis. *De novo* genotyping for the Generation R mothers was performed as part of a GEFOS initiative at Kbiosciences for specific SNPs of interest in the Osteoporosis field among those rs3801387.

Imputation

For the imputation in the discovery Generation R cohort, we built a panel of reference haplotypes using HapMap phase II (release 22, build 36) CEU, YRI and CHB/JPT data. A two-step imputation process was performed, haplotype phasing and genotype imputation were carried out using MACH and minimac software, respectively. Imputation of the replication cohorts was done using MACH v1 based on the Phase II CEU HapMap data (release 22, build 36). Detailed descriptions of quality control and imputation procedures are summarized in Table S8.

Statistical Methods

Association between Total Body BMD and GWAS SNPs was carried out using a regression framework adjusting for age, gender, weight and population stratification in the Generation R discovery cohort using MACH2QTL as implemented in GRIMP[32]. Since this is a population-based study on unrelated individuals of different ethnic background, 20 genomic principal components obtained after SNP quality exclusion criteria and LD pruning were used to adjust for population sub-structure reaching a Genomic Inflation Factor (λ) of 1. We selected the most associated SNP and SNPs located at +/- 500 kb from the top SNP for replication including all markers with a MAF>0.01 and an r^2 imputation quality score >0.3 in all the participating studies. Additionally, rs3801387 a proxy of the 'top hit' (r^2=1 in HapMap CEU populations) was genotyped in a subset of mothers of the Generation R Study of Dutch Northern European background. All replication cohorts included only individuals of North European ancestry and thus the correction for stratification was not as astringent as for the discovery cohort.

In the genome-wide association study, the association test of SNPs with standardized residuals of total body (skull) BMD after adjusting for age, gender, population stratification and weight (height) was implemented via Mach2QTL for all cohorts. Moreover, association in the mature adults and elderly cohorts, in which ample ranges of age are seen (Rotterdam Studies I, II, III and Mothers of Generation R) allowed for a non-linear relationship between age and BMD by inclusion of a squared term. On the conditional analysis we selected the

most associated genotyped SNP (rs3801382) and applied a regression model including that marker besides the mentioned covariates in order to evaluate its effect in the originally detected signal(s).

We carried out regional meta-analyses in METAL using the minor allele from HapMap CEU genotypes as the coding allele, and applying inverse-variance methodology assuming fixed effects. A P value less than $5x10^{-8}$ was considered genome-wide significant (GWS). Heterogeneity was evaluated using Cochran's Q statistic and was quantified by I^2. (http://www.sph.umich.edu/csg/abecasis/Metal/). For the meta-regression the absolute value of the effect size of the selected SNP (i.e. rs7801723, rs917727) in each trait was regressed on mean age of each of the six studies. These analyses were weighted by the inverse of variance of the effect; in line with the methodology applied for the meta-analyses.

Knockout mice

Wnt16 and *Fam3c* knockout (KO) mice were obtained from a program scrutinizing targets for drug discovery at Lexicon Pharmaceuticals. F2 hybrid littermates were derived from C57BL/6J and 129 SvEv parental strains. The 28 knockout mice (16 females) were generated by homologous recombination removing the first three exons of *Wnt16.* A three-fold strategy was used to generate *Fam3c* KO mice including a gene-trap (G-T) disrupting the intron between the first two exons of 6 mice (3 females); homologous recombination removing the first two exons (HR#1) of 8 mice (4 females); and homologous recombination replacing the whole gene of 8 mice (4 females) by the human gene (HR#2) resulting in loss of function. Confirmation of the exon disruption in *Wnt16* and *Fam3c* (HR#1) was achieved with Southern blot hybridization analysis while RT-PCR was used to confirm lack of gene expression of *Fam3c* (G-T) in KO mice (see Zheng et al. for details). Successful disruption of the *Fam3c* gene in all three KO strategies was demonstrated by a consistent hematological phenotype (data not shown). Male and female mice were scanned using a PIXImus DXA at 24 weeks (*Wnt16* KO comparison) and 14 weeks (*Fam3c* KO comparison) of age. BMD (and body composition) measurements were obtained from total body (excluding skull), femur and spine scans. Bone mineral content (BMC) was derived from the projected bone area (BA) as BMC (mg) = BMD mg/cm^2 X BA (cm^2). Student's t-tests were used to assess statistical significance ($P<0.05$) of the differences within each sex.

Gene transcripts expression levels from trans-iliacal bone biopsies

Gene expression profiles from all transcripts located within +/- 500kb of the rs917727 SNP in locus 7q31.31 were analyzed within an eQTL dataset of 78 Norwegian women, who make part of the set published by Reppe and colleagues [17]. Of these 78 women, 40 had osteoporosis (T-score less than -2.5), 7 had osteopenia (T-score between -2.5 and -1) and 31 were normal (T-score greater than -1) as ascertained by the BMD measurement at the total

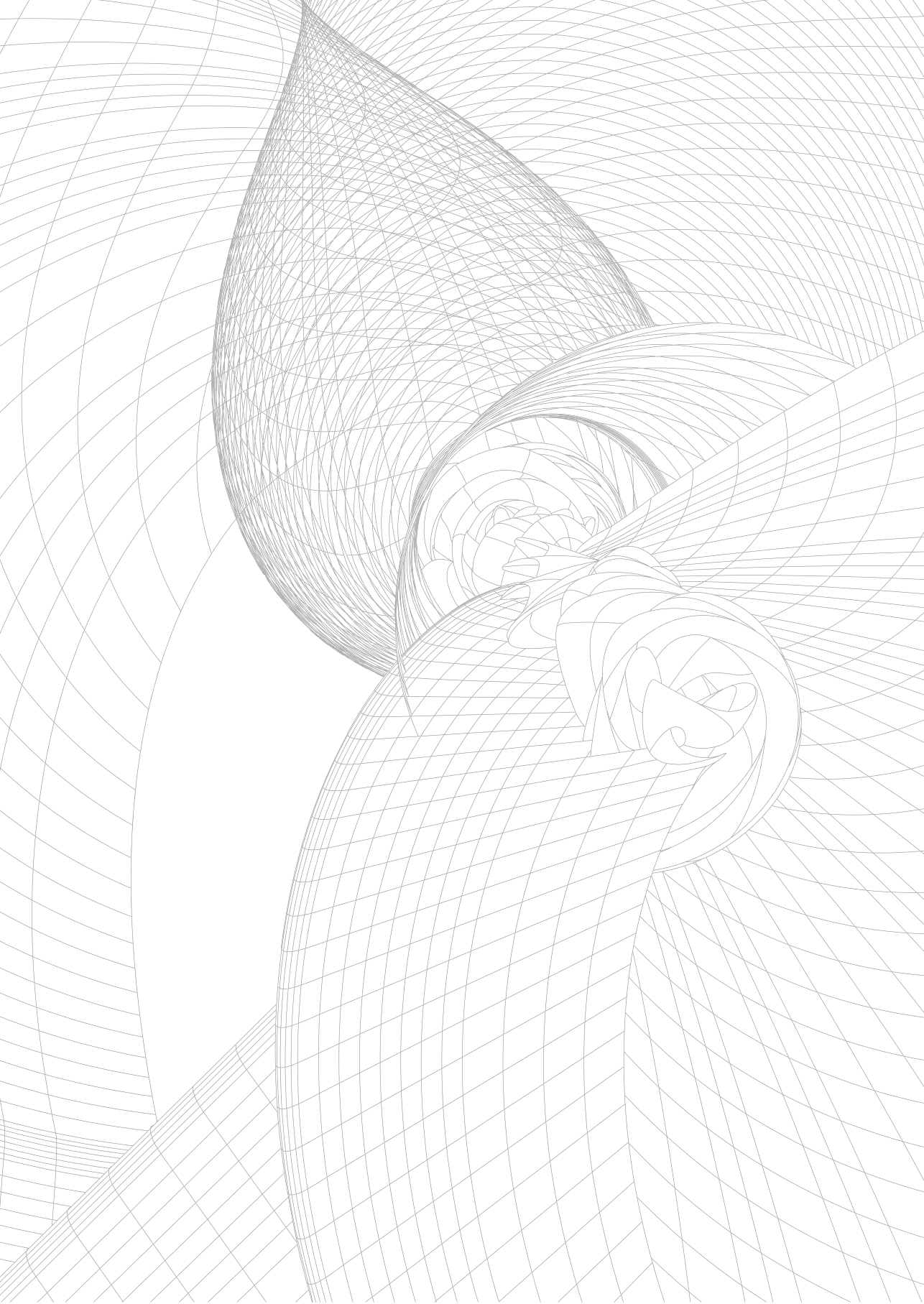
hip or lumbar spine (L1-L4 verterbrae). The Affymetrix HG U133 2.0 plus array was used for the expression analysis. The Affymetrix Cel files were imported into Partek Genomics Suite (Partek Inc., St Louis, MO, USA), and normalized using the RMA (Robust Multichip Average) algorithm. Further normalization was done by removing batch effects and patterns of gene expression levels due to differences in synthesis times across samples.

Detailed acknowledgements and online resources can be found in the published article online: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002718

REFERENCES

1. Siris ES, Miller PD, Barrett-Connor E, et al. Identification and fracture outcomes of undiagnosed low bone mineral density in postmenopausal women: results from the National Osteoporosis Risk Assessment. **Jama**. 2001;286(22):2815-22.
2. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
3. Clarke B. Normal bone anatomy and physiology. **Clin J Am Soc Nephrol**. 2008;3 Suppl 3:S131-9.
4. Yerges LM, Klei L, Cauley JA, et al. High-density association study of 383 candidate genes for volumetric BMD at the femoral neck and lumbar spine among older men. **J Bone Miner Res**. 2009;24(12): 2039-49.
5. Duncan EL, Cardon LR, Sinsheimer JS, Wass JA, Brown MA. Site and gender specificity of inheritance of bone mineral density. **J Bone Miner Res**. 2003;18(8):1531-8.
6. Rizzoli R, Bonjour JP, Ferrari SL. Osteoporosis, genetics and hormones. **J Mol Endocrinol**. 2001;26(2): 79-94.
7. Loro ML, Sayre J, Roe TF, Goran MI, Kaufman FR, Gilsanz V. Early identification of children predisposed to low peak bone mass and osteoporosis later in life. **J Clin Endocrinol Metab**. 2000;85(10):3908-18.
8. Holroyd C, Harvey N, Dennison E, Cooper C. Epigenetic influences in the developmental origins of osteoporosis. **Osteoporos Int**. 2012;23(2):401-10.
9. Timpson NJ, Tobias JH, Richards JB, et al. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. **Hum Mol Genet**. 2009;18(8):1510-7.
10. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. **Nat Genet**. 2009;41(11):1199-206.
11. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. **N Engl J Med**. 2008;358(22):2355-65.
12. Cho YS, Go MJ, Kim YJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. **Nat Genet**. 2009;41(5):527-34.
13. Zhang LS, Hu HG, Liu YJ, et al. A follow-up association study of two genetic variants for bone mineral density variation in Caucasians. **Osteoporos Int**. 2011.
14. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. **Nat Genet**. 2010;42(7):565-9.
15. Zheng HF, Tobias JH, Duncan E, et al. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. **PLoS Genet**. 2012;8(7):e1002745.
16. Rawlinson SC, McKay IJ, Ghuman M, et al. Adult rat bones maintain distinct regionalized expression of markers associated with their development. **PLoS One**. 2009;4(12):e8358.
17. Fear MW, Kelsell DP, Spurr NK, Barnes MR. Wnt-16a, a novel Wnt-16 isoform, which shows differential expression in adult human tissues. **Biochem Biophys Res Commun**. 2000;278(3):814-20.
18. Clements WK, Kim AD, Ong KG, Moore JC, Lawson ND, Traver D. A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. **Nature**. 2011;474(7350):220-4.
19. Dell'accio F, De Bari C, Eltawil NM, Vanhummelen P, Pitzalis C. Identification of the molecular response of articular cartilage to injury, by microarray screening: Wnt-16 expression and signaling after injury and in osteoarthritis. **Arthritis Rheum**. 2008;58(5):1410-21.
20. Wood AR, Hernandez DG, Nalls MA, et al. Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. **Hum Mol Genet**. 2011;20(20):4082-92.
21. Jaddoe VW, van Duijn CM, van der Heijden AJ, et al. The Generation R Study: design and cohort update 2010. **Eur J Epidemiol**. 2010;25(11):823-41.
22. Golding J, Pembrey M, Jones R, Team AS. ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. **Paediatr Perinat Epidemiol**. 2001;15(1):74-87.

23. Paternoster L, Howe LD, Tilling K, et al. Adult height variants affect birth length and growth rate in children. **Hum Mol Genet**. 2011;20(20):4069-75.
24. Lorentzon M, Swanson C, Andersson N, Mellstrom D, Ohlsson C. Free testosterone is a positive, whereas free estradiol is a negative, predictor of cortical bone size in young Swedish men: the GOOD study. **J Bone Miner Res**. 2005;20(8):1334-41.
25. Lorentzon M, Mellstrom D, Ohlsson C. Age of attainment of peak bone mass is site specific in Swedish men--The GOOD study. **J Bone Miner Res**. 2005;20(7):1223-7.
26. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. **Eur J Epidemiol**. 2011;26(8):657-86.
27. Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. **Bioinformatics**. 2009;25(20):2750-2.
28. Reppe S, Refvem H, Gautvik VT, et al. Eight genes are highly associated with BMD variation in postmenopausal Caucasian women. **Bone**. 2010;46(3):604-12.
29. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. **Annu Rev Genomics Hum Genet**. 2009;10: 387-406.
30. Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. **PLoS One**. 2009;4(9):e7038.
31. Milat F, Ng KW. Is Wnt signalling the final common pathway leading to bone formation? **Mol Cell Endocrinol**. 2009;310(1-2):52-62.
32. Krishnan V, Bryant HU, Macdougald OA. Regulation of bone mass by Wnt signaling. **J Clin Invest**. 2006;116(5):1202-9.
33. Zhu Y, Xu G, Patel A, et al. Cloning, expression, and initial characterization of a novel cytokine-like gene family. **Genomics**. 2002;80(2):144-50.



Chapter 4.2

Genetic studies of TB-BMD yield eighteen novel loci involved in bone biology

Carolina Medina-Gomez*, John Kemp*, Alessandra Chesi, Eskil Kreiner-Møller, Tarun Ahluwalia, Dennis Mook, Youfang Liu, Fernando P. Hartwig, Dan Evans, Raimo Joro, Cornelia van Duijn, Ivana Nedeljkovic, Benjamin Mullin, Joel Eriksson, Brent Richards, Rebecca Jackson, David Karasik, Nathalie Van der Velde, Arfan Ikram, Babette Zemel, Tamara Harris, Yanhua Zhou, John Robins, Ruifang Li, Bruce Psaty, Carrie Nielson, Bram van der Eerden, Jeroen van de Peppel, Wilson Scott, Bernardo L Horta, Timo Lakka, Struan Grant, Fiona McGuigan, Jim Wilson, Unnur Styrkársdóttir, Dan Koller, Kun Zhu, Doug Kiel, Claes Ohlsson, Cheryl L Ackert-Bicknell, Andre G. Uitterlinden, Vincent Jaddoe, Jon H. Tobias, Dave M. Evans, Fernando Rivadeneira

*These authors equally contributed to this manuscript

In preparation

ABSTRACT

Background: Bone mineral density (BMD) is used for the assessment of bone health in children and diagnose osteoporosis in the elderly population. While in pediatric populations, BMD is preferably measured from total body scans (TB-BMD); in adults, lumbar spine (LS) and femoral neck (FN) are traditionally used. Genetic factors explain 50-80% of BMD, depending on the site of measurement and study design. To date, more than sixty genetic loci have been associated with this trait, most of them associated with LS and FN-BMD in the elderly population. Here, we investigated if TB-BMD may constitute an adequate phenotype for genetic studies when applied to the whole population. **Methods:** We performed a meta-analysis of 26 genome-wide association studies (GWAS) on TB-BMD including 52,680 individuals. Also, we applied conditional analysis to determine independent association signals and evaluate their novelty. **Results:** We identified genome-wide significant variants in 55 loci of which 18 are novel and enriched for genes with a proven role in bone biology mainly as factors in the Wnt signalling pathways. Also, we identified independent signals in nine known loci associated with bone traits. **Conclusion:** TB-BMD is a relevant trait for genetic studies of osteoporosis, capable of identifying (novel) variants influencing different bone compartments at different skeletal sites. We identified 18 novel genetic loci influencing BMD variation, including several with factors embedded in Wnt signaling and a handful unknown to bone biology. These results notoriously contribute to our knowledge of BMD genetic determination and highlight the efficiency of increasing sample size for the discovery of new genetic variants.

INTRODUCTION

Osteoporosis is a disease characterized by low bone mass and microarchitectural deterioration of bone tissue leading to increased risk[1]. Clinically, osteoporosis is diagnosed through the measurement of bone mineral density (BMD) utilizing dual-energy X-ray absorptiometry (DXA), which remains the single best predictor of fracture[2,3]. A recent study included BMD in the global burden estimates as a risk factor for fractures after fall and concluded that low BMD is responsible for a growing global health burden, probably related to the global growth of the aged population[4].

Bone is a dynamic tissue that is constantly undergoing bone resorption and bone formation. BMD at any age is the result of bone accrued during growth, until the third decade of life and the subsequent bone loss thereafter. The International Society for Clinical Densitometry recommend DXA measurements at the lumbar spine (LS), femoral neck (FN) and total hip to predict osteoporosis in postmenopausal women and men 50 years or older [5]. Consequently, studies of modifiable and non-modifiable factors affecting BMD are based on measurements at these skeletal sites. On the other hand, for the assessment of bone health in children, from infancy to adolescence, total body less head and LS are the preferred sites for measurement [6]. As DXA is a projection of a 3-D structure into a 2-D image, the continual growth of bone in children introduce an error in this assessment and thus, larger skeletal sites with large areas are preferred[7]. Nevertheless, also in adults, total body DXA scans are commonly available, as they are used to measure total body composition and fat content with a high degree of accuracy.

In twin and family studies, the heritability of BMD has been estimated to be 50%-85%[8-10]. These estimates can vary due to the analytical model used, the skeletal site measured and the population under study [11]. Up to date, around 60 different loci have been shown to be robustly associated with BMD variability in adults [12-15]. Several of the associated variants display significant effects on a site-specific manner. It is plausible that these differences can be the result of bone composition dissimilarities across different skeletal sites (e.g. cortical bone vs. trabecular bone) or response to mechanical loading, which also varies across skeletal sites[15].

The purpose of this study was to identify gene variants associated with total body BMD (TB-BMD). The selection of TB-BMD as phenotype of interest allowed us to reach the largest GWAS of BMD to date. Also, as the total body BMD measurement incorporates components of both cortical and to a lesser extent trabecular bone, we expected to be able to identify variants which would not be necessarily discovered in a site-specific study.

METHODS

Study Populations

Subjects

This study comprised 26 GWA studies comprising ~52,680 individuals from populations across America, Europe, and Australia, with a variety of epidemiological designs (**Supplementary Table 1**) and participant characteristics (**Supplementary Table 2**). All research aims and the specific measurements in the participating studies 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/cm^2) was measured in all participants using dual-energy X-ray absorptiometry (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 [16] (e.g., 0-15 years). Detailed information on this assessment per study can be found in **Supplementary Table 1**.

GWAS data and imputation

All individuals included in this study had genome-wide array data, except for the PANIC study genotyped with the Metabochip. Quality control of the obtained genotypes is summarized in **Supplementary Table 1**. 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 [17], via MACH/Minimac[18,19] or SHAPEIT/Impute2[20] suites yielding ~ 30,000,000 SNPs for analysis.

Association Analyses

TB(LH)-BMD was corrected for age, weight, height and the 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 subject to inverse normal transformation. SNP association was tested for autosomal variants, in which the additive effect of each SNP on the normalized BMD-residual was estimated via linear regression using MACH2QTL[18], SNPtest[21], PLINK[22] or ProbABEL[23].

Quality control of TB-BMD association summary statistics

A centralized quality-control procedure implemented in EasyQC [24] was applied to all study-specific association files to identify cohort-specific problems: (1) assessment of possible problems in BMD-residuals normalization by inspection of standard error vs. sample size plots; (2) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues; (3) examination of quantile-quantile (QQ) plots and genomic inflation factors per study to identify problems arising from population stratification, cryptic relatedness and genotype biases; (4) analytical problems in the computation of beta estimates, standard errors and P-values by evaluation of P-Z plots.

In addition, we excluded variants if they had missing information (e.g., missing association P-value, effect size (beta) estimate, alleles, allele frequency), or nonsensical values (e.g., absolute beta estimates or standard errors >10, association P-values (P) >1 or <0; or imputation quality < 0; infinite beta estimates or standard errors); minor allele frequency (MAF) less than 0.5%; Imputation quality <0.4 (when Impute2 was used for imputation) and <0.3 (when Minimac was used for imputation). Moreover, markers were flagged if they had large allele frequency deviations from reference populations (>0.6 for admixed studies and >0.3 for ancestry-specific studies).

The meta-analysis of all individuals per participant cohort consisted of 26 studies comprising ~52,680 individuals. Although no exclusion criteria were applied based on the ethnicity of the individuals included in each study, more than 80% of them were reported as of European ancestry (**Supplementary Table 1**). We discarded markers present in less than three studies; approximately 23,700,000 markers (both SNPs and INDELS) were assessed for association. We used the conventional genome-wide significance level (GWS, $P<5x10^{-8}$) for SNP discovery.

Conditional meta-analyses

Conditional analyses were undertaken based on the combined meta-analysis of the studies of European ancestry only (N=42,352). To note, only those loci which reached GWS in the described meta-analysis of cohorts of European ancestry were assessed. The Rotterdam Study I (n=6,291) was used as a referent for precise calculation of the linkage disequilibrium (LD) between the analyzed markers. We used an iterative strategy as implemented in GCTA [25,26] ($P<5x10^{-8}$) 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. 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 (**Supplementary Table 3**). To this end, we performed the association analysis conditional on 74 bone-traits associated variants (--massoc-cond routine). These 74 SNPs were selected from different GWAS publications [12-15,27-29], assuring their inter-independence to avoid collinearity issues.

Search for biological and functional knowledge of the identified association regions

For all those SNPs not within 500Kb of the bone GWAS SNPs (hereinafter referred as novel) we checked in PubMed and Web of Science 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 KO 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 assessed through Bioconductor [30].

Pathway analysis

We used a recently developed method, DEPICT[31] to identify enriched genes-sets as well as tissues/cell types where genes from associated loci are highly expressed. The methodology first selects all lead SNPs below a certain threshold with respect to a target P-value (GWS SNPs, $P<5x10^{-8}$). Additionally, we used clusterProfiler [32]developed in R[33] for a Gene Ontology (GO) enrichment analysis[34]. This analysis opposite to DEPICT included not only the genes in LD r>0.5 with the index SNP but also the nearest transcribed element as input. Given the redundancy of GO terms, we have used the REVIGO tool to prune the results employing the simRel algorithm and a threshold of 0.5 for allowed similarity [35]. Adjusted P-value of all enrichment analysis were obtained by False Discovery Rate (FDR) method.

RESULTS

GWAS meta-analysis the overall population

Meta-analysis of all cohorts resulted in 52 different loci associated with TB-BMD at GWS level (**Figure 1**). Genomic inflation factor for the meta-analysis was 1.06 providing no evidence for population stratification, cryptic relatedness or technical artifacts as illustrated in the Q-Q plots. The highest inflation of the test statistic was observed for the common SNPs (0.2<MAF<0.5) reaching 1.14. From the 52 identified loci, 16 (31%) were not located within 500 kb from index SNPs previously associated with any bone trait through GWAS nor in regions of extended LD with them (**Table 1**). Sixty-seven out of the 74 index SNPs in loci known to be associated with bone traits, were associated with TB-BMD in our study at a nominal level (P<0.05) (**Supplementary Table 3**).

Index SNPs of association signals located in the 16 novel loci were, in general, common non-coding markers. Twelve of them map nearby genes likely to influence bone metabolism (**Table 1, Supplementary Figure 1**). Fourteen of these sixteen index SNPs had concordant effect direction and were nominally associated with either LS-BMD or FN-BMD in the last published GEFOS meta-analysis for these traits[13] (**Table 1**).

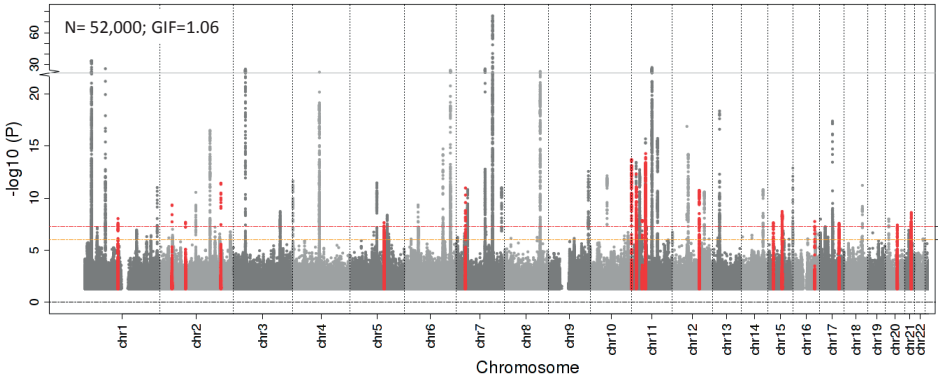


Figure 1. Manhattan plot of association statistics (-log10(P values)) for TB-BMD trans-ethnic meta-analysis. Each dot represents an SNP and the x-axis indicates its chromosomal position (built 37 NCBI). Red dots represent SNPs not within ±500Kb of leading SNPs in previous GWAS with different bone traits. 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.

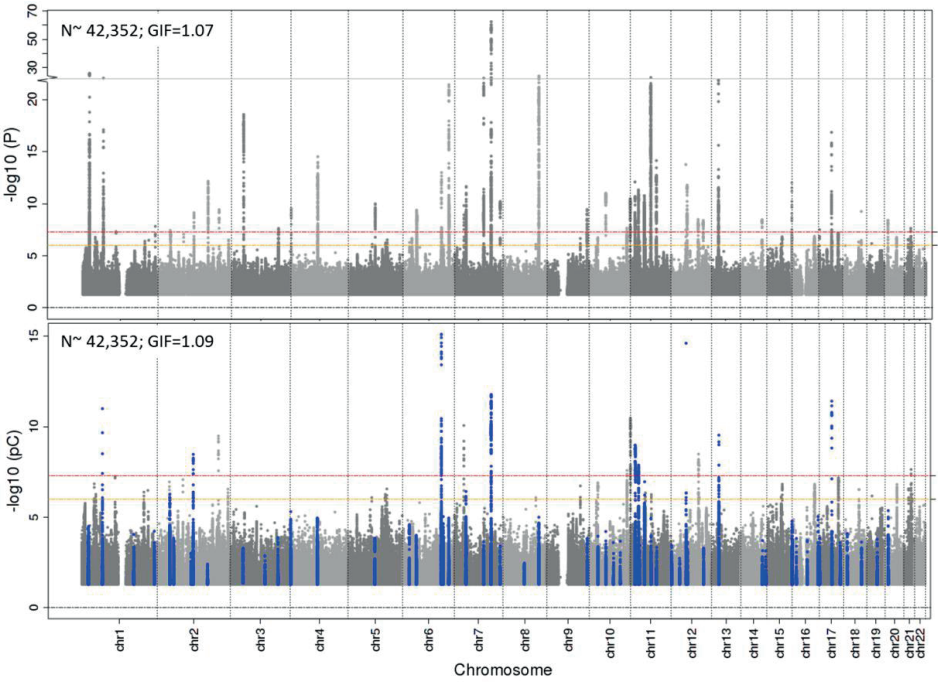


Figure 2. Manhattan plots of association statistics (-log10(P values)) for TB-BMD only-European meta-analysis. Each dot represents an SNP and the x-axis indicates its chromosomal position (Build 37 NCBI). Dashed horizontal red and yellow lines mark the GWS threshold ($P<5x10^{-8}$) and suggestive threshold ($P<1x10^{-6}$), respectively. **Top**: The association Pvalue (on $-\log_{10}$ scale) in the meta-analysis including only studies comprising individuals of European ancestry. **Bottom:** The association Pvalue (on $-\log_{10}$ scale) after conditional analysis on all variants described in Supplementary Table 3. Highlighted in blue already known loci (SNPs within ±500Kb of leading SNPs in previous GWAS with different bone traits).

Table 1. Index SNPs of novel locus 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. Associations with lumbar Spine (LS) and Femoral Neck (FN)-BMD were retrieved from Zheng et al. [13].

CHR	BP	rsID	A1	A2	Freq1	Effect	P	HetP	N	annotation	close gene	Notes	LS-β	LS-P	FN-β	FN-P
1	110475971	rs7548588	T	C	0.612	-0.037	9.29E-09	0.009	52308	intergenic	*CSF1*	Osteoclast differentiation [49]	-0.030	0.001	-0.022	0.005
2	40630678	rs10490046	A	C	0.757	0.047	4.55E-10	0.030	52029	intron	*SLC8A1*	Bone mineralization[50]	0.015	0.162	0.021	0.025
2	85484818	rs11904127	A	G	0.545	-0.036	2.12E-08	0.556	52629	intron	*TCF7L1*	Factors in Wnt signaling[51,52]	-0.021	0.023	-0.015	0.054
2	202799604	rs2350085	T	C	0.871	-0.066	3.67E-12	0.567	52480	intergenic	*FZD7*	Factors in Wnt signaling[53]	-0.042	0.002	-0.044	1.96E-04
5	112205298	rs438252	A	C	0.302	0.041	2.27E-08	0.943	47735	intron	*APC*	Bone metabolism[46]	0.004	0.846	0.021	0.161
5	122847622	rs11745493	A	G	0.747	0.043	4.31E-09	0.588	52665	promoter	*CSNK1G3*	Novel Biology	0.010	0.326	0.025	0.005
7	30957702	rs28362721	T	C	0.176	-0.061	1.07E-11	0.068	52342	intron	*AQP1*	Bone metabolism[38]	-0.037	0.002	-0.049	1.39E-06
10	124086757	rs6585812*	T	C	0.518	0.040	2.24E-08	0.260	41825	Intron	*BTBD16*	Novel Biology	0.042	2.56E-06	0.033	1.35E-05
11	237209	rs56312618	A	G	0.2137	0.0595	2.11E-14	0.084	52241	Intron	*RIC8A*	Ossification process[40]	0.045	2.26E-05	0.026	0.004
11	35981346	rs113964474**	A	G	0.025	0.485	1.41E-08	0.042	6748	Intron	*LDLRAD3*	Novel Biology	.	.	.	.
12	90337046	rs4842697	A	C	0.591	-0.043	1.79E-11	0.939	52679	intergenic	*ATP2B1*	Calcium absorption[54]	-0.027	0.003	-0.025	0.001
12	92597520	rs11106464*	A	G	0.234	0.049	1.14E-08	0.607	42289	intergenic	*BTG1*	Ossification process[43]	0.040	1.36E-04	0.015	0.091
15	38332568	rs7170368	A	T	0.807	-0.047	2.31E-08	0.426	52484	intergenic	*TMCO5A*	Novel Biology	-0.044	2.15E-04	-0.049	2.02E-06
15	67420680	rs1545161	A	G	0.564	0.039	1.91E-09	0.361	52072	Intron	*SMAD3*	Osteoblast differentiation[55]	0.034	1.27E-04	0.035	5.78E-06
16	73100308	rs6416749	T	C	0.656	-0.045	1.76E-08	0.685	46686	intergenic	*ZFHX3*	Novel Biology	-0.041	6.28E-05	-0.022	0.015
17	63771079	rs9972944	A	G	0.413	0.037	2.65E-08	0.402	52663	Intron	*CEP112*	Novel Biology	0.028	0.003	0.004	0.576
20	39103882	rs6029130	T	C	0.295	0.039	3.85E-08	0.625	52565	intergenic	*MAFB*	Osteoclast differentiation[56]	0.027	0.007	0.015	0.083
21	36970350	rs9976876	T	G	0.442	-0.039	2.44E-09	0.296	52582	Intron	*RUNX1*	osteoclast differentiation[57]	-0.019	0.031	-0.016	0.041

*Results for this variants are derived from only-European meta-analysis instead of the trans-ethnic meta-analysis. ** Only polymorphic in Africans.

Meta-analysis based in studies with European ancestry

We conducted a step-wise conditional approach in order to determine the independent SNPs associated with TB-BMD including studies comprising only individuals of European ancestry. The first step consisted of a meta-analysis based only in these 20 cohorts. We identified 45 loci associated with the trait (**Figure 2**). Of these, three loci surpassed the GWS threshold in the only-European analysis despite only being suggestive in the larger trans-ethnic meta-analysis. Two of them (e.g., 10q26.13 (*BTBD16*) and 12q21.33 (*BTG1*)) were novel (**Table 1, Supplementary Figure 1**). Whereas the remaining signal mapped to the already known *PKDCC* locus.

Conditional analysis based in studies with European ancestry

To perform adequate conditional analysis, we opted to exclude non-European cohorts leaving for this analysis ~80% of the original sample size (N=42,352). As a consequence of this lower sample size, 9 of the 52 initially discovered loci in the trans-ethnic meta-analysis did not reach the GWS threshold and were excluded from the conditional analysis. All of the excluded signals were novel and included 2p11.2 (*TCF7L1*), 5q22.2 (*APC*), 5q23.2 (*CNK1G3*), 11p13 (*LDLRAD3*), 15q14 (*TMCO5A*), 15q22.33 (*SMAD3*), 16q22.3 (*ZFHX3*), 17q24.1 (*CEP112*) and 20q12 (*MAFB*). Yet, index SNPs from eight of these loci were all suggestively associated with TB-BMD in this meta-analysis ($5x10^{-8}<P<5x10^{-6}$).

We performed an approximate conditional analysis which resulted in the identification of 55 independent SNPs mapping to 45 different loci. Loci harboring 1p36 (*WNT4/ZBTB40)*, 6q22.33 (*RSPO3*), 6q25.1 (*ESR1/CCDC170)*, 7q31.31 (*WNT16/CPED1*), 13q14.11 (*AKAP11/TNFSF11*), 17q21.31 (*SOST/UBTF*), 11p14.1 (*CCDC34/LINC7),* and 11p15.2 (*SOX6*) depicted multiple distinct signals attaining GWS (**Table 2**). The TB-BMD variance explained by these 55 variants reached 8.1%.

We next investigated if variants reaching GWS in our analysis were independent of previous variants associated with bone traits in well-powered GWAS meta-analysis (**Supplementary Table 3**). Our analysis revealed that in nine of the loci previously reported as associated with bone traits, we identified statistically independent signals (**Figure 2**). None of the subset of new loci inquired in the approximate conditional analyses: 2q33.1 (*FZD7*), 7p14.3 (*AQP1*), 10q26.13 (*BTBD16),* 11p15.5 (*RIC8A),* 12q21.33 (*ATP2B1*), 12q21.33 (*BTG1*), 21q22.12 (*RUNX1*) and 1p13.3 (*CSF1*), showed secondary signals, and only *CSF1* did not reach GWS when conditioning for known variants ($P_{before\_cond}=3.7x10^{-8}$, $P_{after\_cond}=1.1x10^{-7}$).

Biological and functional knowledge of the identified association regions

Novel loci and their potential role in bone metabolism are summarized in **Table 1**. Across these novel loci, we identified only one coding marker associated with TB-BMD,

Table 2. Independent SNPs associated with BMD in the only –European meta-analysis. Each locus is named according to either the most biologically relevant candidate gene in the region, or the gene that is physically closest to the most strongly associated SNP. Note that in the vast majority of instances neither the identity of the true functional variant(s), nor the particular gene responsible for the association is known with certainty

Chr	BP	rsID	A1	freq1	Effect	P	Effect_joint	P_joint	± 500 Kb	Locus
1	22484575	rs3971300	T	0.70	0.072	1.33E-18	0.075	1.02E-19	yes	*WNT4*
1	22692315	rs12728589	C	0.18	0.102	1.05E-26	0.104	5.77E-28	yes	*ZBTB40*
1	68656697	rs2566752	T	0.61	-0.074	2.71E-23	-0.074	5.61E-23	yes	*WLS*
1	2.41E+08	rs12044944	T	0.19	0.052	1.48E-08	0.052	1.78E-08	yes	*FMN2*
2	40630678	rs10490046	A	0.76	0.047	3.69E-08	0.048	2.70E-08	no	*SLC8A1*
2	42194758	rs2374385	A	0.22	-0.049	4.72E-08	-0.050	3.23E-08	yes	*PKDCC*
2	1.2E+08	rs62159873	T	0.21	-0.061	7.39E-10	-0.061	8.32E-10	yes	*EN1*
2	1.67E+08	rs7586085	A	0.52	0.052	6.89E-13	0.052	5.82E-13	yes	*GALNT3*
2	2.03E+08	rs6716216	A	0.88	-0.069	3.70E-10	-0.069	3.20E-10	no	*FZD7*
3	41118898	rs419918	C	0.52	0.065	2.62E-19	0.065	3.51E-19	yes	*CTNNB1*
3	1.57E+08	rs11717138	T	0.23	-0.048	2.32E-08	-0.048	2.32E-08	yes	*LEKR1*
4	1008972	rs74569579	A	0.15	-0.070	2.87E-10	-0.070	3.33E-10	yes	*IDUA*
4	88831249	rs11934731	A	0.68	-0.061	3.00E-15	-0.061	3.68E-15	yes	*MEPE*
5	88376061	rs1366594	A	0.52	0.047	1.03E-10	0.047	1.23E-10	yes	*MEF2C*
6	44639184	rs11755164	T	0.40	-0.047	5.42E-10	-0.047	4.45E-10	yes	*SUPT3H/RUNX2*
6	1.27E+08	rs13220518	A	0.22	-0.060	1.66E-12	-0.055	7.99E-11	yes	*RSPO3*
6	1.27E+08	rs4580892	T	0.30	0.060	1.01E-13	0.055	5.72E-12	yes	*RSPO3*
6	1.52E+08	rs6557155	T	0.42	-0.073	3.64E-22	-0.075	2.31E-23	yes	*ESR1/CCDC170*
6	1.52E+08	rs7765040	A	0.84	0.056	2.76E-08	0.061	1.48E-09	yes	*ESR1/CCDC170*
7	30957702	rs28362721	T	0.18	-0.065	1.48E-10	-0.065	1.09E-10	no	*AQP1*
7	38136277	rs1524058	T	0.40	-0.051	2.21E-12	-0.052	1.39E-12	yes	*STARD3NL*
7	96133319	rs6965122	A	0.69	0.076	3.05E-23	0.076	4.76E-23	yes	*SLC25A13*
7	1.21E+08	rs6950680	A	0.63	0.055	1.29E-13	0.072	8.84E-22	yes	*CPED1*
7	1.21E+08	rs3801387	A	0.74	-0.135	4.27E-63	-0.165	5.81E-82	yes	*WNT16*
7	1.21E+08	rs2041490	C	0.18	0.018	0.06782	0.067	2.90E-11	yes	*WNT16*
7	1.21E+08	rs6952513	T	0.07	0.104	4.71E-12	0.104	4.34E-12	yes	*FAM3C*
7	1.51E+08	rs113810201	A	0.88	-0.071	7.68E-11	-0.071	6.60E-11	yes	*ABCF2*
8	1.2E+08	rs1485307	T	0.44	0.075	8.52E-25	0.075	1.58E-24	yes	*TNFRSF11B*
9	1.33E+08	rs10901216	A	0.35	-0.048	3.56E-10	-0.048	3.61E-10	yes	*FUBP3*
10	54420223	rs7070913	A	0.12	-0.078	9.89E-12	-0.078	1.01E-11	yes	*MBL2/DKK1*
10	1.24E+08	rs6585812	T	0.52	0.040	2.24E-08	0.040	2.58E-08	no	*BTBD16*
11	249408	rs17655188	A	0.22	0.056	3.36E-11	0.056	3.30E-11	no	*RIC8A*
11	15695331	rs9787942	T	0.21	0.054	9.69E-10	0.056	1.77E-10	yes	*SOX6*
11	15814794	rs11023718	T	0.04	0.128	7.61E-10	0.119	1.17E-08	yes	*SOX6*
11	16249659	rs10766308	A	0.26	0.054	2.14E-11	0.052	2.22E-10	no	*SOX6*
11	27308483	rs10450586	C	0.61	-0.051	4.69E-12	-0.053	5.66E-13	yes	*CCDC34*

Table 2 (continued)

Chr	BP	rsID	A1	freq1	Effect	P	Effect_joint	P_joint	±500 Kb	Locus
11	27593899	rs1352479	A	0.28	0.043	3.32E-07	0.046	4.13E-08	yes	*LINC7*
11	46856536	rs10838622	T	0.37	0.053	1.66E-11	0.053	1.69E-11	yes	*CKAP5*
11	68218290	rs11228240	T	0.26	-0.086	1.00E-23	-0.086	9.37E-24	yes	*LRP5*
11	86887931	rs634277	A	0.67	0.058	2.18E-13	0.058	1.74E-13	yes	*TMEM135*
12	49385679	rs10875906	T	0.27	0.048	3.42E-08	0.049	1.68E-08	yes	*DDN*
12	53743064	rs10735851	A	0.71	-0.057	1.62E-12	-0.057	7.98E-13	yes	*SP7*
12	90334829	rs10777212	T	0.35	0.045	3.25E-09	0.045	3.22E-09	no	*ATP2B1*
12	92597520	rs11106464	A	0.23	0.049	1.14E-08	0.049	1.37E-08	no	*BTG1*
12	1.07E+08	rs759603	T	0.50	0.042	5.92E-09	0.042	4.65E-09	yes	*RIC8B/TMEM263*
13	42952145	rs9594738	T	0.47	-0.071	1.30E-22	-0.064	8.38E-19	yes	*AKAP11*
13	43077245	rs116926994	A	0.97	-0.173	5.12E-14	-0.145	4.67E-10	yes	*TNFSF11*
14	91445162	rs1286079	T	0.19	0.055	3.39E-09	0.055	3.85E-09	yes	*RPS6KA5*
16	392318	rs8047501	A	0.49	0.054	9.58E-13	0.054	1.02E-12	yes	*AXIN1*
17	2064702	rs2873195	A	0.32	-0.043	3.15E-08	-0.043	3.56E-08	yes	*SNG6*
17	41798621	rs66838809	A	0.08	0.125	1.36E-17	0.127	7.01E-18	yes	*SOST*
17	42281282	rs4473241	T	0.31	0.045	3.30E-08	0.046	1.71E-08	yes	*UBTF*
18	60054857	rs884205	A	0.25	-0.053	5.51E-10	-0.053	5.73E-10	yes	*TNFRSF11A*
20	10640877	rs6040063	A	0.50	0.042	3.90E-09	0.042	3.07E-09	yes	*JAG1*
21	36963137	rs118052194	T	0.59	0.043	2.30E-08	0.043	1.89E-08	no	*RUNX1*

rs77447196/G (11:280464, MAF=0.21) in the *NLRP6* gene (Pro244Ala), which is cataloged as benign both by SIFT and polyphen2. Also, we report 32 GWS coding variants in known loci from which 20 are non-synonymous. A low-frequency variant in *LRP5*, rs4988321/A (11:68174189, MAF=0.04), is predicted to be damaging by both polyphen2 and SIFT (Val-667Met). Moreover, a homozygous G-to-A transition of this variant was identified in a patient with osteoporosis-pseudoglioma syndrome [36]. Another common variant in *ESPL1* rs1318648/C (12: 53670545, MAF=0.36) is also cataloged as probably damaging and damaging in polyphen2 and SIFT, respectively; nevertheless, this variant has no clinical annotation. Mice and rat knockout models of several genes in the implicated loci, show an impaired skeletal phenotype (e.g., Csf1[37], Aqp1[38], Apc[39], Ric8a[40], Runx1[41], Smad3[42], Btg1[43]).

Pathway analysis

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 were highly expressed we used DEPICT. DEPICT prioritized 43 genes and 271 pathways (**Supplementary Table 4**). Also, there was significant enrichment expression of genes within loci associated with TB-BMD (FDR<0.05) in muscular and cartilaginous tissues and cells.

Based only on the novel variants DEPICT defined 17 loci -the *LDLRAD3* locus was not included as reference data derived from European ancestry- and genes mapping to these loci showed enrichment of expression only in eight cells and tissues, from which five are related to the musculoskeletal system (**Figure 3**). No significant enrichment for pathways or prioritized genes were uncovered (FDR<0.05). Nevertheless, *C15orf61, ZFHX3, REEP5* and *FZD7* were nominally prioritized. Additionally, we used the transcribed elements defined by DEPICT per locus to assess GO enrichment. The *ATP2B1*, *BTG1* and *TMCO5* genes in **Table 1** were not included as they failed both conditions regardless of being less than 500Kb from the index SNP. *LDLRAD3* was added to the list as the index SNP is intronic to this gene.

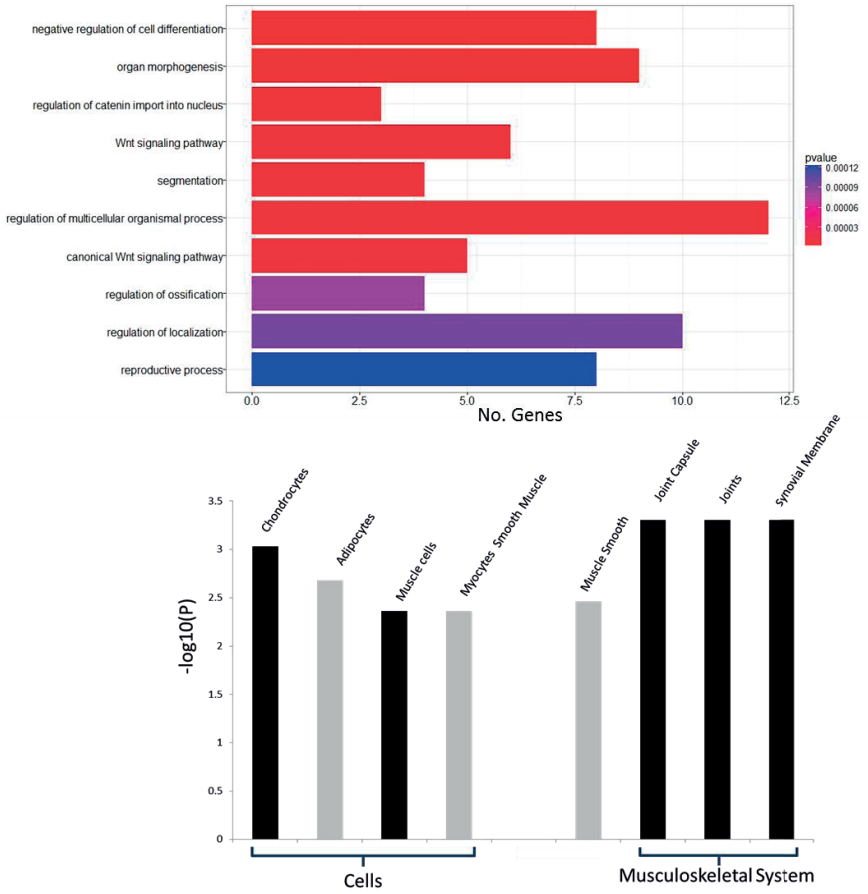


Figure 3. Pathway enrichment and tissue expression of genes at the novel TB-BMD-associated loci (FDR<0.05). Top panel: Gene Ontology (GO) enrichment analysis for 27 genes only the top 10 results are shown. **Bottom panel:** Enrichment and tissue expression of genes on the basis of expression patterns in 37,427 human microarray samples, annotations found to be significantly enriched by DEPICT are shown, grouped by type and significance. Specific genes at the loci were significantly enriched for expression in musculoskeletal-related cells and tissues (highlighted in black).

Pruned analysis of GO enrichment resulted in 55 enriched GOterms at an FDR<0.05 threshold (**Supplementary Table 5**). The GOterm with the highest enrichment was "negative regulation of cell differentiation" (*APC, SMAD3, ZFHX3, AXIN2, FZD7, TCF7L1, RUNX1*,and *MAFB*). Among the ten GO terms more significantly enriched three were related to Wnt signaling pathways including: "Catenin import into the nucleus" (*SMAD3, AXIN2* and *FZD7*), "Wnt signaling pathway" (*CSNK1G3, APC, SMAD3, AXIN2, FZD7* and *TCF7L1*) and "canonical Wnt signaling" (*APC, SMAD3, AXIN2, FZD7* and *TCF7L1*). Also, the term "regulation of ossification" (*CSF1, APC, SMAD3* and *AXIN2*) was significantly enriched (**Figure 3**).

DISCUSSION

This meta-analysis of TB-BMD in up to 52,000 individuals emerges as the largest discovery GWAS ran for BMD to date. We have identified 18 novel loci and replicated at GWS several known association signals with diverse bone phenotypes. Loci identified for the first time by our study comprised genes enriched for expression, almost exclusively, in musculoskeletal-related cells and tissues, and solidly represented in the Wnt Signaling pathway. Conditional analysis confined to studies of European ancestry identified secondary signals in 9 of the known associated loci. We have identified six loci unknown to bone biology whose role in BMD variation reminds to be elucidated. These results emphasize the suitability of TB-BMD as a phenotype to study the genetic factors influencing bone quality.

Traditionally, DXA-BMD measurements 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 derived from total body scans, widely available because of their appropriateness to study body composition. Not only we show a high overlap of association signals with previous GWAS of different bone traits but also identified novel loci, demonstrating that TB-BMD offers a powerful alternative to identify genetic variants associated with bone metabolism. We have scrutinized 74 previously reported signals of association with diverse bone traits and 39 of them surpassed GWS threshold in our meta-analysis. From these, 36 were previously associated with DXA-derived measurements of BMD at specific skeletal sites: 23 with both FN-BMD and LS-BMD[12], 5 only with FN-BMD[12], another 5 only with LS-BMD[12,13], 1 with Hip-BMD[29], and other 2 with pediatric skull BMD[15] and TB-BMD[14]. Moreover, we also replicated association signals obtained from bone phenotypes acquired with techniques other than DXA, as ultrasound BUA and VOS measures[28] or cortical and trabecular vBMD derived from pQCT[27] (**Supplementary Table 3**). Additionally, other 28 reported association signals, have consistent effect direction and a nominally significant evidence of association in our analysis.

Association signals at seven loci failed to replicate in our study (index SNPs: rs148771817 (7q31.31), rs754388 (14q32.12), rs7017914 (8q13.3), rs7071206 (10q22.3), rs7953528 (12p11.22), rs1878526 (2q14.2), rs3905706 (10p.11.23)). The last four markers reported significant site effect heterogeneity in Estrada et al. [12]. Of note, these markers were only nominally associated in the latest GEFOS discovery meta-analysis of FN-BMD and LS-BMD [13] with the same BMD trait for which they were discovered. These results suggest that variants whose effect is highly site-specific cannot be detected in our TB-BMD meta-analysis. Whilst, this event involved a limited number of the tested known variants (<10%), it is plausible that more variants of this type exist and will be discovered as site-specific BMD meta-analysis increase in size. Furthermore, the rs7017914 mapping to the *XRK9* locus was identified as associated with FN-BMD only in females[12], while our study is sex-combined. Also, the rs754388 mapping to the *RIN3* locus has only been associated with pediatric BMD [15]. This particular SNP shows high effect heterogeneity in our meta-analysis (P_{het}=3x10^{-5}) probably as a result of the different age ranges of the participating cohorts. Also, even if rs148771817 (*WNT16/CPED1)* and rs1878526 (*INSIG2)* did not replicate in our study, neighboring variants (within 500Kb), not in LD with them, reached GWS evidence of association.

The transethnic meta-analysis presented no evidence of population stratification and identified 52 loci associated with BMD. Index SNPs from these loci were in a vast majority common non-coding variants. In six of these loci, index SNPs show high effect heterogeneity (P_{het}<0.05), however, neighboring variants also reaching GWS exist in all these but the *LDLRAD3* locus. The index SNP of this association signal, the rs113964474 variant, located in the first *LDLRAD3* intron*,* is monomorphic in all our European studies, ergo the association was confined to four studies. The low allele frequency (MAF= 0.025) of this variant and our limited statistical power (N=6,748), allude this might be a false-positive association signal, and replication of this finding is necessary. Nevertheless, plasma lipids have been previously associated with BMD[44], and several other low-density lipoprotein receptors have been described as playing a role in the maintenance of bone (e.g., *LRP4, LRP5* and *LRP6*)[45].

Index SNPs in the remaining discovered loci, all but one had consistent effect direction and nominally significant association with FN-BMD or LS-BMD, implying a low probability for false-positives among these findings (**Table 1**). The association signal in 5q21, whose index SNP rs438252 maps to *SREP19*, although, in high linkage disequilibrium with variants in *REEP5* and *APC* was not significantly associated with LS- or FN-BMD. None of the variants reaching GWS in this locus is functional, complicating the identification of the gene underlying the association. DEPICT nominally significantly prioritize *REEP5* in this locus. Nevertheless, *APC* can affect osteoblast viability and signaling [46] and its KO mouse model shows impaired skeletogenesis (all endochondral bones were misshaped and lacked structural integrity)[39], making it a strong candidate to drive the unveiled association.

We also performed a meta-analysis including only studies from European ancestry identifying three loci which did not reach GWS in the trans-ethnic meta-analysis. Moreover, ten of the novel BMD-loci did not reach GWS in the European only meta-analysis. Both differences in MAF or effect size for these variants suggest that the differences observed between these analyses is a consequence of the change in sample size and not to the different population structure embedded in each of them. The conditional analysis identified 55 independent SNPs associated with TB-BMD mapping to 45 different loci which together explained 8.03% of the TB-BMD variance. Loci harboring *CCDC34/LINC7, RSPO3, WNT4/ZBTB40, ESR1/CCDC170, WNT16/CPED1/FAM3C, AKAP11/TNFSF11, SOST/UBTF,* and *SOX6* depicted multiple distinct signals attaining GWS.

Allelic heterogeneity in the latter six loci has already been established by previous meta-analyses[12,13], Conditioning on the known variants associated with bone traits demonstrated the existence of allelic heterogeneity additionally in loci harboring *WLS, EN1* and *WNT1/DHH* were we report new secondary signals. Noteworthy, *WNT1* has never previously been implied in bone phenotypes through GWAS. Although, variant rs12821008, previously associated with LS-BMD is only 98.2 Kb downstream of the gene. The independent variant reported in our study is only 3 Kb downstream of *WNT1*. Yet, functional studies will be needed in order to confirm the involvement of *WNT1* in normal BMD variation. This gene is responsible for rare monogenetic forms of early-onset osteoporosis and osteogenesis imperfecta[47].

We have shown that the genes mapping to the novel loci identified as associated with TB-BMD in our analysis are enriched for expression in cells and tissues relevant to the musculoskeletal system, whereas enrichment in cells or tissues, particularly coming from smooth muscle are less clearly connected to the studied phenotype. DEPICT did not identify any enriched geneset. Using the same genes considered by DEPICT but including always the closest transcripted element to each index SNPs we assessed several GOterms to be enriched. Our GWS results were overrepresented by genes involved in Wnt signaling pathway. On top of markers mapping to Wnt factors known to be associated with BMD (e.g., *WNT4, WNT16, WLS, CTNNB1, SOST, DKK1, MEF2C, AXIN1, LRP4, LRP5*) our analysis implicated additional Wnt signaling factors (e.g., *TCF7L1, FZD7, APC, SMAD3, AXIN2, CSNK1G3*). In addition, *CSF1, APC, SMAD3* and *AXIN2* are annotated in the GO term of regulation of ossification. To note, enrichment analysis are biased towards what it is already known.

Our study has limitations. The identified SNPs are in their vast majority, not coding variants, increasing the probability that the causal genes are different from the candidate genes we have prioritized here, based on the current biological knowledge and prediction bioinformatic tools. Further functional studies are required to unveil the potential role of the genes in the identified loci. Despite the large sample size of our study, we did not identify several

variants in the low-frequency spectrum pointing out that comprehensive surveys of rare variation influencing BMD would need an even larger sample size. Results in other traits as height or BMI, have already shown that GWAS with better coverage and in larger populations nearly all the proportion of genetic variability of a trait can be accounted for [48].

Altogether, this study is the largest-scale genome-wide survey for association with BMD to date. Opposite to previous studies, we used TB-BMD and not DXA measurements on skeletal sites prone to fracture to identify genetic factors influencing BMD variation. We demonstrate that TB-BMD is a suitable phenotype to study, as more than 90% of previously reported signals, regardless of site or compartment specificity, are nominally associated with TB-BMD in our meta-analysis. Most importantly, we identified 55 different loci associated with TB-BMD from which 18 have not been previously reported in GWAS of bone phenotypes. Our association signals in 45 of the identified loci explained 8% of TB-BMD variation. The novel loci are enriched for genes highly expressed in chondrocytes and affecting bone metabolism, mainly through WNT signaling pathways. On the basis of our results, we anticipate that increasing the sample size of GWAS of bone phenotypes will yield important insights in bone biology.

Additional detailed methods, results and acknowledgements are to be provided in the Online Supplement of the journal containing this publication.

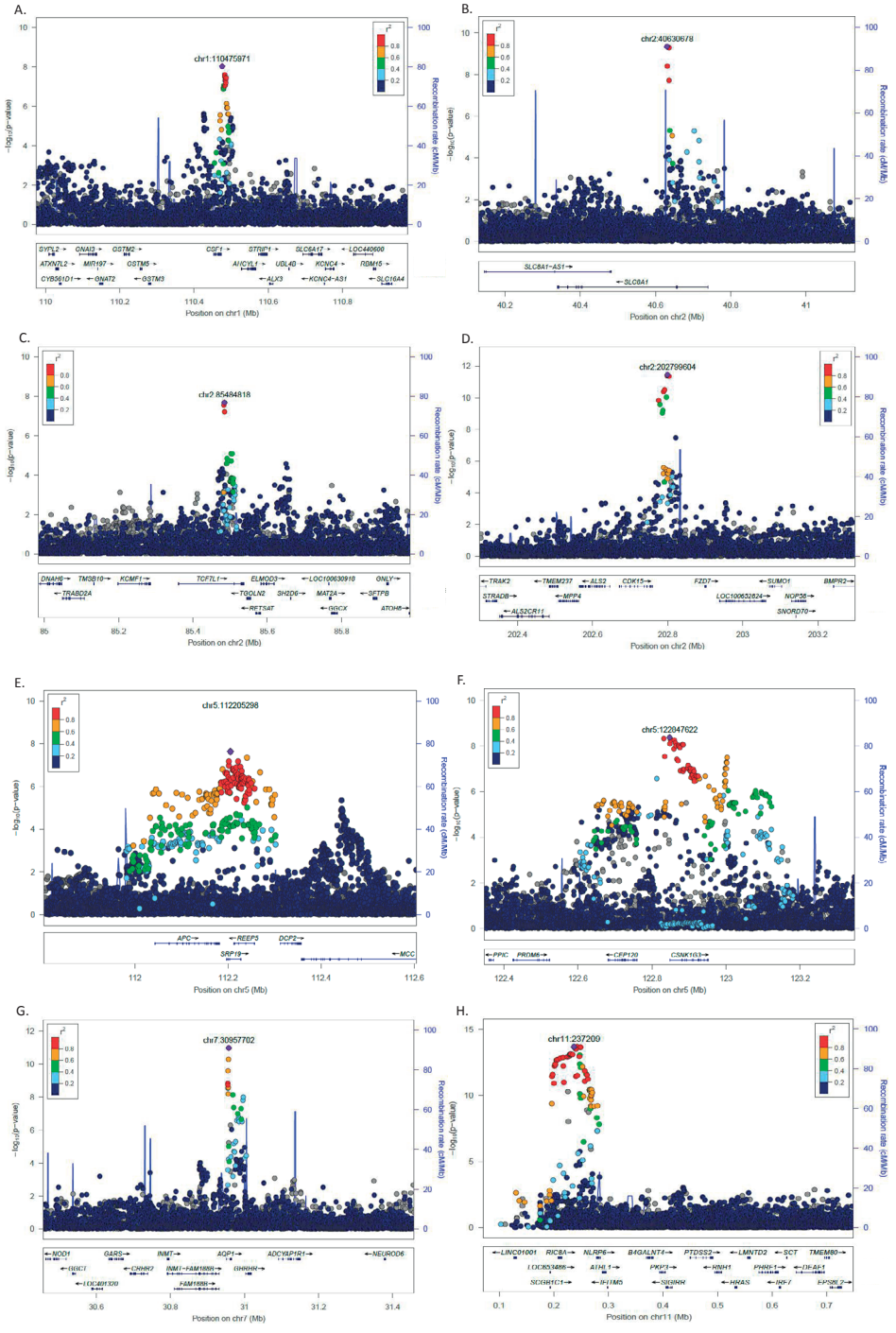
REFERENCES

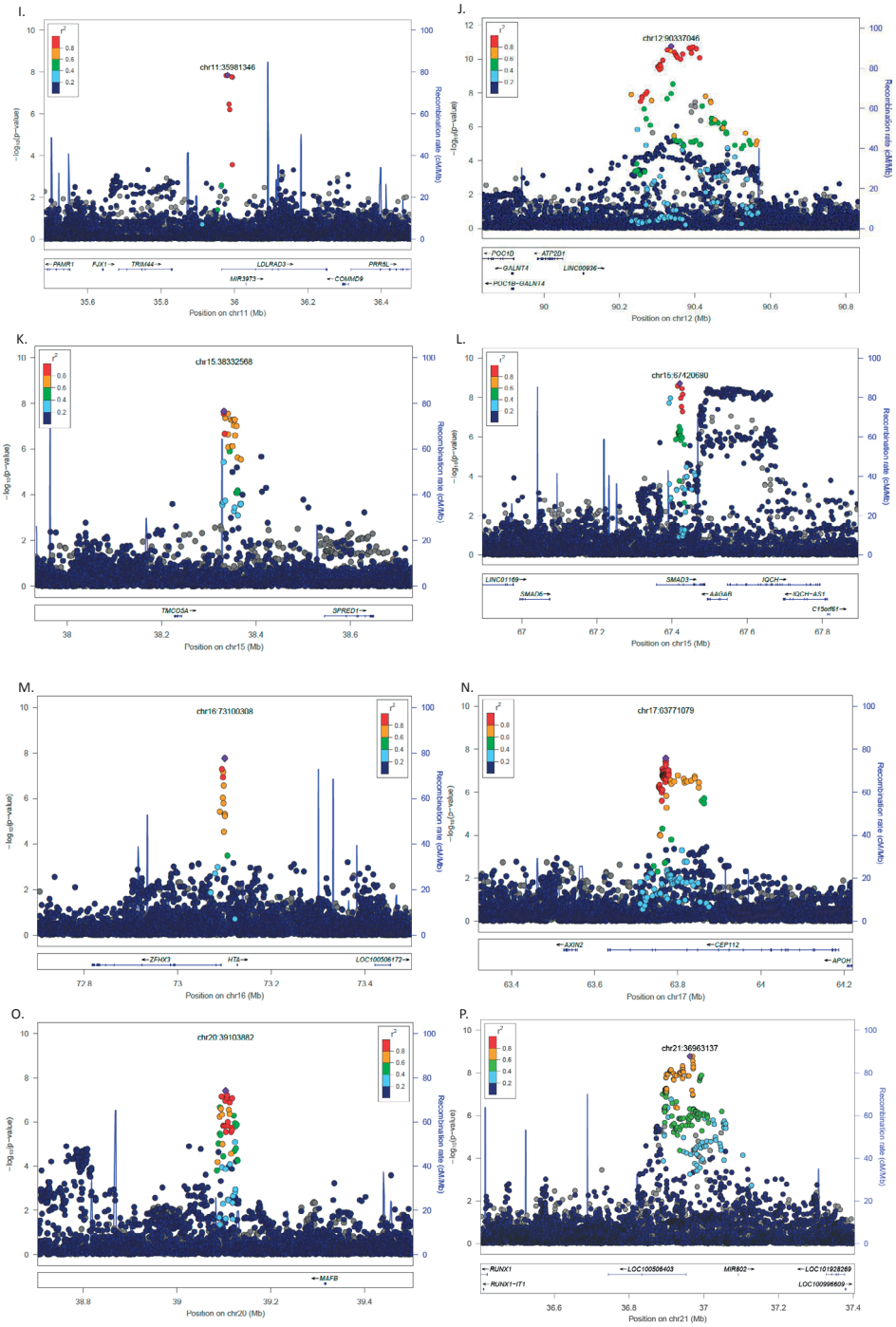
1. Cauley JA. Public health impact of osteoporosis. **J Gerontol A Biol Sci Med Sci**. 2013;68(10):1243-51.
2. Johnell O, Kanis JA, Oden A, et al. Predictive value of BMD for hip and other fractures. **J Bone Miner Res**. 2005;20(7):1185-94.
3. Kanis JA, Oden A, Johnell O, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. **Osteoporosis Int**. 2007;18(8): 1033-46.
4. Sanchez-Riera L, Carnahan E, Vos T, et al. The global burden attributable to low bone mineral density. **Ann Rheum Dis**. 2014;73(9):1635-45.
5. International Society for Clinical Densitometry. Official ISCD Positions – Adult. 2015.
6. International Society for Clinical Densitometry. Official ISCD Positions – Pediatric. 2015.
7. Binkovitz LA, Henwood MJ. Pediatric DXA: technique and interpretation. **Pediatr Radiol**. 2007;37(1): 21-31.
8. Arden NK, Baker J, Hogg C, Baan K, Spector TD. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. **J Bone Miner Res**. 1996;11(4): 530-4.
9. Videman T, Levalahti E, Battie MC, Simonen R, Vanninen E, Kaprio J. Heritability of BMD of femoral neck and lumbar spine: a multivariate twin study of Finnish men. **Journal of Bone and Mineral Research**. 2007;22(9):1455-62.
10. Gueguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, Siest G. Segregation analysis and variance components analysis of bone mineral density in healthy families. **J Bone Miner Res**. 1995;10(12):2017-22.
11. Kemp JP, Medina-Gomez C, Tobias JH, Rivadeneira F, Evans DM. The case for genome-wide association studies of bone acquisition in paediatric and adolescent populations. **BoneKEy Reports**. 2016; [In press]
12. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
13. Zheng HF, Forgetta V, Hsu YH, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. **Nature**. 2015;526(7571):112-7.
14. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **Plos Genet**. 2012;8(7):e1002718.
15. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **Plos Genet**. 2014;10(6):e1004423.
16. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
17. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**. 2012;491(7422):56-65.
18. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. **Genet Epidemiol**. 2010;34(8):816-34.
19. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. **Nat Genet**. 2012;44(8):955-+.
20. 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(6).
21. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. **Nat Genet**. 2007;39(7):906-13.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**. 2007;81(3):559-75.

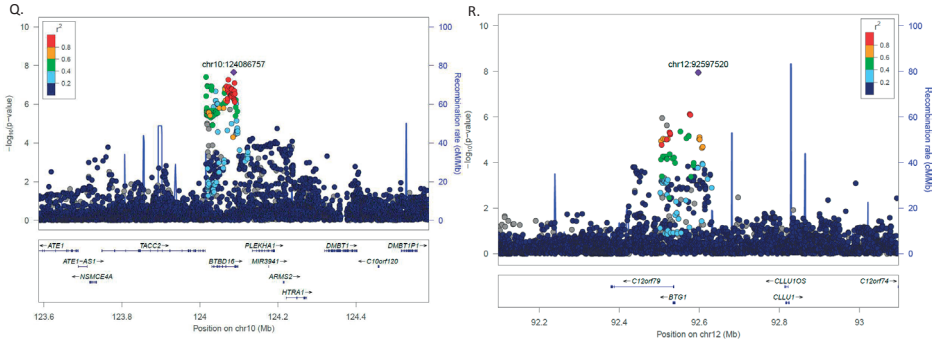
23. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. **Bmc Bioinformatics**. 2010;11.
24. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. **Nat Protoc**. 2014;9(5):1192-212.
25. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. **Nature Genetics**. 2012;44(4):369-U170.
26. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. **Am J Hum Genet**. 2011;88(1):76-82.
27. Paternoster L, Lorentzon M, Lehtimaki T, et al. Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure. **PLoS Genet**. 2013;9(2):e1003247.
28. Moayyeri A, Hsu YH, Karasik D, et al. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. **Hum Mol Genet**. 2014;23(11):3054-68.
29. Yang TL, Guo Y, Liu YJ, et al. Genetic variants in the SOX6 gene are associated with bone mineral density in both Caucasian and Chinese populations. **Osteoporos Int**. 2012;23(2):781-7.
30. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biol**. 2004;5(10):R80.
31. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. **Nat Commun**. 2015;6.
32. Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. **Omics**. 2012;16(5):284-7.
33. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2014.
34. Boyle EI, Weng SA, Gollub J, et al. GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. **Bioinformatics**. 2004;20(18):3710-5.
35. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. **PLoS One**. 2011;6(7):e21800.
36. Gong Y, Slee RB, Fukai N, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. **Cell**. 2001;107(4):513-23.
37. Dobbins DE, Sood R, Hashiramoto A, Hansen CT, Wilder RL, Remmers EF. Mutation of macrophage colony stimulating factor (Csf1) causes osteopetrosis in the tl rat. **Biochem Biophys Res Commun**. 2002;294(5):1114-20.
38. Wu QT, Ma QJ, He CY, et al. Reduced bone mineral density and bone metabolism in aquaporin-1 knockout mice. **Chem Res Chinese U**. 2007;23(3):297-9.
39. Miclea RL, Karperien M, Bosch CA, et al. Adenomatous polyposis coli-mediated control of beta-catenin is essential for both chondrogenic and osteogenic differentiation of skeletal precursors. **BMC Dev Biol**. 2009;9:26.
40. Ruisu K, Kask K, Meier R, et al. Ablation of RIC8A function in mouse neurons leads to a severe neuromuscular phenotype and postnatal death. **PLoS One**. 2013;8(8):e74031.
41. Liakhovitskaia A, Lana-Elola E, Stamateris E, Rice DP, van 't Hof RJ, Medvinsky A. The essential requirement for Runx1 in the development of the sternum. **Dev Biol**. 2010;340(2):539-46.
42. Yang X, Chen L, Xu X, Li C, Huang C, Deng CX. TGF-beta/Smad3 signals repress chondrocyte hypertrophic differentiation and are required for maintaining articular cartilage. **J Cell Biol**. 2001;153(1):35-46.
43. Tijchon E, van Ingen Schenau D, van Opzeeland F, et al. Targeted Deletion of Btg1 and Btg2 Results in Homeotic Transformation of the Axial Skeleton. **PLoS One**. 2015;10(7):e0131481.
44. Yamaguchi T, Sugimoto T, Yano S, et al. Plasma lipids and osteoporosis in postmenopausal women. **Endocr J**. 2002;49(2):211-7.

45. Lara-Castillo N, Johnson ML. LRP receptor family member associated bone disease. **Rev Endocr Metab Disord**. 2015;16(2):141-8.
46. Shen K, Murphy CM, Chan B, et al. Activated protein C (APC) can increase bone anabolism via a protease-activated receptor (PAR)1/2 dependent mechanism. **J Orthop Res**. 2014;32(12):1549-56.
47. Laine CM, Joeng KS, Campeau PM, et al. WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. **N Engl J Med**. 2013;368(19):1809-16.
48. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. **Nat Genet**. 2015;47(10):1114-20.
49. Cecchini MG, Hofstetter W, Halasy J, Wetterwald A, Felix R. Role of CSF-1 in bone and bone marrow development. **Mol Reprod Dev**. 1997;46(1):75-83; discussion -4.
50. Ousingsawat J, Wanitchakool P, Schreiber R, Wuelling M, Vortkamp A, Kunzelmann K. Anoctamin-6 controls bone mineralization by activating the calcium transporter NCX1. **J Biol Chem**. 2015;290(10): 6270-80.
51. Shy BR, Wu CI, Khramtsova GF, et al. Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/beta-catenin signaling. **Cell Rep**. 2013;4(1):1-9.
52. Liu W, Liu Y, Guo T, et al. TCF3, a novel positive regulator of osteogenesis, plays a crucial role in miR-17 modulating the diverse effect of canonical Wnt signaling in different microenvironments. **Cell Death Dis**. 2013;4:e539.
53. Li Y, Dudley AT. Noncanonical frizzled signaling regulates cell polarity of growth plate chondrocytes. **Development**. 2009;136(7):1083-92.
54. Ryan ZC, Craig TA, Filoteo AG, et al. Deletion of the intestinal plasma membrane calcium pump, isoform 1, Atp2b1, in mice is associated with decreased bone mineral density and impaired responsiveness to 1, 25-dihydroxyvitamin D3. **Biochem Biophys Res Commun**. 2015;467(1):152-6.
55. Borton AJ, Frederick JP, Datto MB, Wang XF, Weinstein RS. The loss of Smad3 results in a lower rate of bone formation and osteopenia through dysregulation of osteoblast differentiation and apoptosis. **J Bone Miner Res**. 2001;16(10):1754-64.
56. Kim K, Kim JH, Lee J, et al. MafB negatively regulates RANKL-mediated osteoclast differentiation. **Blood**. 2007;109(8):3253-9.
57. Soung do Y, Kalinowski J, Baniwal SK, et al. Runx1-mediated regulation of osteoclast differentiation and function. **Mol Endocrinol**. 2014;28(4):546-53.

SUPPLEMENTARY MATERIAL




I.
chr11:35981346
PAMR1
FJX1
TRIM44
LDLRAD3
PRRSL
MIR3973
COMMD9
Position on chr11 (Mb)
J.
chr12:90337046
POC1B
ATP2D1
GALNT4
LINC00936
POC1B-GALNT4
Position on chr12 (Mb)
K.
chr15:38332568
TMC03A
SPRED1
Position on chr15 (Mb)
L.
chr15:67420680
LINC01169
SMAD6
SMAD3
IQCH
AAGAB
IQCH-AS1
C15orf61
Position on chr15 (Mb)
M.
chr16:73100308
ZFHX3
HTA
LOC100506172
Position on chr16 (Mb)
N.
chr17:63771079
AXIN2
CEP112
APOH
Position on chr17 (Mb)
O.
chr20:39103882
MAFB
Position on chr20 (Mb)
P.
chr21:36963137
RUNX1
LOC100506403
MIR802
LOC101928269
RUNX1-IT1
LOC100996609
Position on chr21 (Mb)
r²
0.8
0.6
0.4
0.2
Recombination rate (cM/Mb)



Supplementary Figure 1. Regional association plots for each of the novel BMD loci. SNPS are plotted by position on hg19. Blue peaks indicate recombination rates. The SNPs surrounding the most significant SNP are color coded to reflect their LD with this SNP, pairwise r^2 values from the CEU 1000 Genomes population except for figure I. based on the YRI population. Figures Q. and R. represent GWS in the only CEU meta-analysis **A. 1p13.3, B. 2p22.1, C. 2p11.2, D. 2q33.1, E. 5q22.2, F. 5q23.2, G. 7p14.2, H. 11p15.5, I. 11p13, J. 12q21.3, K. 15q14, L. 15q22.33, M. 16q22.3, N. 17q24.1, O. 20q12, P. 21q22.12, Q.10q26.13, R. 12q21.33**

Supplementary Table 1. Descriptives of the participating cohorts

STUDY	RS1	RS2	RS3	ALSPAC	deCode	OPRA cohort	HABC	BPROOF	GOOD	ERF	FHS
GWAS SAMPLE NUMBER	**6,291**	**2,157**	**3,048**	8,165	**90,298**	**993**	**2,658**	**2,789**	**938**	**2,786**	**8,481**
Ethnicity	Caucasian	Caucasian	Caucasian	Caucasian	Icelandic	Caucasian	European Ancestry (EA) and African American (AA)	Caucasian	Caucasian	Caucasian	Caucasian
GENOTYPING											
Genotyping platform & SNP panel	Ilumina HapMap 550	Ilumina HapMap 550	Ilumina HapMap 550 & 610	Illumina HumanHap 550K	Illumina Hapmap and Omni chips	Illumina Omni Express Exome chip	Illumina 1M	Omniexpress	Ilumina HapMap 550	Illumina 310, 370, 610K	Affymetrix Nsp, Sty and 50K gene centric
Genotyping calling algorithm	Genome-Studio	Genome-Studio	Genome-Studio	Laboratory Corporation of America	Genome-Studio	Genome Studio	Genome-Studio	Genome-Studio	Genome-Studio	Bead-studio/ Genome-Studio	
SAMPLE QC											
Call rate	≤97.5% Excluded	≤97.5% Excluded	≤97.5% Excluded	≤97% Excluded	<97% excluded	<95% Excluded	≤95% Excluded	≤97.5% Excluded	≤97.5% Excluded	≤95% Excluded	≤95% Excluded
Heterozygosity	F-values < -0.096	F-values < -0.096	F-values < -0.096	(F< .32) or (F> .345)	None	None	6sd from population mean	F-values < -0.096	F-values < -0.096	FDR < 1%	None
Ethnic outliers excluded	4SD CEU HapMap	4SD CEU HapMap	4SD CEU HapMap	Individuals with non European Ancestry	Non-icelandic individuals excluded	No exclusion under this criteria	6sd from reference popualtion mean for C1 and C2	4SD CEU HapMap	4SD CEU HapMap	ethnic outliers removed	None
Stratification Adjustment	4 PCs	4 PCs	4 PCs	None	None	None	2 (EA) / 10 (AA) PCs	4 PCs	4 PCs	Relationship matrix in probabel	None
Incorrect Gender Allocation	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	None

NEO STUDY	COP-SAC2000	COP-SAC2010	GEN-ERATION R STUDY	TwinsUK	1982 Pelotas	PANIC	RAINE STUDY	ORCADES	CHOP's	MrOS	SOF
5,744	**271**	**446**	**5733 individuals with GWAS data**	**5654 individuals with GWAS data**	**3736 individuals with GWAS data**	469	**1494 individuals with GWAS data**	**2215 individuals with GWAS data**	1885 individuals with GWAS data	**5130**	**3625**
Caucasian	Danish, European Descent	Danish, European Descent	Multiethnic	North-Western European	Multiethnic	Caucasian	Mainly Caucasian, a small number of admixed	European descent	Multiethnic	Non-hispanic white	Non-hispanic white
Illumina Human-CoreExom-e24v1-0	Illumina Omni-Expres-sExome	Illumina Omni-Expres-sExome	Illumina Hapmap 610K	Illumina Human-Hap300, Human-Hap610Q and 1M-Duo	HumanOmni2.5-8v1	Illumina Custom Infinium chemistry, Array - Cardio-Metabo_Chip_11395247_A, Illumina HumanCoreExome-12v1-0_B	Illu-mina Hu-man660W Quad Array	Illumina Hap300, Omni1, OmniX	Illumina Infinium™ II OMNI Express plus Exome BeadChip	Illumina HumanOmni1_Quad_v1-0 B	Illumina HumanOmni1_Quad_v1-0 B
Genome Studio	Genome Studio	Genome Studio	Genome Studio	Illuminus	Genome Studio	BeadStudio version 3.3.7, Gentrain version 1.0 (Metabochip), GenomeStudio (CoreExome)	Bead Studio	Bead Studio, Genome Studio	Genome Studio	BeadStudio	BeadStudio
≤98% Excluded	Individu-als with call rate <95% were excluded	Individu-als with call rate <95% were excluded	≤97.5% Excluded	<98% excluded	≤97% Excluded	≤95% excluded	≤97% Excluded	≤98% Excluded	≤95% Excluded	≤97% Excluded	<97% excluded
F>0.35-0.45	F<0.0375 &F >0.27	F<0.0375 &F >0.27	Individuals with F-values < -0.096 were excluded	Individuals with het-erozygosity across all SNPs >2SD from the sample mean excluded	Individuals with heterozygos-ity rates outside the range of median±1.5*IQR were excluded	All included.	h-values <0.3	NA	No exclu-sion under this criteria	Not applied	Not applied
+/- 3.5 SD of NEO mean	Yes	Yes	No exclu-sion under this criteria	Indi-viduals with evidence of non-European ancestry excluded	No exclusion un-der this criteria	Non-caucasians were excluded	No exclu-sion	No exclu-sion under this criteria	No exclu-sion under this criteria	4 SD from mean of PC1 or PC2	4 SD from mean of PC1 or PC2
3 PCs	No	No	Correction for 20 PCs generated by MDS in Plink	Corrected for 2 PCs gener-ated using Eigenstrat	20 PCs	None	10PCs		10 PCs	4 PCs	4 PCs
Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromosome heterozygosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromo-some heterozy-gosity	Self reported gender vs. X chromosome heterozy-gosity	Self reported gender vs. X chromosome heterozy-gosity

Supplementary Table 1 (continued)

STUDY	RS1	RS2	RS3	ALSPAC	deCode	OPRA cohort	HABC	BPROOF	GOOD	ERF	FHS
Cryptic Relatedness	PI HAT>0.3	PI HAT>0.3	PI HAT>0.3	PI HAT> 0.11	Related individuals included	PI HAT > 0.2	exclusion at 12.5% sharing	PI HAT>0.3	PI HAT>0.3	Related individuals included	Related individuals included
SNP QC (prior to imputation)											
MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 1 % MAF	No exclusion on MAF performed	SNPs with ≤ 1 % MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 0.5 % MAF	SNPs with ≤ 0.5 % MAF
HWE	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 0.001	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6
Call rate	SNPs with a callrate ≤ 98%	SNPs with a callrate ≤ 98%	SNPs with a callrate ≤ 98%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 98%	SNPs with a callrate ≤ 98%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 97%
IMPUTATION AUTOSOMAL											
Imputation software	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	In-house software	IMPUTE version 2.1.2	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (MACH) Imputation (Minimac)
Phased Reference Data	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3
SNPs used for Imputation	502,668	490,409	517,658	464,311	699,107		899,818	730,525	502,668	678,524	389,956
Total number of Imputed SNPs	30,072,738	30,072,738	30,072,738	30,072,738	21106834		36,890,632	30,072,738	30,072,738	29,974,334	8,221,074
ANALYSIS OF SNP ASSOCIATIONS											
Association Software	MACH2qtl	MACH2qtl	MACH2qtl	MACH2qtl	SNPTEST	GEMMA	MACH2qtl	MACH2qtl	MACH2qtl	ProbABEL	In House
DXA MACHINE SPECIFICATION											
Device	GE-Lunar (Lunar Prodigy)	GE-Lunar (Lunar Prodigy)	GE-Lunar (Lunar Prodigy)	GE-Lunar (Lunar Prodigy)	HologicQ-DR4500A	GE-Lunar DPX-L	Hologic QDR 4500	GE-Lunar (Lunar Prodigy)	GE-Lunar (Lunar Prodigy)	GE-Lunar-DPX-L	GE-Lunar-DPX-L
Total Number Scanned	2436	758	2488	5,434	24,397	931	2658	968	938	2,786	5,627

NEO STUDY	COP-SAC2000	COP-SAC2010	GEN-ERATION R STUDY	TwinsUK	1982 Pelotas	PANIC	RAINE STUDY	ORCADES	CHOP's	MrOS	SOF
PI HAT (>0.25)	removed with IBM calcula-tion	removed with IBM calcula-tion	287 close related individuals (pi^>0.3), from which 28 indi-viduals with a twin in the dataset was excluded	Unre-lated but estimated pi hat > 0.12 (19 samples excluded)	IBD≥0.1	Twins were excluded	PI HAT> 0.19	excess IBS incompat-ible with pedigree	PI HAT> 0.45	Only one member of a family	Only one member of a family
NA	SNPs with ≤ 1 % MAF	SNPs with ≤ 1 % MAF	SNPs with ≤ 1 % MAF	SNPs with MAF <1% were excluded	SNPs with ≤ 0.1 % MAF	SNPs with ≤ 1 % MAF	SNPs with ≤ 1 % MAF	SNPs with ≤ 1 % MAF	SNPs with ≤ 0.5% MAF	SNPs with ≤ 1 % MAF	SNPs with ≤ 1 % MAF
SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-6	SNPs with P value < 1E-7	SNPs with P value < 1E-6	SNPs with P value < 6E-7	SNPs with P value < 1E-6	No exclu-sion under this criteria	SNPs with P value < 1E-4	SNPs with P value < 1E-4
SNPs with a callrate ≤ 98%	SNPs with a call rate < 95%	SNPs with a call rate < 95%	SNPs with a callrate ≤ 98%	MAF >5% C.R <97% or MAF 1-5% C>R <99%	SNPs with a call rate < 95%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 97%	SNPs with a callrate ≤ 95%	SNPs with a callrate ≤ 97%	SNPs with a callrate ≤ 97%
IMPUTE2	SHA-PEIT2 & IMPUTE2	SHA-PEIT2 & IMPUTE2	Phasing genotypes (MACH) Imputation (Minimac)	SHAPEIT2 & IMPUTE2	IMPUTE2 v2.3.1	SHAPEIT, IM-PUTE2.	Phasing genotypes (MACH) Imputation (Minimac)	Phasing genotypes (ShapeIt2) Imputation (Impute2)	Phasing genotypes (SHAPEIT) Imputation (IMPUTE2)	IMPUTE2	IMPUTE2
1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	1000G Phase I V. 3	UK10K/1000G Phase I V.3	UK10K/1000G Phase I V.3
349,605	657,699	657,699	489,878	303,940 -878,319	2,073,816	390,669	521,554	287,208 (Hap300) 615658 (Omni)	739,284	740,713	740,713
28,427,527	27,446,975	27,446,975	30,072,738	37,428,090	37,759,470	12,250,634	29,540,342	37,500,000	15,689,930	41,992,162	41,992,162
SNPTEST	Quicktest v0.94	Quicktest v0.94	MACHqtl	Genome-wide Effi-cient Mixed Model Association (GEMMA)	SNPTEST v2.5	SNPTEST2	ProbAbel	ProbABEL	SNPTEST2	GEMMA	GEMMA
Hologic Discovery A, Tromp Medical BV, Castricum, The Nether-lands	Lunar iDXA densitom-eter (GE Health-care, Fairfield, Connecti-cut, USA)	Lunar iDXA densitom-eter	GE-Lunar iDXA (Lunar Prodigy)	Ho-logic QDR 4500W	Lunar GE® (Prodigy)	Lunar GE® (Prodigy)	Norland XR-36 den-sitometer	Hologic QDR-4500	Hologic bone densi-tometers	Hologic QDR 4500	Hologic QDR 1000
802	271	446	6,490	7,711	3,438	469	1183	1,302	1885	5955	609

Supplementary Table 2 Anthopometric description of the individuals in each of the participating studies

Study	Trait	All					
		n	mean	SD	median	min	max
DECODE	Age (yrs)	6794	54.825	16.745	55.400	15.100	92.100
	Weight (kg)	6794	74.520	15.042	73.000	35.700	129.300
	Height (cm)	6794	168.782	8.756	168.000	116.500	203.000
	(TB/TBLH)BMD (g/cm^2)	6794	1.067	0.129	1.072	0.669	1.710
BPROOF	Age (yrs)	968	74.594	5.547	74.000	66.000	96.000
	Weight (kg)	968	78.094	13.152	77.500	46.700	133.500
	Height (cm)	968	170.024	8.879	169.650	145.600	194.700
	(TB/TBLH)BMD (g/cm^2)	968	1.117	0.152	1.119	0.675	1.638
FHS	Age (yrs)	2386	53.532	12.836	52.000	24.000	92.000
	Weight (kg)	2386	86.193	14.270	84.370	48.535	148.327
	Height (cm)	2386	175.674	7.033	175.895	151.765	199.390
	(TB/TBLH)BMD (g/cm^2)	2386	1.264	0.098	1.264	0.905	1.598
MROS	Age (yrs)	4670	73.951	5.955	73.000	65.000	100.000
	Weight (kg)	4670	83.421	13.003	82.000	50.800	136.400
	Height (cm)	4670	174.494	6.633	174.450	147.150	198.900
	(TB/TBLH)BMD (g/cm^2)	4670	1.167	0.127	1.158	0.765	2.047
SOF	Age (yrs)	314	79.557	3.949	79.000	74.000	97.000
	Weight (kg)	314	66.716	12.443	65.650	37.000	115.700
	Height (cm)	314	158.223	5.902	158.575	143.050	175.350
	(TB/TBLH)BMD (g/cm^2)	314	1.014	0.122	1.011	0.749	1.531
CHOP	Age (yrs)	1320	9.568	3.013	9.496	5.000	14.995
	Weight (kg)	1320	35.244	14.001	31.800	15.300	78.400
	Height (cm)	1320	137.261	18.588	136.216	101.867	182.400
	(TB/TBLH)BMD (g/cm^2)	1320	0.712	0.148	0.693	0.416	1.168
Orcades	Age (yrs)	1162	56.408	13.076	58.040	25.720	87.340
	Weight (kg)	1162	77.329	15.435	75.450	45.600	148.400
	Height (cm)	1162	166.473	9.160	165.600	140.600	199.400
	(TB/TBLH)BMD (g/cm^2)	1162	1.059	0.130	1.064	0.688	1.468
PANIC	Age (yrs)	447	7.640	0.375	7.630	6.770	9.000
	Weight (kg)	447	26.830	4.815	26.020	15.240	51.440
	Height (cm)	447	128.700	5.597	129.200	110.700	144.700
	(TB/TBLH)BMD (g/cm^2)	447	0.720	0.048	0.717	0.610	0.960
PELOTAS	Age (yrs)	2688	30.079	0.188	30.076	29.430	31.136
	Weight (kg)	2688	74.533	16.016	72.350	33.700	119.500
	Height (cm)	2688	167.232	8.969	166.800	140.600	191.800
	(TB/TBLH)BMD (g/cm^2)	2688	1.217	0.101	1.212	0.917	1.568
COPSAC2000	Age (yrs)	271	6.894	0.718	6.550	6.354	10.227

Supplementary Table 2 (continued)

Study	Trait	All					
		n	mean	SD	median	min	max
	Weight (kg)	271	24.590	4.251	24.000	15.500	40.600
	Height (cm)	271	124.400	6.162	124.000	108.200	143.100
	(TB/TBLH)BMD (g/cm^2)	271	0.583	0.053	0.577	0.462	0.777
COPSAC2010	Age (yrs)	436	3.458	0.320	3.400	2.650	4.760
	Weight (kg)	436	15.880	1.930	15.800	11.700	23.500
	Height (cm)	436	99.650	5.064	99.650	59.900	115.800
	(TB/TBLH)BMD (g/cm^2)	436	0.450	0.033	0.449	0.344	0.546
Generation R	Age (yrs)	4070	6.215	0.501	6.072	4.890	9.080
	Weight (kg)	4070	23.080	4.085	22.370	13.660	54.950
	Height (cm)	4070	119.600	5.909	119.200	99.000	147.000
	(TB/TBLH)BMD (g/cm^2)	4070	0.555	0.049	0.551	0.392	0.784
TUK	Age (yrs)	4650	48.075	13.885	50.000	16.000	84.000
	Weight (kg)	4650	68.100	13.046	66.000	35.100	128.100
	Height (cm)	4650	163.309	7.224	163.000	141.000	204.000
	(TB/TBLH)BMD (g/cm^2)	4650	1.151	0.117	1.152	0.732	1.602
RAINE	Age (yrs)	4070	6.215	0.501	6.072	4.890	9.080
	Weight (kg)	4070	23.080	4.085	22.370	13.660	54.950
	Height (cm)	4070	119.600	5.909	119.200	99.000	147.000
	(TB/TBLH)BMD (g/cm^2)	4070	0.555	0.049	0.551	0.392	0.784
OPRA	Age (yrs)	887	75.220	0.140	75.189	75.030	76.000
	Weight (kg)	887	66.881	10.596	66.000	41.000	109.000
	Height (cm)	887	160.490	5.670	160.000	140.000	179.000
	(TB/TBLH)BMD (g/cm^2)	887	1.007	0.098	1.002	0.718	1.422
HABC_EU	Age (yrs)	1585	73.750	2.838	73.000	69.000	80.000
	Weight (kg)	1585	74.270	14.328	73.200	40.800	134.500
	Height (cm)	1585	166.900	9.307	166.600	141.600	194.800
	(TB/TBLH)BMD (g/cm^2)	1585	1.071	0.135	1.068	0.730	1.605
HABC_AA	Age (yrs)	1073	73.420	2.872	73.000	68.000	80.000
	Weight (kg)	1073	77.660	14.898	77.100	33.500	133.800
	Height (cm)	1073	165.300	9.478	164.600	137.000	200.700
	(TB/TBLH)BMD (g/cm^2)	1073	1.115	0.146	1.105	0.763	1.706
ERF	Age (yrs)	2786	48.958	14.345	49.518	16.653	86.501
	Weight (kg)	2631	75.457	15.414	74.200	41.900	161.000
	Height (cm)	2631	167.591	9.398	167.000	141.000	196.500
	(TB/TBLH)BMD (g/cm^2)	2786	1.151	0.113	1.148	0.745	1.557
RS1	Age (yrs)	2436	75.270	5.930	74.400	65.230	98.490
	Weight (kg)	2436	76.080	13.170	75.300	36.800	138.000

Supplementary Table 2 (continued)

Study	Trait	All					
		n	mean	SD	median	min	max
	Height (cm)	2436	166.600	9.219	166.000	135.300	196.000
	(TB/TBLH)BMD (g/cm^2)	2436	1.109	0.124	1.108	0.726	1.473
RS2	Age (yrs)	750	67.160	6.457	65.170	49.470	91.190
	Weight (kg)	750	77.230	12.770	77.200	46.400	123.200
	Height (cm)	750	168.800	8.830	168.400	146.900	192.400
	(TB/TBLH)BMD (g/cm^2)	750	1.146	0.114	1.147	0.770	1.467
RS3	Age (yrs)	2483	57.310	6.940	57.100	45.740	97.220
	Weight (kg)	2483	80.230	15.439	78.900	31.500	153.200
	Height (cm)	2483	170.500	9.511	169.800	145.000	201.000
	(TB/TBLH)BMD (g/cm^2)	2483	1.178	0.107	1.176	0.845	1.515

Supplementary Table 3 Known independent markers associated with bone phenotypes. All effect sizes (β) are reported for the minor allele (MA). MAF=Minor Allele Frequency, TB= total body BMD, assessed in this study, LS=lumbar spine BMD, assessed in Zheng et al. [13], FN=Femoral Neck BMD, assessed in Zheng at al. [13]. Phenotype for which association was previously reported and the correspondent reference are given.

Chr	Pos	rsID	Gene	Phenotype	Reference	MA	MAF	β_TB	P_TB	β_LS*	P_LS*	β_FN*	P_FN*
1	22490724	rs7521902	*WNT4*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.240	-0.072	8.32E-20	-0.034	0.002	-0.037	6.60E-05
1	22711473	rs6426749	*ZBTB40*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.182	0.100	1.21E-33	0.088	1.15E-13	0.082	5.9E-16
1	68639385	rs17482952	*WLS*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.077	-0.063	2.75E-08	-0.045	0.008	-0.043	0.003
1	68647716	rs12407028	*WLS*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.382	-0.051	2.56E-13	-0.063	4.33E-12	-0.046	5.07E-09
1	172199573	rs479336	*DNM3*	FN-BMD	Estrada et al. (12)	G	0.289	0.029	1.92E-04	0.034	9.74E-04	0.043	9.08E-07
1	240597214	rs9287237	*FMN2*	Trabecular vBMD	Paternoster et al (27)	T	0.176	0.051	3.12E-10	0.037	0.002	0.039	1.47E-04
2	42250549	rs7584262	*LOC91461*	FN-BMD	Estrada et al. (12)	T	0.237	0.034	2.31E-05	0.009	0.392	0.044	1.34E-06
2	54659707	rs4233949	*SPTBN1*	LS-BMD	Estrada et al. (12)	C	0.365	0.022	1.67E-03	0.055	1.49E-03	0.044	0.002
2	112500035	rs17040773	*ANAPC1*	FN-BMD	Estrada et al. (12)	C	0.211	-0.016	0.040	-0.004	0.714	-0.023	0.012
2	119038598	rs1878526	*INSIG2*	LS-BMD	Estrada et al. (12)	A	0.225	0.021	0.248	0.038	4.29E-04	0.006	0.529
2	119154872	rs6542457	*EN1*	LS-BMD	Zheng et al. (13)	C	0.096	0.049	4.30E-05	0.083	6.53E-06	-0.005	0.757
2	119545994	rs11692564	*EN1*	LS-BMD	Zheng et al. (13)	T	0.014	0.179	2.27E-08	0.238	4.1E-09	0.116	7.23E-04
2	166601046	rs1346004	*GALNT3*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.465	-0.049	5.56E-16	-0.052	8.71E-09	-0.058	7.15E-14
3	41128564	rs430727	*CTNNB1*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.453	-0.061	3.00E-25	-0.056	5.34E-10	-0.061	2.02E-15
3	113370010	rs1026364	*KIAA2018*	FN-BMD	Estrada et al. (12)	T	0.370	0.030	7.58E-04	0.013	0.171	0.024	0.003
3	156555984	rs344081	*LEKR1*	LS-BMD	Estrada et al. (12)	C	0.168	-0.051	2.68E-08	-0.047	4.28E-04	-0.035	0.002
4	994414	rs3755955	*IDUA*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.154	-0.066	2.16E-11	-0.049	5.05E-05	-0.044	2.87E-05
4	88773849	rs6532023	*MEPE*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.341	0.054	3.32E-18	0.048	2.34E-07	0.034	2.78E-05
5	88376061	rs1366594	*MEF2C*	FN-BMD	Estrada et al. (12)	C	0.494	-0.047	3.84E-12	-0.007	0.435	-0.079	5.44E-25
6	21384613	rs9466056	*CDKAL1/SOX4*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.386	-0.034	2.20E-06	-0.039	1.81E-05	-0.038	8.86E-07
6	44639184	rs11755164	*SUPT3H/ RUNX2*	LS-BMD	Estrada et al. (12)	T	0.408	-0.047	1.74E-08	-0.024	0.010	-0.013	0.108

Supplementary Table 3 (continued)

Chr	Pos	rsID	Gene	Phenotype	Reference	MA	MAF	β_TB	P_TB	β_LS*	P_LS*	β_FN*	P_FN*
6	127167072	rs13204965	*RSPO3*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.222	-0.064	6.65E-15	-0.039	2.81E-04	-0.052	1.39E-08
6	133350936	rs3012465	*EYA4*	Skull BMD	Kemp et al. (15)	A	0.325	-0.024	4.65E-04	0.016	0.079	0.025	0.002
6	151895456	rs6909279	*ESR1*	cortical vBMD	Paternoster et al. (27)	G	0.444	-0.069	1.05E-23	-0.055	1.27E-09	-0.051	8.07E-11
6	151907748	rs4869742	*C6orf97*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.341	-0.069	1.59E-19	-0.061	2.79E-08	-0.052	1.75E-08
6	151946658	rs7751941	*C6orf97*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.204	-0.038	1.09E-05	-0.059	9.82E-08	-0.031	0.001
7	37938422	rs10226308	*TXNDC3*	LS-BMD	Estrada et al. (12)	G	0.178	0.038	1.74E-04	0.059	2.48E-07	0.025	0.012
7	38128326	rs6959212	*STARD3NL*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.338	-0.047	2.42E-09	-0.066	2.55E-12	-0.033	5.12E-05
7	96120675	rs4727338	*SLC25A13*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.320	-0.073	9.30E-25	-0.059	5.79E-10	-0.063	5E-15
7	120742980	rs148771817	*WNT16*	FA-BMD	Zheng et al. (13)	T	0.009	0.135	0.053	0.136	0.006	0.011	0.798
7	120785064	rs13245690	*CPED1*	LS-BMD	Estrada et al. (12)	G	0.366	-0.054	2.48E-21	-0.028	0.002	-0.023	0.003
7	120903815	rs4609139	*CPED1*	TB-BMD	Medina-Gomez et al. (14)	T	0.344	-0.037	4.77E-12	-0.015	0.117	-0.013	0.114
7	120974765	rs3801387	*WNT16*	FN-BMD	Estrada et al. (12)	G	0.271	0.135	1.23E-76	0.073	1.7E-13	0.054	3.26E-10
7	150919829	rs7812088	*ABCF2*	FN-BMD	Estrada et al. (12)	A	0.116	0.069	2.94E-11	0.035	0.010	0.044	1.61E-04
8	71591203	rs7017914	*XKR9*	Fem FN-BMD	Estrada et al. (12)	G	0.474	-0.005	0.463	0.011	0.219	-0.016	0.045
8	120007420	rs2062377	*TNFRSF11B*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.406	0.072	2.36E-22	0.081	6.61E-19	0.060	1.64E-14
9	133478827	rs7851693	*FUBP3*	FN-BMD	Estrada et al. (12)	G	0.350	-0.047	2.54E-12	-0.017	0.074	-0.040	9.62E-07
10	28479942	rs3905706	*MPP7*	LS-BMD	Estrada et al. (12)	T	0.244	0.004	0.380	0.055	7.69E-07	-0.014	0.157
10	54427825	rs1373004	*MBL2/DKK1*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.162	-0.077	3.86E-12	-0.056	1.10E-04	-0.045	2.81E-04
10	79401316	rs7071206	*KCNMA1*	LS-BMD	Estrada et al. (12)	C	0.208	0.018	0.061	0.053	8.60E-07	-0.014	0.122
10	101813802	rs7084921	*CPN1*	FN-BMD	Estrada et al. (12)	T	0.410	0.028	9.73E-05	0.018	0.043	0.026	9.35E-04
11	15710084	rs7108738	*SOX6*	FN-BMD	Estrada et al. (12)	G	0.185	0.053	1.01E-11	0.043	2.05E-04	0.083	8.07E-17
11	16296412	rs1347677	*SOX6*	Hip BMD	Yang TL (29)	C	0.211	0.050	3.11E-08	0.030	0.006	0.039	3.37E-05
11	27505677	rs10835187	*LIN7C*	LS-BMD	Estrada et al. (12)	C	0.485	0.046	2.72E-11	0.026	0.004	0.009	0.275

Supplementary Table 3 (continued)

Chr	Pos	rsID	Gene	Phenotype	Reference	MA	MAF	β_TB	P_TB	β_LS*	P_LS*	β_FN*	P_FN*
11	30951674	rs163879	*DCDC5*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.339	0.038	2.55E-06	0.039	6.17E-05	0.018	0.026
11	46722221	rs7932354	*ARHGAP1*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.348	0.050	4.66E-10	0.036	4.23E-04	0.041	1.64E-06
11	68201295	rs3736228	*LRP5*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.143	-0.094	4.14E-26	-0.078	2.9E-10	-0.049	4.79E-06
11	68263370	rs12272917	*PPP6R3*	SK-BMD	Kemp et al. (15)	C	0.247	-0.077	1.11E-24	-0.074	4.58E-13	-0.045	2.89E-07
11	86853997	rs597319	*TMEM135*	BUA and VOS	Moayyeri et al (28)	G	0.328	-0.053	8.68E-14	-0.042	1.25E-05	-0.026	0.002
12	1638171	rs2887571	*ERC1/WNT5B*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.239	0.036	2.31E-07	0.034	8.57E-04	0.022	0.012
12	28017159	rs7953528	*KLHDC5*	FN-BMD	Estrada et al. (12)	A	0.169	0.019	0.067	-0.011	0.349	0.038	1.35E-04
12	49474605	rs12821008	*DHH*	LS-BMD	Estrada et al. (12)	T	0.374	0.029	1.58E-05	.	.	0.012	0.525
12	53727955	rs2016266	*SP7*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.339	0.048	1.23E-11	0.056	6.7E-09	0.039	1.89E-06
12	54417576	rs736825	*HOXC6*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.376	-0.019	6.89E-03	-0.062	2.72E-11	-0.043	8.70E-08
12	107367225	rs1053051	*C12orf23*	FN-BMD	Estrada et al. (12)	C	0.492	0.035	7.43E-08	0.024	0.006	0.023	0.003
13	42951449	rs9533090	*AKAP11*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.450	-0.069	2.39E-18	-0.082	5.15E-20	-0.035	6.19E-06
14	91442779	rs1286083	*RPS6KA5*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.199	0.052	1.05E-10	0.065	9.99E-09	0.038	7.66E-05
14	93114787	rs754388	*RIN3*	TB-BMD, LL-BMD	Estrada et al. (12)	A	0.169	-0.001	0.139	-0.007	0.543	-0.004	0.670
14	103883633	rs11623869	*MARK3*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.338	-0.025	5.07E-04	-0.020	0.029	-0.029	2.96E-04
16	375782	rs9921222	*AXIN1*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.484	-0.051	8.08E-13	-0.053	3.16E-09	-0.050	6.36E-11
16	1532463	rs13336428	*C16orf38*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.443	-0.013	1.00E-03	-0.027	0.003	-0.039	8.33E-07
16	15129459	rs4985155	*NTAN1*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.345	0.017	0.021	0.035	2.09E-04	0.022	0.007
16	50986308	rs1564981	*CYLD*	LS-BMD	Estrada et al. (12)	A	0.497	-0.033	5.57E-05	-0.044	7.95E-07	-0.032	2.38E-05
16	51021803	rs1566045	*SALL1*	FN-BMD	Estrada et al. (12)	C	0.201	0.021	1.15E-03	0.012	0.317	0.043	2.90E-05
16	86710660	rs10048146	*FOXL1*	FN-BMD, LS-BMD	Estrada et al. (12)	G	0.183	-0.044	3.24E-07	-0.064	4.46E-08	-0.056	2.31E-08
17	2068932	rs4790881	*SMG6*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.293	-0.039	8.21E-07	-0.035	3.41E-04	-0.050	2.92E-09
17	41798824	rs4792909	*SOST*	FN-BMD, LS-BMD	Estrada et al. (12)	T	0.403	0.051	1.86E-09	0.047	3.07E-07	0.048	1.52E-09

Supplementary Table 3 (continued)

Chr	Pos	rsID	Gene	Phenotype	Reference	MA	MAF	β_TB	P_TB	β_LS*	P_LS*	β_FN*	P_FN*
17	42225547	rs227584	*C17orf53*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.350	0.039	3.34E-07	0.043	4.05E-05	0.048	2.69E-08
17	43977827	rs1864325	*MAPT*	LS-BMD	Estrada et al. (12)	T	0.197	-0.022	0.022	-0.052	2.48E-04	-0.019	0.103
17	69949016	rs7217932	*SOX9*	FN-BMD	Estrada et al. (12)	A	0.485	0.022	1.32E-03	0.006	0.501	0.033	1.89E-05
18	13708574	rs4796995	*C18orf19*	FN-BMD	Estrada et al. (12)	G	0.368	-0.020	0.010	-0.025	0.006	-0.037	2.56E-06
18	60054857	rs884205	*TNFRSF11A*	FN-BMD, LS-BMD	Estrada et al. (12)	A	0.239	-0.053	5.96E-12	-0.062	2.77E-09	-0.042	2.71E-06
19	33599127	rs10416218	*GPATCH1*	LS-BMD	Estrada et al. (12)	C	0.302	0.027	1.48E-03	0.070	8.32E-09	0.042	2.93E-05
20	10639988	rs3790160	*JAG1*	FN-BMD, LS-BMD	Estrada et al. (12)	C	0.497	-0.042	4.37E-08	-0.051	1.50E-08	-0.029	1.94E-04

Supplementary Table 4. Prioritized genes in the associated TB-BMD loci. Analysis was carried out including all markers GWS associated with TB-BMD ($P<5x10^{-8}$) using DEPICT.

No. Genes	Locus Location	Gene Prioritized	Top cis eQTL SNP	False discovery rate
5	chr1:22350487-22470462	*WNT4*	rs6688182	<0.05
1	chr1:22778344-22857650	*ZBTB40*	-	<0.05
1	chr1:68564142-68698803	*WLS*	rs11209243;rs11209243; rs11209243;rs2039153;rs6691613	<0.05
3	chr11:15987995-17035990	*SOX6*	-	<=0.01
3	chr11:15987995-17035990	*PLEKHA7*	-	<0.05
2	chr11:27068733-27385415	*CCDC34*	rs3781677	<0.05
3	chr11:27387508-27719721	*LGR4*	rs725015	<0.05
3	chr11:27387508-27719721	*BDNF-AS1*	-	<0.05
31	chr11:46299212-47736941	*CREB3L1*	-	<=0.01
31	chr11:46299212-47736941	*AMBRA1*	-	<0.05
31	chr11:46299212-47736941	*LRP4*	-	<0.05
31	chr11:46299212-47736941	*SLC39A13*	-	<0.05
31	chr11:46299212-47736941	*CELF1*	rs12419692;rs10838708	<0.05
31	chr11:46299212-47736941	*CKAP5*	-	<0.05
2	chr11:68080077-68382802	*LRP5*	rs10750821	<0.05
1	chr12:49372398-49375459	*WNT1*	-	<0.05
4	chr12:49388932-49463808	*MLL2*	-	<=0.01
8	chr12:53662083-53874945	*SP7*	-	<=0.01
8	chr12:53662083-53874945	*SP1*	-	<=0.01
8	chr12:53662083-53874945	*PCBP2*	-	<0.05
8	chr12:53662083-53874945	*ESPL1*	-	<0.05
1	chr12:90501516-90506066	-	-	<=0.01
4	chr15:67356101-67819628	*SMAD3*	rs12443279	<=0.01
1	chr16:238968-279462	*LUC7L*	rs1203974;rs8466	<0.05
1	chr16:72816784-73093597	*ZFHX3*	-	<=0.01
4	chr17:1963133-2240678	*SMG6*	-	<0.05
5	chr17:41831103-41940994	*SOST*	rs1731902	<=0.01
7	chr17:42154121-42298994	*UBTF*	rs11079983;rs7222501	<=0.01
1	chr2:119599766-119605254	*EN1*	-	<=0.01
1	chr2:202899310-202903160	*FZD7*	-	<0.05
1	chr2:85360533-85537511	*TCF7L1*	-	<=0.01
1	chr20:10618332-10654608	*JAG1*	rs8708	<=0.01
1	chr20:39314488-39317880	*MAFB*	rs2902941	<0.05
1	chr21:36160098-37357047	*RUNX1*	-	<0.05
3	chr3:156391024-156697347	*TIPARP*	rs13070271;rs7644005	<=0.01
1	chr3:41236328-41301587	*CTNNB1*	rs11711946	<0.05

Supplementary Table 4 (continued)

No. Genes	Locus Location	Gene Prioritized	Top cis eQTL SNP	False discovery rate
1	chr4:88742550-88767969	*MEPE*	-	<=0.01
2	chr4:980785-1020685	*FGFRL1*	-	<0.05
1	chr6:127439749-127518910	*RSPO3*	-	<=0.01
2	chr6:44777054-45518818	*RUNX2*	rs2772395	<=0.01
4	chr7:120628731-121038821	*CPED1*	rs1949803	<=0.01
1	chr7:37945543-38065297	*SFRP4*	rs6953908	<=0.01
3	chr8:119935796-120257913	*TNFRSF11B*	rs1389541	<0.05

Supplementary Table 5. Gene Ontology (GO) enrichment analysis. Genes in the 18 novel loci as defined by DEPICT where used to perform a GO enrichment (p-adjust, FDR<0.05).

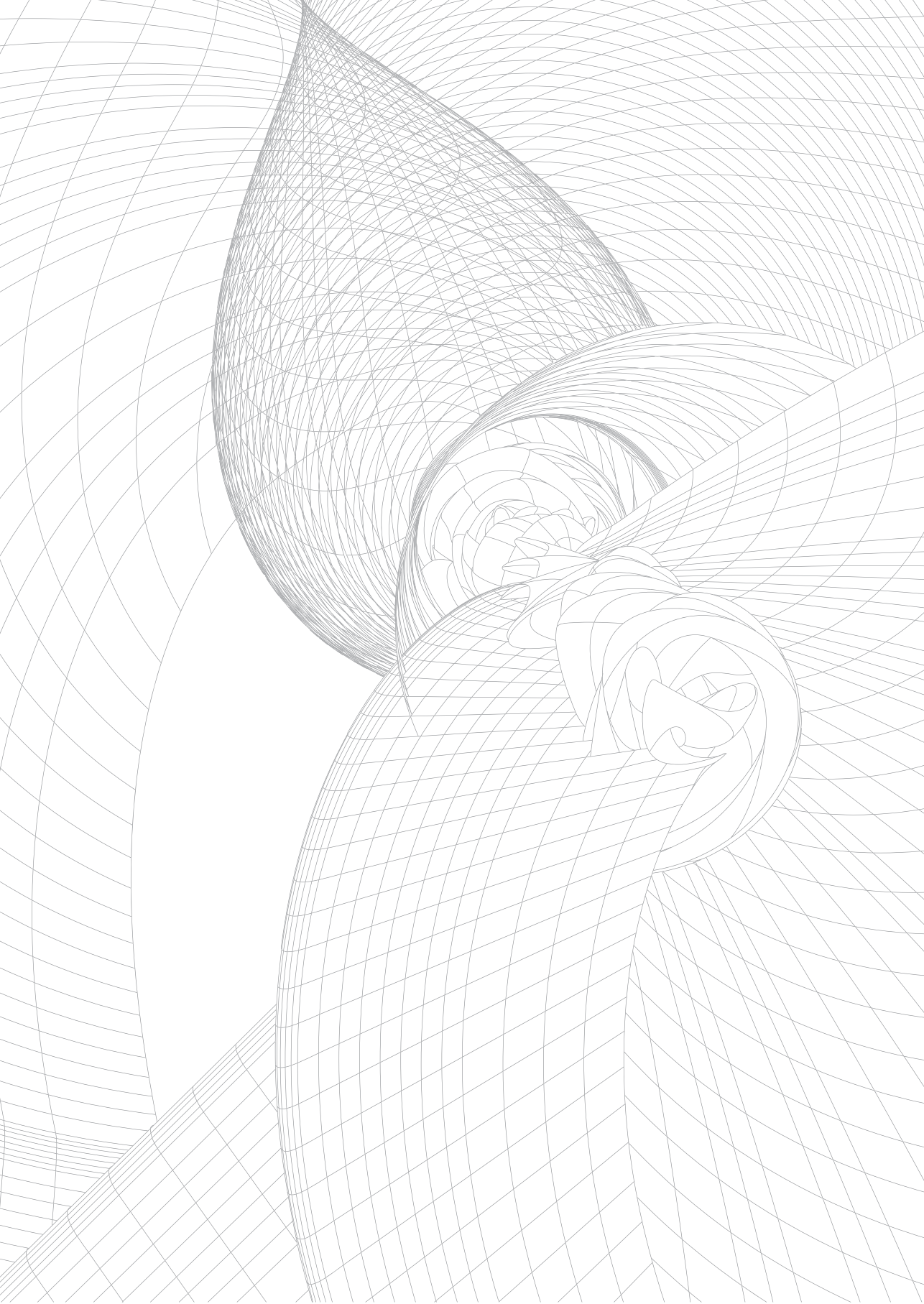
ID	Description	p.adjust	geneID
GO:0045596	negative regulation of cell differentiation	5.78E-04	APC/SMAD3/ZFHX3/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0009887	organ morphogenesis	7.86E-04	CSF1/APC/AQP1/SMAD3/PLEKHA1/SLC8A1/AXIN2/RUNX1/MAFB
GO:0035412	regulation of catenin import into nucleus	0.001	SMAD3/AXIN2/FZD7
GO:0016055	Wnt signaling pathway	0.001	CSNK1G3/APC/SMAD3/AXIN2/FZD7/TCF7L1
GO:0035282	segmentation	0.001	SMAD3/AXIN2/TCF7L1/MAFB
GO:0051239	regulation of multicellular organismal process	0.001	CSF1/GHRHR/APC/AQP1/SMAD3/ZFHX3/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0060070	canonical Wnt signaling pathway	0.001	APC/SMAD3/AXIN2/FZD7/TCF7L1
GO:0030278	regulation of ossification	0.004	CSF1/APC/SMAD3/AXIN2
GO:0032879	regulation of localization	0.005	CSF1/CEP120/GHRHR/APC/AQP1/SMAD3/SLC8A1/AXIN2/FZD7/RUNX1
GO:0022414	reproductive process	0.005	ODF3/CSF1/GHRHR/APC/PLEKHA1/SLC8A1/FZD7/RUNX1
GO:0003006	developmental process involved in reproduction	0.005	CSF1/APC/PLEKHA1/SLC8A1/FZD7/RUNX1
GO:0051179	localization	0.006	LDLRAD3/CSF1/CEP120/CEP112/GHRHR/APC/AQP1/SMAD3/BET1L/PLEKHA1/RIC8A/SLC8A1/AAGAB/AXIN2/FZD7/RUNX1
GO:0032502	developmental process	0.007	ODF3/CSF1/CEP120/GHRHR/APC/AQP1/SMAD3/ZFHX3/PLEKHA1/RIC8A/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0007050	cell cycle arrest	0.008	APC/SMAD3/ZFHX3/PSMD13
GO:0007389	pattern specification process	0.008	APC/SMAD3/AXIN2/TCF7L1/MAFB

Supplementary Table 5 (continued)

ID	Description	p.adjust	geneID
GO:0042127	regulation of cell proliferation	0.008	CSF1/GHRHR/APC/AQP1/SMAD3/AXIN2/FZD7/RUNX1
GO:0008283	cell proliferation	0.009	CSF1/CEP120/GHRHR/APC/AQP1/SMAD3/AXIN2/FZD7/RUNX1
GO:0033036	macromolecule localization	0.009	CEP120/CEP112/APC/SMAD3/BET1L/PLEKHA1/AAGAB/AXIN2/FZD7/RUNX1
GO:0034614	cellular response to reactive oxygen species	0.009	AQP1/PLEKHA1/SLC8A1
GO:0048518	positive regulation of biological process	0.010	CSF1/NLRP6/GHRHR/APC/AQP1/SMAD3/ZFHX3/PSMD13/PLEKHA1/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0048519	negative regulation of biological process	0.010	NLRP6/APC/AQP1/SMAD3/ZFHX3/PSMD13/PLEKHA1/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0008104	protein localization	0.012	CEP120/CEP112/APC/SMAD3/BET1L/PLEKHA1/AAGAB/AXIN2/FZD7
GO:0031122	cytoplasmic microtubule organization	0.012	CEP120/APC
GO:0002682	regulation of immune system process	0.014	CSF1/NLRP6/APC/SMAD3/PLEKHA1/RUNX1/MAFB
GO:0051983	regulation of chromosome segregation	0.015	APC/AXIN2
GO:0009893	positive regulation of metabolic process	0.015	CSF1/GHRHR/APC/SMAD3/PSMD13/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0035264	multicellular organism growth	0.018	CSF1/GHRHR/PLEKHA1
GO:0051641	cellular localization	0.018	CEP120/GHRHR/APC/AQP1/SMAD3/BET1L/SLC8A1/AXIN2/FZD7/RUNX1
GO:0030155	regulation of cell adhesion	0.018	CSF1/APC/SMAD3/FZD7
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	0.020	GHRHR/RIC8A/FZD7
GO:0060986	endocrine hormone secretion	0.020	AQP1/RUNX1
GO:0043009	chordate embryonic development	0.020	SMAD3/RIC8A/SLC8A1/AXIN2/RUNX1
GO:0001503	ossification	0.020	CSF1/APC/SMAD3/AXIN2
GO:0048538	thymus development	0.021	APC/MAFB
GO:0035556	intracellular signal transduction	0.023	CSF1/NLRP6/GHRHR/APC/SMAD3/PSMD13/PLEKHA1/SLC8A1/FZD7
GO:0030104	water homeostasis	0.023	GHRHR/AQP1
GO:0010604	positive regulation of macromolecule metabolic process	0.023	CSF1/APC/SMAD3/PSMD13/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0042060	wound healing	0.025	NLRP6/AQP1/SMAD3/SLC8A1/FZD7
GO:0019220	regulation of phosphate metabolic process	0.026	CSF1/NLRP6/GHRHR/APC/SMAD3/RIC8A/AXIN2/FZD7

Supplementary Table 5 (continued)

ID	Description	p.adjust	geneID
GO:0002376	immune system process	0.026	CSF1/NLRP6/APC/SMAD3/PSMD13/PLEKHA1/FZD7/RUNX1/MAFB
GO:0048522	positive regulation of cellular process	0.026	CSF1/GHRHR/APC/AQP1/SMAD3/ZFHX3/PSMD13/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0006950	response to stress	0.031	CSF1/NLRP6/APC/AQP1/SMAD3/PSMD13/PLEKHA1/SLC8A1/AXIN2/FZD7/RUNX1
GO:0023052	signaling	0.031	CSF1/CSNK1G3/NLRP6/GHRHR/APC/AQP1/SMAD3/PSMD13/PLEKHA1/RIC8A/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1
GO:0010817	regulation of hormone levels	0.031	GHRHR/AQP1/PLEKHA1/RUNX1
GO:0051817	modification of morphology or physiology of other organism involved in symbiotic interaction	0.033	AQP1/SMAD3
GO:0007154	cell communication	0.034	CSF1/CSNK1G3/NLRP6/GHRHR/APC/AQP1/SMAD3/PSMD13/PLEKHA1/RIC8A/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1
GO:0032501	multicellular organismal process	0.036	ODF3/CSF1/CEP120/GHRHR/APC/AQP1/SMAD3/ZFHX3/PLEKHA1/RIC8A/SLC8A1/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0050886	endocrine process	0.041	AQP1/RUNX1
GO:0065008	regulation of biological quality	0.041	CSF1/GHRHR/APC/AQP1/SMAD3/PLEKHA1/SLC8A1/AXIN2/RUNX1/MAFB
GO:0050790	regulation of catalytic activity	0.041	CSF1/GHRHR/APC/AQP1/SMAD3/PSMD13/RIC8A/AXIN2
GO:0035821	modification of morphology or physiology of other organism	0.044	AQP1/SMAD3
GO:0019222	regulation of metabolic process	0.044	LDLRAD3/CSF1/NLRP6/GHRHR/APC/AQP1/SMAD3/ZFHX3/PSMD13/RIC8A/AXIN2/FZD7/TCF7L1/RUNX1/MAFB
GO:0048583	regulation of response to stimulus	0.044	CSF1/NLRP6/GHRHR/APC/SMAD3/PLEKHA1/AXIN2/FZD7/TCF7L1/RUNX1
GO:0040007	Growth	0.044	CSF1/GHRHR/SMAD3/PLEKHA1/FZD7
GO:0001775	cell activation	0.046	CSF1/APC/SMAD3/FZD7/MAFB



Chapter 4.3

A Large-Scale Genome-Wide Investigation of Age-dependent Effects in Genetic Associations with TB-BMD

Carolina Medina-Gomez*, John Kemp*, Alessandra Chesi, Eskil Kreiner-Møller, Tarun Ahluwalia, Dennis Mook, Youfang Liu, Fernando P. Hartwig, Dan Evans, Raimo Joro, Cornelia van Duijn, Ivana Nedeljkovic, Benjamin Mullin, Joel Eriksson, Brent Richards, Rebecca Jackson, David Karasik, Nathalie Van der Velde, Arfan Ikram, Babette Zemel, Tamara Harris, Yanhua Zhou, John Robins, Ruifang Li, Bruce Psaty, Carrie Nielson, Bram van der Eerden, Jeroen van de Peppel, Wilson Scott, Bernardo L Horta, Timo Lakka, Struan Grant, Fiona McGuigan, Jim Wilson, Unnur Styrkársdóttir, Dan Koller, Kun Zhu, Doug Kiel, Claes Ohlsson, Cheryl L Ackert-Bicknell, Evangelia Ntzani, Evangelis Evangelou, Andre G. Uitterlinden, Vincent Jaddoe, Jon H. Tobias, Vangelis Evangeli Dave M. Evans, Fernando Rivadeneira

*These authors equally contributed to this manuscript

In preparation

ABSTRACT

The skeleton is a dynamic organ increasing in mass and density until approximately the third decade of life, and subsequently after the age of fifty-yearsbeing resorpt, particularly in women after menopause. Bone Mineral Density (BMD) a highly heritable trait is used to assess skeletal health in children and risk of osteoporosis later in life. To date >60 loci associated with BMD have been described arising from studies focusing primary in old adults. Nevertheless, age-specific regulation of BMD might drive the dynamic nature of this trait. To evaluate possible age-dependent effects of genetic variants on BMD, we conducted a genome-wide association study (GWAS) meta-analysis of total body (TB-)BMD by age-strata of 15 years, in more than 50,000 individuals across the life span. We identified variants in 24 loci, associated with TB-BMD at different age-strata ($P<5x10^{-8}$), including variants in three loci not identified in an overall meta-analysis, including *RIN3, ZNF789/ZSCAN5*, and *PTPN5*. In addition, we identified variants in 18 loci exerting age dependent effects on BMD. The strongest age specific effects were found for variants in the ESR1 locus only exerting a GWS effect ($\beta=0.07$ SD, $P=2.6x10^{-12}$) in adults ($beta_{age}=|1.5x10^{-3}|$, $P=2.42x10^{-10}$); and variants in *RIN3* with GWS effect ($\beta=0.1$ SD, $P=1x10^{-8}$) only in children ($beta_{age}=|1.8x10^{-3}|$, $P=1x10^{-8}$). In conclusion, applying an age-stratified GWAS approach allowed us to identify novel loci and further characterize variants exerting effects that change with aging, helping to unveil further the complex genetic mechanisms underlying bone health and partly the risk of osteoporosis.

INTRODUCTION

The skeleton is a rigid yet dynamic organ that provides support to the body, protects vital organs, serves as a reservoir for calcium, growth factors and cytokines. One of the most important properties of bone is its ability to regenerate and repair; bones continuously adjusts skeletal mass and architecture to changing mechanical environments [1]. Generation and maintenance of the shape of bone during skeletal growth depends on bone modeling; although throughout life, bones also undergo remodeling to remove old, microdamaged bone and replace it with new, mechanically stronger tissue in order to help preserve bone strength [2]. In fact, this process is so dynamic that it is estimated that the entire adult human skeleton is replaced every ten years [3].

Bone mass increases steadily during childhood and markedly during adolescence growth [4]. Peak bone mass is attained at approximately the third decade of life. Thereafter, until an age of about fifty years, bone mineral density (BMD) levels remain more or less stable, preserving the coupling between bone formation and resorption (e.g., bone remodeling). From there, bone loss occurs at a higher rate after fifty-years of age[5], particularly in women after menopause.

Twin and family studies indicate that BMD is a highly heritable trait, with heritability estimates ranging between 50 - 85%[6-8]. These estimates can vary due to the analytical model used in their calculation, the skeletal site measured and the population under study. Importantly, the porportion of variance explained by genetic factors, also hanges with age. Gueguen and colleagues assessed the heritability of BMD in 129 families and found that heritability estimates increase during childhood and reach a maximum of 85% in early adulthood, being of 63% and 34% at 14 and 53 years, respectively[9]. In addition, Moayyeri and colleagues showed that heritability estimates decreased with aging based on a large twin-based longitudinal study with participants 35 years or older and average follow up of ten years[10]. To date more than 60 loci have been described as associated with this trait [11]. The vast majority of these discoveries come from genome-wide association studies (GWAS) in elderly individuals [12,13], while in children 14 loci have been identified [14,15]. Previous studies have shown for specific variants in *CPED1* [14], *ESR1* and *ESR2* [16] the magnitude of the genetic effect varies with age. Nevertheless, it is plausible that several genetic variants would exert differential age effects considering that BMD shows an age-dependent pattern of an initial increase in bone accrual until achieving peak bone mass followed by a subsequent decline in later life, driven by different metabolic processes (e.g. menopause in women). In this study, we applied a comprehensive approach to investigate the possible modulation of genetic effects by age, through the use of aggregated GWAS summary results from cross-sectional studies analyzed across age-bins. Besides the specific meta-analyses, we assesed the changes of magitud of genetic effects with respect to age, by coparing the extreme age groups and by employing a meta-regression methodology.

METHODS

Study Populations

Subjects

This study included 26 GWAS comprising ~53,427 individuals from populations across America, Europe, and Australia, with a variety of epidemiological designs (**Chapter 4.2, Supplementary Table 1**) and participant characteristics (**Chapter 4.2, Supplementary Table 2**). The correspondent Medical Ethical Committee of each participating study approved all research aims and the specific measurements performed on human subjects. Written informed consent was provided by all subjects or their parents (or legal guardian) in the case of children.

BMD measurement

Total body BMD (g/cm^2) was measured in all participants using dual-energy X-ray absorptiometry (DXA) following standard manufacturer protocols. As recommended by the International Society for Clinical Densitometry [17] total body less head (TBLH) was the measurement used in pediatric cohorts. Detailed information on this assessment per study can be found in **Chapter 4.2, Supplementary Table 1**.

GWAS data and imputation

All individuals included in this study had genome-wide array data, except for the PANIC study genotyped with the Metabochip. Quality control of the obtained genotypes is summarized in **Chapter 4.2, Supplementary Table 1**. To enable meta-analysis, each study performed genotype imputation using the cosmopolitan (all ethnic panels combined) 1000 genomes phase 1 version 3 (March 2012) reference panel [18], via MACH/Minimac[19,20] or SHAPEIT/Impute2[21] suites yielding ~30,000,000 SNPs.

Association Analyses

TB(LH)-BMD was corrected for age, weight, height and the genomic principal components (derived from GWAS data), as well as any additional study-specific covariates (e.g. recruiting center) using linear regression models. For studies with non-related individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included as a covariate in the model. Finally, residuals were subject to inverse normal transformation to diminish the impact of BMD extreme values in the analysis. The analyses were performed in each study for 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-residual was estimated via linear regression using MACH2QTL[19], SNPtest[22], PLINK[23] or ProbABEL[24].

Quality control of TB-BMD association summary statistics

A centralized quality-control procedure implemented in EasyQC [25] was applied to all study-specific association files to identify cohort-specific problems: (1) assessment of possible problems in BMD-residuals normalization by inspection of standard error vs. sample size plots; (2) comparison of allele frequency alignment against the 1000 Genomes Project phase1 reference data to pinpoint any potential strand issues; (3) examination of quantile-quantile (QQ) plots and genomic inflation factors per study to identify problems arising from population stratification, cryptic relatedness and genotype biases; (4) Analytical issues in the computation of beta estimates, standard errors and P-values by evaluation of (P-value – Test statistic) P-Z plots.

In addition, we excluded variants if they had missing information (e.g., missing association P-value, effect size (beta) estimate, alleles, allele frequency), or nonsensical values (e.g., absolute beta estimates or standard errors >10, association P-values (P) >1 or <0; or imputation quality < 0; infinite beta estimates or standard errors); minor allele frequency (MAF) less than 0.5%; Imputation quality <0.4 (when Impute2 was used for imputation) and <0.3 (when Minimac was used for imputation). Moreover, markers were flagged if they had large allele frequency deviations from reference populations (>0.6 for admixed studies and >0.3 for ancestry-specific studies).

GWAS meta-analyses

Analyses were performed in each study and meta-analyzed by age strata (minimum sample size per bin N=200) resulting in: 1) 0-15 years (N=11,807), 15-30 years (N=4,180), 30-45 years (N=6,604), 45-60 years (N=10,482), and 60 or more years (N=20,354). Although there was not exclusion based on the ethnicity of the individuals included in each study, more than 80% of them were reported as of European ancestry (**Chapter 4.2, Supplementary Table 1**). We discarded markers per age-bin analysis if they were present in less than three studies or the overall MAF was less than 1%; approximately 23,700,000 markers were assessed for association. We used the conventional genome-wide significance level (GWS, $P<5x10^{-8}$) for SNP discovery.

Age-dependent effects screening

We selected SNPs which were suggestively ($P<5x10^{-6}$) associated with BMD either in the children or the elderly population, whose MAF difference across these meta-analyses was lower than 0.5, assigning to each SNP the lowest P-value reached in any of these two association meta-analyses. Then, in PLINK[23], we clumped the dataset with an $r^2 \geq 0.8$, so that variants in the final set will not be correlated at this high level. Linkage disequilibrium (LD) calculations were based on the RS1 study best guess data (**Chapter 4.2, Supplementary Table 1**), using as reference the most BMD-associated SNPs and finally, pruning remaining

SNPs within 0.5 Mb of each other. The assessment of an age-dependent effect was performed by testing for these SNPs, the difference between beta-estimates in children vs. elderly meta-analyses (P_{diff}) using Easy-strata[26]. We used a Bonferroni-corrected P-value to determine the significance of the differences. Also, we evaluated the gene-age interaction of these SNPs with significant differences in effect in children and elderly, through 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). Standard errors of the effect estimates of each subgroup were multiplied by the square root of the genomic inflation factor, if this was higher than 1, as proposed before[27]. We performed this meta-regression in R[28] using the Metafor package[29], statistical evidence of linear association was adjusted for multiple testing. For those markers whose effect displayed evidence for age-dependency, we assessed the 2 degrees of freedom results, meaning the significance of intercept (basal effect of the SNP) and interaction effect, in order to test whether the inclusion of an interaction term would enhance the discovery of associated variants in an overall meta-analysis.

DEPICT Pathway analysis

We used the recently developed DEPICT[30] method to identify enriched genes-sets as well as tissues/cell types where genes from associated loci are highly expressed. DEPICT first selects all lead SNPs below a certain threshold with respect to a target P-value. We tested the set of GWS SNPs in the elderly, and those reaching the suggestive threshold ($P<5x10^{-6}$) on the other subgroups, where statistical power was more limited.

Search for biological and functional knowledge of the identified association regions

We examined whether GWS SNPs are in regions known to be associated with previous GWAS of bone traits. In addition, for all those SNPs not within 500Kb of the bone GWAS-SNPs we checked in PubMed and Web of Science if nearby genes (within 500Kb) were known to play a role in bone metabolism. Genomic annotation for all SNPs was made based on UCSC HG19 assessed through Bioconductor [31].

RESULTS

GWAS meta-analyses

Meta-analyses across age strata resulted in 24 different loci associated with TB-BMD as summarized in **Figure 1**. Most of the loci were GWS associated in the elderly population, where the statistical power for discovery was the highest (i.e., largest sample size). Genomic inflation factors in the five age bins were all close to unity, providing no substantial evidence for population stratification issues. Markers surpassing the GWS threshold were in the vast majority common non-coding SNPs (**Supplementary Table 1**). Missense variants in *RIN3* (rs117068593,

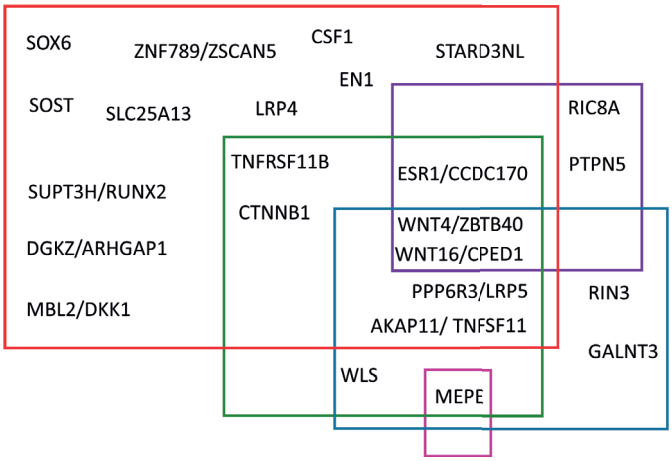


Figure 1. Loci GWS associated with TB-BMD at the primary meta-analysis per age-bin. Each locus is characterized by one or two representative genes chosen either because of their biological relevance or physically proximity to the most strongly associated SNP. The colored squares represent the different age bins enclosing the genes from loci harboring GWS variants: 60 yrs or more (N=20,354) in red; 45-60 yrs, (N=10,482) in green; 30-45 yrs (N=6,604) in purple; 15-30 yrs (N=4,180) in pink; and 0-15 yrs (N=11,807) in blue.

GWS in children), *TNFRSF11B* (rs2073618, GWS in elderly), *LRP4* (rs2306029 and rs6485702, GWS in elderly), *CCDC170* (rs12205837 and rs3734804, GWS in elderly, and rs6929137 GWS in the 45-60 years and elderly groups), and *WNT16* (rs2707466 and rs2908004 associated in all age-strata but 15-30 years); associated with TB-BMD are all classified as benign both in SIFT[32] and polyphen2[33] algorithms. Additionally, a low-frequency variant in *LRP5,* GWS in the elderly strata - rs4988321/A (11:68174189, MAF=0.04)- is annotated in ClinVar, as a patient with osteoporosis-pseudoglioma syndrome was a homozygous carrier of this mutation[34].

When comparing the results obtained in the age-specific meta-analyses with those reported in the overall meta-analysis, including all age category sub-populations (Chapter 4.2), three additional novel loci were identified (**Figure 1, Supplementary Table 1**). In children (i.e., 0-15 age bin), the 14q32.12 locus, harboring *RIN3*, had seven common (MAF=17%) variants, including one missense marker, which reached the GWS threshold. In the 30-45 years age-bin, one rare variant rs71486878 (MAF=1.5%) mapping intronic in the *PTPN5* gene was GWS associated. Moreover, in the elderly, three variants (MAF ~5%) mapping to 7q22.1, containing *ZNF789* and *ZSCAN5* among others were also associated with TB-BMD.

Age-dependent effects screening

Association signals, rising from the meta-analyses of the children and elderly individuals were different, with the elderly meta-analysis showing more association peaks even if signals in chromosome 2 and 14 were absent (Figure 2). We sought to investigate if these distinct signals were a consequence of variation in sample size across these age-strata

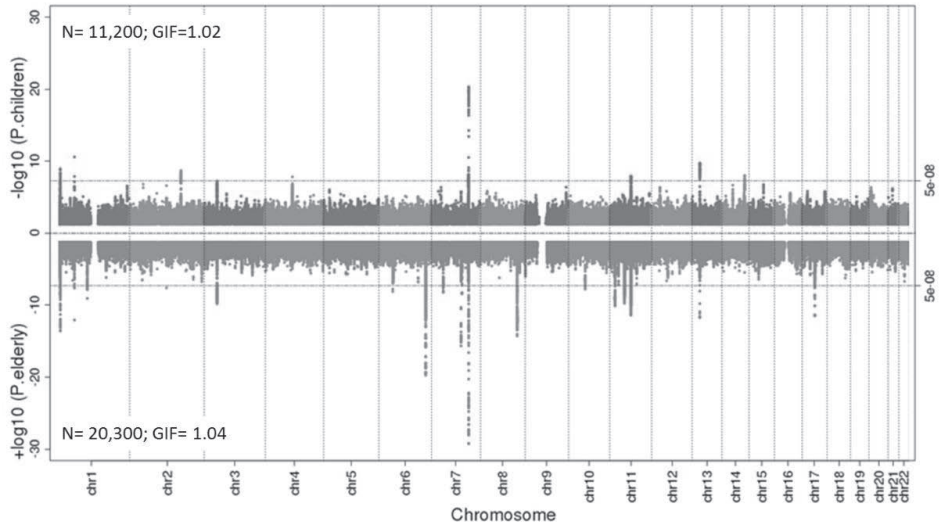


Figure 2. Miami plot of association statistics (-log10(P values)) for TB-BMD contrasting children (top) and elderly (bottom) populations. Each dot represents an SNP and the x-axis indicates its chromosomal position. Dashed horizontal red and yellow lines mark the GWS threshold ($P<5x10^{-8}$) and suggestive threshold ($P<1x10^{-6}$), respectively.

(i.e., children vs. elderly) or rather due to age-dependency of the effect of the associated variants on BMD. We selected 3,529 SNPs suggestively associated ($P<10^{-6}$) with BMD either in the children or in the elderly population meta-analyses, from which 377 were pruned by LD ($r^2<0.8$). These 377 variants were tested for differences in effect size between the children and elderly groups. We chose suggestive (rather than GWS) SNPs, to allow capturing potentially true genetic associations with BMD either in the children or the elderly population meta-analyses, without particularly favoring the markers identified in the elderly, where the statistical power was the highest. Fifty-two markers, mapping to 18 different loci (**Table 1**), displayed significant effect differences after multiple testing adjustment ($P=0.05/377=1.33x10^{-4}$). Markers within ten loci presented with larger effects in children (i.e., 2p21, 5p14.3, 5q11.2, 7p15.3, 7q31.31, 11p12, 12p11.22, 17q25.1, 20p12 and 21q21.3). Whereas, larger effects in the elderly were observed for markers in six loci (i.e., 2q37, 6q21.1, 6q25.1, 7q22.1, 11p15.3 and 17q12). The remaining two loci exhibited allelic heterogeneity, with markers exerting greater effects on the children while others exerted greater effects in the elderly individuals. In locus 14q32.12, two RIN3 intronic variants were tested. While a common variant (rs72699866 with MAF=17%) had greater effect on the children ($beta_{A:0\text{-}15yrs}=-0.09$ $P_{0\text{-}15yrs}=1x10^{-8}$) than in the elderly ($beta_{A:60>yrs}=0.03$ $P_{60>yrs}=0.08$), a low frequency variant (rs77002573 with MAF=3.6%) located 11Kb away, had greater effect on the elderly ($beta_{A:60>yrs}=-0.15$ $P_{60>yrs}=3.8x10^{-6}$) than on the children ($beta_{A:0\text{-}15yrs}=0.09$ $P_{0\text{-}15yrs}=0.02$). Likewise, in the 13q14 locus rs7318320, rs9525638 and rs875625 in moderate LD with each

Table 1. Genetic Markers showing significant (P<9x10-4) difference in effect magnitude on TBBMD between children and elderly populations A1, represents the coded allele the effect (β) is based on and f1 its frequency in elderly (f1.eld) and in children (f1.child). Pdiff correspond to the different in effect size between the children and elderly meta-analyses. Poverall corresponds to the P value in the overall meta-analysis shown in Chapter 4.2. In bold significant associations for the effect (βage) of age (significance $P<9x10^{-4}$).

MarkerName	band	A1	f1.eld.	β.eld.	P.eld.	f1.child.	β.child	P.child	Poverall	P.diff	βage	P.age
rs7589428	2p21	a	0.53	0.004	0.739	0.52	0.070	1.54E-07	2.74E-04	9.47E-05	-0.0008	0.001
rs2741023	2q37	a	0.34	0.057	1.91E-07	0.35	-0.010	0.451	2.44E-04	1.15E-04	-0.0009	**3.78E-04**
rs111652685	5p14.3	t	0.95	0.013	0.676	0.95	-0.177	1.01E-06	1.55E-01	5.96E-05	0.0022	0.001
rs72752228	5q11.2	a	0.04	-0.029	0.310	0.04	0.151	3.95E-06	3.28E-01	3.11E-05	0.0026	**4.44E-05**
rs634789	6q21.1	a	0.65	0.056	4.84E-07	0.63	-0.012	0.392	1.29E-04	1.23E-04	-0.0010	**2.91E-05**
rs4711801	6q21.1	c	0.54	0.052	1.44E-06	0.53	-0.013	0.322	6.64E-04	1.17E-04	0.0011	**1.23E-05**
rs62444275	6q25.1	a	0.54	0.081	1.08E-13	0.53	0.004	0.773	4.85E-17	7.96E-06	-0.0010	**7.33E-05**
rs4455680	6q25.1	a	0.21	0.106	5.27E-08	0.21	-0.023	0.247	5.25E-07	3.80E-06	-0.0021	**5.36E-07**
rs6910761	6q25.1	a	0.44	-0.096	4.42E-20	0.45	-0.012	0.362	3.98E-24	4.69E-07	-0.0011	**4.70E-06**
rs4870045	6q25.1	a	0.34	-0.089	2.46E-15	0.34	-0.010	0.487	4.37E-19	7.82E-06	0.0011	**2.88E-05**
rs7768165	6q25.1	a	0.32	-0.077	4.49E-10	0.33	0.010	0.507	6.04E-13	6.02E-06	0.0011	**1.53E-04**
rs9383930	6q25.1	a	0.67	0.072	1.67E-10	0.67	0.004	0.766	1.67E-11	1.29E-04	0.0010	**9.98E-05**
rs9371543	6q25.1	a	0.27	-0.074	1.91E-09	0.26	0.002	0.920	9.55E-09	8.83E-05	0.0010	**2.76E-04**
rs9479083	6q25.1	a	0.30	-0.059	1.01E-07	0.30	0.020	0.156	4.47E-07	1.02E-05	0.0011	**4.10E-05**
rs6929137	6q25.1	a	0.34	-0.071	3.12E-11	0.34	0.013	0.352	3.02E-10	1.47E-06	0.0013	**2.51E-07**
rs56681564	6q25.1	a	0.23	-0.058	8.95E-07	0.24	0.020	0.174	1.99E-06	3.99E-05	0.0012	**2.70E-05**
rs3734804	6q25.1	a	0.52	-0.070	5.35E-12	0.53	0.020	0.125	5.65E-10	5.25E-08	0.0013	**8.42E-08**
rs6917575	6q25.1	t	0.41	0.063	1.13E-09	0.39	-0.005	0.691	4.68E-12	4.82E-05	-0.0010	**9.04E-05**
rs58164038	6q25.1	a	0.75	0.067	8.02E-07	0.76	-0.030	0.080	3.41E-05	7.93E-06	0.0014	**6.35E-06**
rs6930633	6q25.1	a	0.65	0.064	2.94E-09	0.64	-0.017	0.217	3.51E-08	3.00E-06	0.0012	**1.05E-06**
rs7740042	6q25.1	a	0.20	-0.069	5.76E-08	0.21	0.025	0.114	1.30E-06	3.91E-06	0.0014	**5.10E-06**

Table 1 (continued)

MarkerName	band	A1	f1.eld.	β.eld.	P.eld.	f1.child.	β.child	P.child	Poverall	P.diff	βage	P.age
rs11755786	6q25.1	t	0.14	-0.080	6.93E-08	0.15	0.017	0.362	8.40E-07	4.93E-05	0.0015	**2.65E-05**
rs2941741	6q25.1	a	0.41	0.077	9.51E-14	0.40	-0.010	0.450	1.40E-12	2.11E-07	-0.0011	**4.02E-06**
rs851982	6q25.1	t	0.62	-0.071	8.13E-12	0.62	0.011	0.401	1.56E-10	1.06E-06	-0.0011	**5.10E-06**
rs851981	6q25.1	a	0.74	-0.074	2.01E-10	0.75	0.019	0.194	1.86E-07	7.37E-07	-0.0014	**8.42E-07**
rs1101080	6q25.1	t	0.29	0.065	1.24E-08	0.28	-0.018	0.217	1.17E-06	7.06E-06	-0.0011	**2.69E-05**
rs2347638	6q25.1	t	0.24	-0.068	7.71E-08	0.24	0.023	0.153	1.13E-04	7.70E-06	-0.0014	**2.32E-06**
rs2982559	6q25.1	a	0.54	0.073	2.57E-12	0.54	-0.039	0.003	1.37E-07	2.02E-11	-0.0015	**2.42E-10**
rs3020307	6q25.1	a	0.48	-0.064	4.79E-10	0.49	0.035	0.007	1.43E-06	1.68E-09	-0.0013	**2.72E-08**
rs2982553	6q25.1	c	0.65	0.052	1.74E-06	0.64	-0.030	0.029	7.93E-04	2.74E-06	-0.0011	**1.49E-05**
rs1124674	6q25.1	t	0.38	-0.061	9.47E-09	0.38	0.025	0.061	2.84E-07	5.29E-07	-0.0011	**4.77E-06**
rs2504063	6q25.1	a	0.43	-0.061	5.54E-09	0.44	0.024	0.067	3.49E-08	4.65E-07	-0.0011	**6.24E-06**
rs1415193	6q25.1	a	0.48	0.064	8.12E-10	0.46	-0.024	0.062	8.61E-08	1.02E-07	-0.0012	**1.05E-06**
rs2504066	6q25.1	c	0.46	0.057	4.07E-08	0.46	-0.021	0.115	3.96E-05	3.25E-06	-0.0011	**1.19E-05**
rs4722100	7p15.3	t	0.36	-0.005	0.659	0.36	0.070	1.51E-06	1.44E-01	4.05E-05	-0.0010	**2.87E-04**
rs148982377	7q22.1	t	0.96	0.161	3.86E-09	0.96	-0.012	0.725	1.71E-07	9.35E-05	0.0024	**2.09E-04**
rs117118687	7q31.31	a	0.98	0.007	0.843	0.98	-0.237	4.72E-06	3.07E-01	1.25E-04	0.0026	0.005
rs118048672	11p15.3	a	0.96	-0.179	2.33E-06	0.96	0.047	0.257	3.79E-04	5.76E-05	-0.0030	**2.31E-04**
rs7101502	11p15.3	a	0.80	-0.081	8.18E-11	0.80	-0.001	0.954	7.77E-09	7.03E-05	-0.0012	**3.13E-05**
rs10837576	11p12	a	0.30	-0.007	0.537	0.30	0.066	4.74E-06	1.83E-01	6.52E-05	-0.0010	**1.73E-04**
rs4420311	12p11.22	a	0.55	0.012	0.248	0.55	-0.066	5.27E-07	9.28E-02	2.95E-06	-0.0009	**1.22E-04**
rs258413	12p11.22	t	0.58	0.009	0.398	0.58	-0.066	4.45E-07	8.42E-02	7.47E-06	0.0009	**3.49E-04**
rs9533094	13q14	a	0.54	0.072	1.84E-12	0.55	0.008	0.529	7.25E-19	1.07E-04	0.0011	**7.74E-06**
rs7318320	13q14	a	0.45	0.009	0.398	0.45	-0.063	1.87E-06	8.82E-02	1.72E-05	0.0010	**6.62E-05**
rs9525638	13q14	t	0.58	-0.005	0.602	0.58	-0.084	1.90E-10	1.56E-04	2.38E-06	0.0009	**1.10E-04**

Table 1 (continued)

MarkerName	band	A1	f1.eld.	β.eld.	P.eld.	f1.child.	β.child	P.child	Poverall	P.diff	βage	P.age
rs875625	13q14	a	0.48	0.005	0.604	0.49	-0.060	4.05E-06	1.03E-01	7.76E-05	0.0008	0.001
rs72699866	14q32.12	a	0.17	0.036	0.009	0.18	-0.099	1.01E-08	1.39E-01	8.77E-10	-0.0018	**1.00E-08**
rs77002573	14q32.12	a	0.96	-0.156	3.77E-06	0.96	0.093	0.020	4.22E-01	2.10E-06	-0.0039	**3.06E-07**
rs11078892	17q12	t	0.46	-0.057	2.74E-06	0.46	0.021	0.173	2.25E-03	6.32E-05	-0.0010	**3.35E-04**
rs6501911	17q25.1	a	0.46	-0.006	0.586	0.47	0.060	3.93E-06	9.70E-03	8.12E-05	0.0010	**6.48E-05**
rs6107933	20p12	a	0.28	0.005	0.643	0.27	-0.073	4.52E-07	8.06E-03	2.11E-05	-0.0011	**3.58E-05**
rs117911512	21q21.3	t	0.05	-0.010	0.731	0.05	0.182	6.41E-07	6.38E-03	3.70E-05	0.0027	**7.48E-05**

other (r^2~0.6) and mapping in close vicinity or in *TNFSF11*, presented significant effect in the children (e.g., rs9525638, $beta_{A:0\text{-}15yrs}$=0.08 $P_{0\text{-}15yrs}$=1.9x10^{-10}) but not in the elderly (e.g., rs9525638, $beta_{A:>60yrs}$=0.005, $P_{>60yrs}$=0.60). On the contrary, the rs9533094 located 14-20 Kb upstream and closer to *AKAP11* showed significant effect in the elderly ($beta_{A:60>yrs}$=0.07 $P_{60>yrs}$=1.8x10^{-12}) but not in the children ($beta_{A:0\text{-}15yrs}$=-0.008 $P_{0\text{-}15yrs}$=0.52).

In the meta-regression, we evidenced a linear association of age with effect estimates for 48 out of these 52 markers, significant after correction (P=0.05/52=9.61x10^{-4}). By taking into consideration the interaction term in the joint modeling of the SNP effect (e.g., 2df test), we identified association signals, at GWS level, in 32 of these 52 SNPs. Of these, 16 were not identified in the meta-analysis of the same populations without age-stratified subgroup analysis (i.e., overall meta-analysis, Chapter 4.2). Although most of these 16 SNPs map to loci already identified in the overall meta-analysis, rs148982377 and rs72699866, (located in 7q22.1 and 14q32.12, respectively), reach GWS only in our meta-regression approach (**Supplementary Table 2**).

The strongest statistical evidence for age interaction was provided by rs2982559, an intronic variant mapping to *ESR1*, which was not GWS in the overall meta-analysis (P=1.3x10^{-7}) and presented high heterogeneity (P_{het} =4x10^{-5}) (Chapter 4.2). The effect of age accounted for 75% of the rs2982559-effect heterogeneity across cohorts ($beta_{age}$=|1.5x10^{-3}|, P=2.42x10^{-10}) (Figure 3). Nevertheless, other SNPs in the same locus reached GWS already in the overall meta-analysis. In contrast, the intronic variant rs72699866 in *RIN3*, provided a similar level of statistical evidence of effect dependency on age ($beta_{age}$=|1.8x10^{-3}|, P=1x10^{-8}), reaching a 2df P-value of 3.2x10^{-8} even when no evidence of association in the overall meta-analysis was found for this region (Figure 3). Variants with lower frequency (MAF<0.05) attained the largest modulation of TB-BMD by age, as expected due to power limitations. In most of the cases in which age modulated the effect of the genetic variant, the amount of residual heterogeneity was negligible after its inclusion as moderator. Nevertheless, this was not the case for specific variants in 13q14, 20p12, and 6q25.1, for which residual effect heterogeneity was still nominally significant after inclusion of age in the model, probably indicating that other moderators besides age are also influencing their effect.

DEPICT Pathway analysis

DEPICT was used to test for enrichment and define pathways and networks between the genes harbored by the GWS associated loci in elderly, and suggestively associated in each of the other age bins. Genes were prioritized only in the 45-60 years (*SP7*, *BICC1*, *WNT4*, *CD300LG*, *TNFRF11B*, *LGR4*, *PDZRN4*, *SOST*, *PKDCC*, *LRP5*, *TNFSF11*, *FRMD6*, *IDUA*, and *FGFRL1*) and in the elderly (*EN1*, *ESR1*, *CPED1*, *LRP5*, *SOST*, *SFRP4*, *ZNF498*) groups. DEPICT based on the suggestive associated markers with BMD identified enrichment for the GO

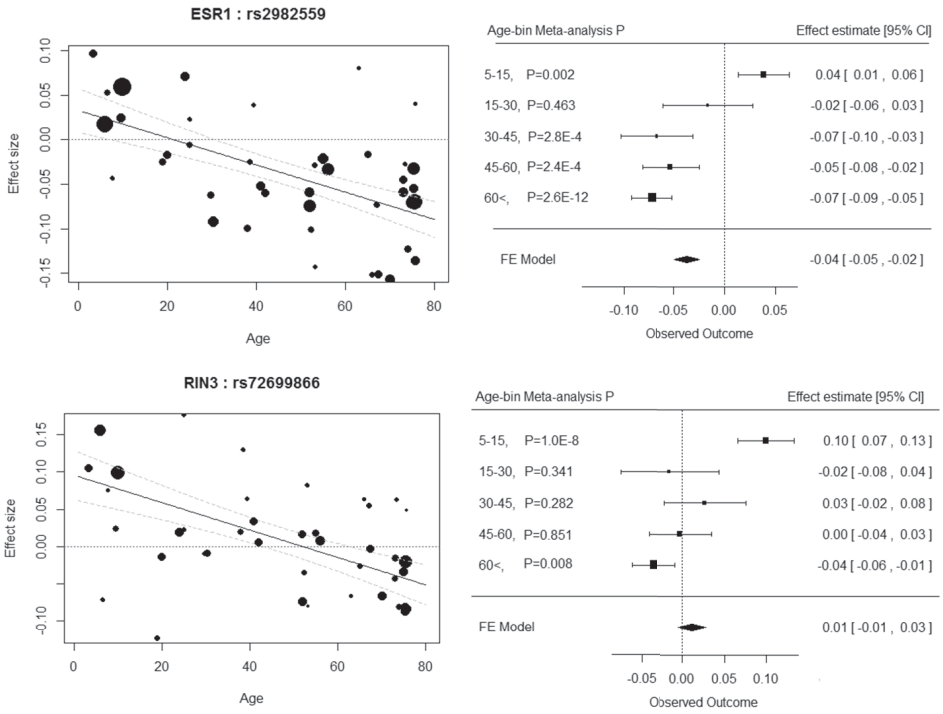


Figure 3. The two loci exhibiting the largest effect of age in the modulation of genetic variant effect in the meta-regression. The panels display the SNP effect as a function of age. At the right, each circle represents an study subgroup (i.e., study divided in age strata), with the circle size proportional to the inverse variance of the SNP main effect. At the left, estimates from each age-bin meta-analysis, with the symbol size proportional to the inverse variance of the SNP main effect.

term "cell-matrix adhesion" in the age group from 15-30 years and the genesets "abnormal long bone epiphyseal plate proliferative zone", "abnormal chondrocyte morphology", "regulation of growth", "ossification, cell growth", "abnormal long bone epiphyseal plate morphology" in the 45-60 years group. The limited power of our analysis may be due to the relatively small number of loci identified in the meta-analyses across age-strata or to limited existing knowledge surrounding bone metabolism.

DISCUSSION

Our meta-analysis of data from more than 50,000 individuals distributed across five different age-strata comprising the whole life-course identified 24 loci significantly associated with TB-BMD. While an overall GWAS including the same populations without age stratification reported 52 different loci associated with TB-BMD, three (7q22.1 [*ZNF789/ZSCAN5*], 14q32.12 [*RIN3*], 11p15.1 [*PTPN5*]) of the 25 associated loci in the different age-strata did not

reach GWS for the TB-BMD overall meta-analysis, despite its larger sample size. This observation supports the contention that genetic effects in these associated loci, influence BMD mainly within an age specific range. In addition, we identified variants in 18 loci with differential effects in children as compared to elderly individuals. Of these, variants in at least 13 loci evidenced a linear trend of change throughout the life-course and the majority of the remaining loci seemed to be only important at an early age. Although preliminary and requiring additional evidence to support this assertion, prominence of different biological pathways across age strata do appear to follow age-specific patterns.

Our findings indicate that there are age-specific genetic effects, maybe as signature of the dynamic nature of gene expression or post-translational protein modifications [35-37].Three different BMD novel loci were detected in the age-bin meta-analyses. The 14q32, harboring the *RIN3* gene that showed only association in children and had previously been identified as influencing pediatric BMD both at the TB- and at the lower limbs [15]. The minor allele of the leading SNP rs726999866/A associated with higher BMD levels in children also shows statistical evidence of an effect decreasing by of 0.001-0.002 per year in the studied cohorts. This change in effect size with aging likely explains why BMD GWAS meta-analyses in adults have not previously detected this association signal. This same rs726999866/A allele has been associated with higher risk of Paget's disease[38]. RIN3 is known to affect osteoclast function[39], yet the reason for the strong association of variants within this gene and BMD in pediatric populations is still unknown.The rs71486878 intronic in *PTPN5* showed GWS association only in the 30-45 years age-bin. Nevertheless, the studies on this age-bin are of modest sample size (N<1,500) and given the MAF of 1.5%, there is a high likelihood for this finding to be spourius and replication in a larger sample size should confirm it. On the ontrary, the rs148982377 (MAF of 4%) mapping to 7q22.1, identified in the elderly population leads an association signal expanding approximately 300Kb where other variants in LD also show evidence of association. Although, this locus has never been associated in previous more powerful GWAS analyses for both lumbar spine (LS) and femoral neck BMD, in the last GEFOS meta-analysis[12], the leading SNP for this association was suggestively associated with LS-BMD ($P=1.9x10^{-6}$). Moreover, rs11761528 ($r^2=0.42$, $P_{60>yrs}=1.9x10^{-5}$) has been previously associated with dehydroepiandrosterone sulphate (DHEAS) levels in the blood [40]. The minor allele associated with lower DHEAS concentration was also associated with lower TB-BMD. The positive correlation between DHEAS and BMD in adult and postmenopausal women have previously been described [41,42]. A longitudinal study showed that this effect might be explained by a decreased rate of bone loss[42].

By comparing summary statistic results from children and elderly subgroups, we found 18 different loci whose effects change with aging. The 6q25.1 locus demonstrated the most compelling evidence for age-dependent effects. Even if this locus, comprising *ESR1* and

CCDC170, has been shown to harbor complex allele heterogeneity, with at least three independent signals associated with BMD [43] and also heel bone properties[44], all the variants studied here showed larger effects in the elderly as compared to the pediatric populations. In the bin 0-15 years, the largest cohorts (i.e., Avon Longitudinal Study of Parents and Children (ALSPAC) and Generation R) comprise predominantly pre-pubertal children. As levels of estradiol in infancy are low [45], a negligible effect of *ESR1* variants on BMD is expected. The rs2982559 showed the strongest association with BMD of this locus in children (P=0.002), the minor allele of this SNP showed a marked increase of magnitude effect on BMD with age, particularly after 45 years, the age of menopause in women. Estrogen and BMD are positively correlated from late puberty onwards [46,47], being estrogen production a key factor in the maintenance of bone mineralization and prevention of osteoporosis in men and women [48]. Nevertheless, before, the influence of estrogen on longitudinal bone growth depends on maturational stage and its serum levels [49]. In line with estrogen exerting an effect in bone longitudinal growth two independent signals in *ESR1* have been identified to be associated with adult height[50].

On top of *RIN3*, detailed above in this section, variants within or in the neighborhood of other known bone-related genes as *BMP2, SUPT3H* or *SOX6* also demonstrated differential effects throughout the life course. Bone morphogenetic protein 2 (*BMP2*) is a growth factor that initiates osteoblast differentiation[51], which even if not GWS in the children meta-analysis, show a larger effect size in this population. *BMP2* variants have been previously associated with risk of osteoporosis in Icelandic and Danish families[51], an association that has not been replicated by later GWAS meta-analyses[12,13,52]. Nonetheless, BMP2 signaling seems to regulate peak bone mass in mice[53]. Variants in *BMP2* have additionally been associated with body mass index[54], waist-to-hip ratio[55] and height[50] through GWAS, underlying the role of this protein in development. Variants in *SUPT3H*, on the other hand, showed a greater effect on old adults, and indeed, it is in this population that association with lumbar spine BMD has been established[13] as well as with height[50]. It has been suggested that markers in this gene could regulate *RUNX2* expression in bone[56]. Together with *RUNX2, SOX6* plays a pivotal role in endochondrial ossification, through the regulation of chondroblast differentiation[57]. The reason for the larger effect of variants in these key factors of skeletal development on the elderly population rather than at an early age remains to be elucidated. The *WNT16/CPED1* age-effect dependency was not confirmed in our study. Nevertheless, this interaction was described for skull-BMD, and not TB-BMD [14]. Site-heterogeneity for variants within this locus have been previously shown in pediatric populations [15].

Variants with age-dependent effects have been detected at a genome-wide significant level for other complex traits using similar approaches [27,58]. By the implementation of the meta-regression, we have demonstrated here that age is a fundamental factor in explaining

the heterogeneity of the genetic effect for variants associated with BMD in our previous study (Chapter 4.2), and should also be investigated as a source of heterogeneity in other BMD meta-analyses. As described earlier, the age-stratification approach provides a chief advantage over the inclusion of an agexSNP interaction term on the study-specific GWAS analysis. Studies typically have a narrow age range (i.e., less of ten or fifteen years) and, therefore, contribute little information by using the interaction as a covariate [27]. This is mainly the case of the participant studies in our overall metanalysis.

However, some methodological issues could have influenced our findings. First of all, we may have introduced a cohort effect, meaning that environmental condition for cohorts at different age ranges might be different, particularly when comparing groups at the extremes of the age distribution. Second, a recent publication has raised the possibility that adjustment for correlated heritable covariates could lead to a (Collider) bias in the assessment of the genetic effect estimate, and this bias would be a function of the correlation of the outcome and the covariate [59]. Our analysis included as covariates height and weight, which are heritable complex traits and whose correlation with BMD varies across the lifespan. Additionally, SNP-by-age interaction for BMI has been shown, and could complicate further the evaluation of the possible effect bias in the reported differences[58]. Third, the linear correlation between age and the SNP effect could have been affected by the relative larger influence of studies in the two extremes of the age distribution (i.e., children and elderly strata), where indeed the largest cohorts included in the meta-analysis are located. This aggregation could have pulled the significance of the identified trends, without the contribution of the smaller set of studies in the middle age-bins. Nevertheless, the different effects of these variants will still be seen in the extremes, but that aggregation makes it difficult to be certain if non-linear effects are present.

In summary, our study provides the first insight into genetic effects changing with aging, at a genome-wide level, on BMD spanning from childhood through late adulthood. Our current knowledge of genetic factors influencing BMD come mostly from large meta-analyses of BMD performed in elderly adults. Our results suggest that increasing the sample size of GWAS in younger populations would be an effective strategy for identification of novel variants affecting bone metabolism at early ages. Several of the age-dependent loci present stronger effects in children as compared to adults, which may reflect a higher accumulation of environmental factors on bone density in older adults, probably masking the genetic effects. Variants with BMD-effects that either increase or decrease during adult age might provide additional knowledge on bone biology mechanisms, which could lead to new strategies for the prevention or treatment of osteoporosis.

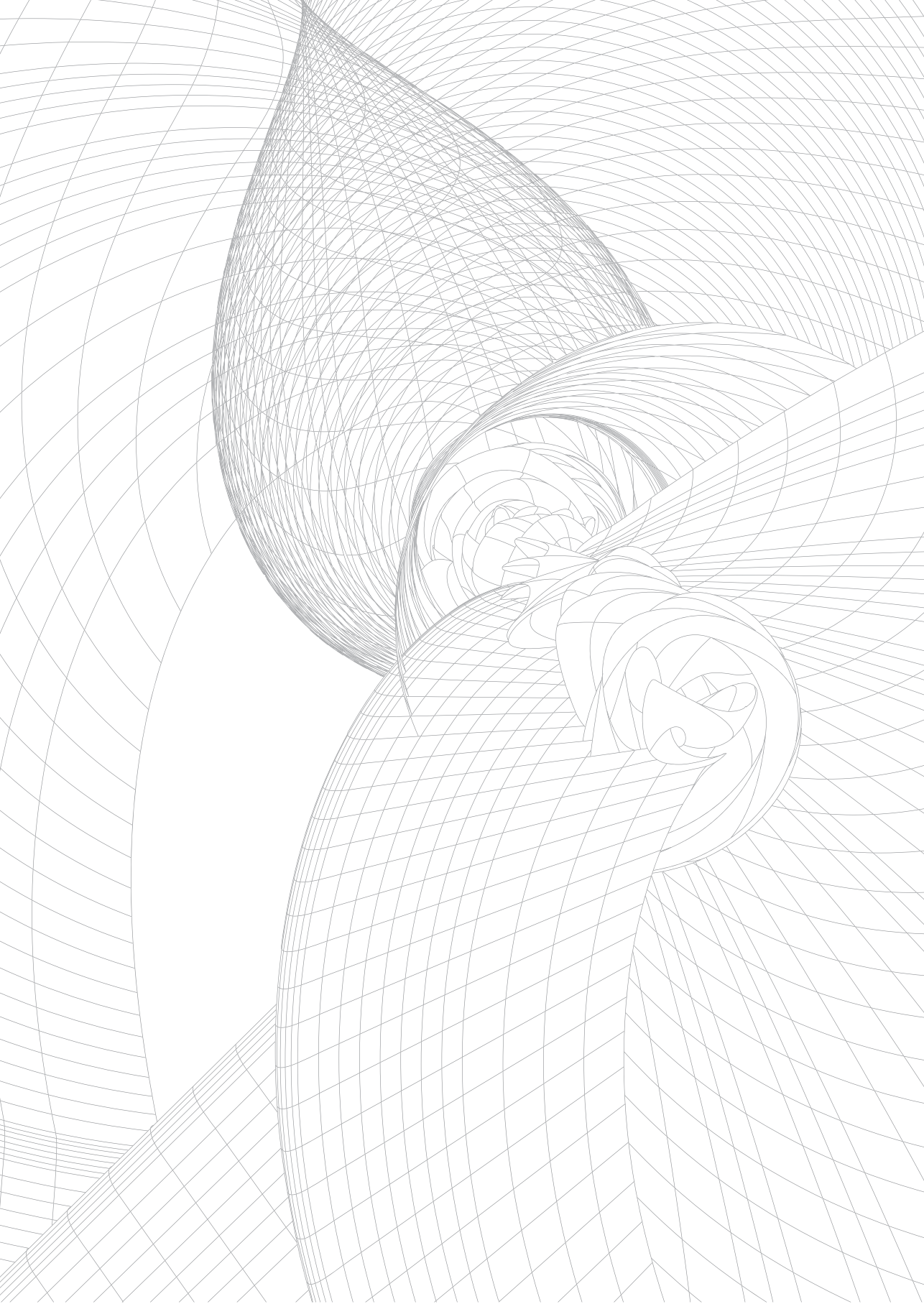
Additional detailed methods, results and acknowledgements are to be provided in the Online Supplement of the journal containing this publication.

REFERENCES

1. Turner CH. Three rules for bone adaptation to mechanical stimuli. **Bone**. 1998;23(5):399-407.
2. Clarke B. Normal bone anatomy and physiology. **Clin J Am Soc Nephrol**. 2008;3 Suppl 3:S131-9.
3. Sims NA, Martin TJ. Coupling the activities of bone formation and resorption: a multitude of signals within the basic multicellular unit. **Bonekey Rep**. 2014;3:481.
4. Farr JN, Khosla S. Skeletal changes through the lifespan--from growth to senescence. **Nat Rev Endocrinol**. 2015;11(9):513-21.
5. Hendrickx G, Boudin E, Van Hul W. A look behind the scenes: the risk and pathogenesis of primary osteoporosis. **Nat Rev Rheumatol**. 2015;11(8):462-74.
6. Arden NK, Baker J, Hogg C, Baan K, Spector TD. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. **J Bone Miner Res**. 1996;11(4): 530-4.
7. Slemenda CW, Turner CH, Peacock M, et al. The genetics of proximal femur geometry, distribution of bone mass and bone mineral density. **Osteoporos Int**. 1996;6(2):178-82.
8. Videman T, Levalahti E, Battie MC, Simonen R, Vanninen E, Kaprio J. Heritability of BMD of femoral neck and lumbar spine: a multivariate twin study of Finnish men. **J Bone Miner Res**. 2007;22(9): 1455-62.
9. Gueguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, Siest G. Segregation analysis and variance components analysis of bone mineral density in healthy families. **J Bone Miner Res**. 1995;10(12):2017-22.
10. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. Effects of age on genetic influence on bone loss over 17 years in women: the Healthy Ageing Twin Study (HATS). **J Bone Miner Res**. 2012;27(10):2170-8.
11. Richards JB, Zheng HF, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges. **Nat Rev Genet**. 2012;13(8):576-88.
12. Zheng HF, Forgetta V, Hsu YH, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. **Nature**. 2015;526(7571):112-7.
13. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
14. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **Plos Genet**. 2012;8(7):e1002718.
15. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **Plos Genet**. 2014;10(6):e1004423.
16. Massart F, Marini F, Bianchi G, et al. Age-specific effects of estrogen receptors' polymorphisms on the bone traits in healthy fertile women: the BONTURNO study. **Reprod Biol Endocrinol**. 2009;7:32.
17. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
18. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**. 2012;491(7422):56-65.
19. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. **Genet Epidemiol**. 2010;34(8):816-34.
20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. **Nat Genet**. 2012;44(8):955-+.
21. 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(6).
22. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. **Nat Genet**. 2007;39(7):906-13.

23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**. 2007;81(3):559-75.
24. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. **Bmc Bioinformatics**. 2010;11.
25. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. **Nat Protoc**. 2014;9(5):1192-212.
26. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. **Bioinformatics**. 2015;31(2):259-61.
27. Simino J, Shi G, Bis JC, et al. Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. **Am J Hum Genet**. 2014;95(1):24-38.
28. R Development Core Team. R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.
29. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. **J Stat Softw**. 2010;36(3):1-48.
30. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. **Nat Commun**. 2015;6.
31. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biol**. 2004;5(10):R80.
32. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. **Nat Protoc**. 2009;4(7):1073-82.
33. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. **Nat Methods**. 2010;7(4):248-9.
34. Gong Y, Slee RB, Fukai N, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. **Cell**. 2001;107(4):513-23.
35. Power RA, Iwaniec UT, Wronski TJ. Changes in gene expression associated with the bone anabolic effects of basic fibroblast growth factor in aged ovariectomized rats. **Bone**. 2002;31(1):143-8.
36. D'Aquila P, Rose G, Bellizzi D, Passarino G. Epigenetics and aging. **Maturitas**. 2013;74(2):130-6.
37. Vrtacnik P, Marc J, Ostanek B. Epigenetic mechanisms in bone. **Clin Chem Lab Med**. 2014;52(5):589-608.
38. Albagha OM, Wani SE, Visconti MR, et al. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. **Nat Genet**. 2011;43(7):685-9.
39. Vallet M, Soares DC, Wani S, et al. Targeted sequencing of the Paget's disease associated 14q32 locus identifies several missense coding variants in RIN3 that predispose to Paget's disease of bone. **Hum Mol Genet**. 2015;24(11):3286-95.
40. Zhai GJ, Teumer A, Stolk L, et al. Eight Common Genetic Variants Associated with Serum DHEAS Levels Suggest a Key Role in Ageing Mechanisms. **Plos Genet**. 2011;7(4).
41. Osmanagaoglu MA, Okumus B, Osmanagaoglu T, Bozkaya H. The relationship between serum dehydroepiandrosterone sulfate concentration and bone mineral density, lipids, and hormone replacement therapy in premenopausal and postmenopausal women. **J Womens Health**. 2004;13(9):993-9.
42. Ghebre MA, Hart DJ, Hakim AJ, et al. Association between DHEAS and Bone Loss in Postmenopausal Women: A 15-Year Longitudinal Population-Based Study. **Calcified Tissue Int**. 2011;89(4):295-302.
43. Koller DL, Zheng HF, Karasik D, et al. Meta-analysis of genome-wide studies identifies WNT16 and ESR1 SNPs associated with bone mineral density in premenopausal women. **J Bone Miner Res**. 2013;28(3):547-58.
44. Moayyeri A, Hsu YH, Karasik D, et al. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. **Hum Mol Genet**. 2014;23(11):3054-68.
45. Courant F, Aksglaede L, Antignac JP, et al. Assessment of circulating sex steroid levels in prepubertal and pubertal boys and girls by a novel ultrasensitive gas chromatography-tandem mass spectrometry method. **J Clin Endocrinol Metab**. 2010;95(1):82-92.

46. Tobias JH, Steer CD, Vilarino-Guell C, Brown MA. Estrogen receptor alpha regulates area-adjusted bone mineral content in late pubertal girls. **J Clin Endocrinol Metab**. 2007;92(2):641-7.
47. Rochira V, Kara E, Carani C. The endocrine role of estrogens on human male skeleton. **Int J Endocrinol**. 2015;2015:165215.
48. Riggs BL. The mechanisms of estrogen regulation of bone resorption. **J Clin Invest**. 2000;106(10): 1203-4.
49. Borjesson AE, Lagerquist MK, Windahl SH, Ohlsson C. The role of estrogen receptor alpha in the regulation of bone and growth plate cartilage. **Cell Mol Life Sci**. 2013;70(21):4023-37.
50. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. **Nat Genet**. 2014;46(11):1173-86.
51. Styrkarsdottir U, Cazier JB, Kong A, et al. Linkage of osteoporosis to chromosome 20p12 and association to BMP2. **PLoS Biol**. 2003;1(3):E69.
52. Ichikawa S, Johnson ML, Koller DL, et al. Polymorphisms in the bone morphogenetic protein 2 (BMP2) gene do not affect bone mineral density in white men or women. **Osteoporos Int**. 2006;17(4):587-92.
53. Bragdon B, Bonor J, Shultz KL, Beamer WG, Rosen CJ, Nohe A. Bone morphogenetic protein receptor type Ia localization causes increased BMP2 signaling in mice exhibiting increased peak bone mass phenotype. **J Cell Physiol**. 2012;227(7):2870-9.
54. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. **Nature**. 2015;518(7538):197-206.
55. Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. **Nature**. 2015;518(7538):187-96.
56. Barutcu AR, Tai PW, Wu H, et al. The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter. **Nucleic Acids Res**. 2014;42(16): 10360-72.
57. Zhang Y, Yang TL, Li X, Guo Y. Functional analyses reveal the essential role of SOX6 and RUNX2 in the communication of chondrocyte and osteoblast. **Osteoporos Int**. 2015;26(2):553-61.
58. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. **Plos Genet**. 2015;11(10): e1005378.
59. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. **Am J Hum Genet**. 2015;96(2):329-39.



Chapter 4.4

Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment

John P Kemp*, Carolina Medina-Gomez*, Karol Estrada, Beate St Pourcain, Denise HM Heppe, Nicole M. Warrington, Ling Oei, Susan M Ring, Claudia J Kruithof, Nicholas J Timpson, Lisa E Wolber, Sjur Reppe, Kaare Gautvik, Elin Grundberg, Bing Ge, Bram van der Eerden, Jeroen van de Peppel, Matthew A Hibbs, Cheryl L Ackert-Bicknell, Kwangbom Choi, Daniel L. Koller, Michael J. Econs, Frances M K Williams, Tatiana Foroud, M. Carola Zillikens, Claes Ohlsson, Albert Hofman, André G. Uitterlinden, George Davey Smith, Vincent WV Jaddoe, Jonathan H Tobias, Fernando Rivadeneira, and David M Evans.

*These authors equally contributed to this manuscript

PLoS Genet. 2014 Jun 19;10(6):e1004423. doi: 10.1371/journal.pgen.1004423.

ABSTRACT

Heritability of bone mineral density (BMD) varies across skeletal sites, reflecting different relative contributions of genetic and environmental influences. To quantify the degree to which common genetic variants tag and environmental factors influence BMD, at different sites, we estimated the genetic (r_g) and residual (r_e) correlations between BMD measured at the upper limbs (UL-BMD), lower limbs (LL-BMD) and skull (SK-BMD), using total-body DXA scans of ~4,890 participants recruited by the Avon Longitudinal Study of Parents and their Children (ALSPAC). Point estimates of r_g indicated that appendicular sites have a greater proportion of shared genetic architecture (LL-/UL-BMD r_g=0.78) between them, than with the skull (UL-/SK-BMD r_g=0.58 and LL-/SK-BMD r_g=0.43). Likewise, the residual correlation between BMD at appendicular sites (r_e=0.55) was higher than the residual correlation between SK-BMD and BMD at appendicular sites (r_e=0.20 – 0.24). To explore the basis for the observed differences in r_g and r_e, genome-wide association meta-analyses were performed (n~9,395), combining data from ALSPAC and the Generation R Study identifying 15 independent signals from 13 loci associated at genome-wide significant level across different skeletal regions. Results suggested that previously identified BMD-associated variants may exert site-specific effects (i.e., differ in the strength of their association and magnitude of effect across different skeletal sites). In particular, variants at *CPED1* exerted a larger influence on SK-BMD and UL-BMD when compared to LL-BMD (P=2.01×10^{-37}), whilst variants at *WNT16* influenced UL-BMD to a greater degree when compared to SK- and LL-BMD (P=2.31×10^{-14}). In addition, we report a novel association between *RIN3* (previously associated with Paget's disease) and LL-BMD (rs754388: β=0.13, SE=0.02, P=1.4×10^{-10}). Our results suggest that BMD at different skeletal sites is under a mixture of shared and specific genetic and environmental influences. Allowing for these differences by performing genome-wide association at different skeletal sites may help uncover new genetic influences on BMD.

AUTHOR SUMMARY

The heritability of bone mineral density (BMD) varies across skeletal sites, reflecting different relative contributions of genetic and environmental influences. To investigate whether the genes underlying bone acquisition act in a site-specific manner, we quantified the shared genetic influences across axial and appendicular skeletal sites by estimating the genetic and residual correlation of BMD at the upper limb, lower limb and the skull. Our results suggest that different skeletal sites as measured by total-body Dual-Energy X-Ray Absorptiometry are to a certain extent under distinct genetic and environmental influences. To further explore the basis for these differences, genome-wide association meta-analyses were performed to identify genetic loci that are preferentially associated with one or more

skeletal regions. Variants at 13 loci (including RIN3, a novel BMD associated locus) reached genome-wide significance and several displayed evidence of differential association with BMD across the different skeletal sites in particular *CPED1* and *WNT16*. Our results suggest that it may be advantageous to decompose the total-body BMD measures and perform GWAS at separate skeletal regions. By allowing for site-specific differences, new genetic variants affecting BMD and future risk of osteoporosis may be uncovered.

INTRODUCTION

Bone mineral density (BMD) at the femoral neck and lumbar spine [as measured by dual-energy X-ray absorptiometry, (DXA)], represents the primary diagnostic marker for osteoporosis as it serves as a good predictor of bone strength and fracture risk in adults [1]. Bone strength and fracture risk are influenced by: i) bone acquisition in childhood, adolescence and young adulthood ii) the subsequent maintenance of bone mass over the life course and iii) the progressive loss of bone in later life [2,3]. Large-scale genome-wide association studies (GWAS) using adult-BMD measured at the femoral neck (FN) and lumbar spine (LS) have successfully identified variants in 56 loci explaining 4-5% of the phenotypic variance in adult-BMD [4-6]. However, it is possible that the genetic variants influencing bone acquisition are different from the ones involved in bone maintenance and bone loss across the life course. Consequently, GWAS using paediatric-BMD measurements have recently been performed with the goal of identifying novel genetic variants primarily associated with bone acquisition, whilst limiting the noise introduced by bone maintenance and bone loss [7]. This approach has resulted in the successful identification of novel BMD associated variants in the *WNT16* [7] and Osterix (*SP7*) loci [8] and it is highly likely that more variants will be discovered as the sample size of these paediatric studies increases.

In growing children, changes in bone area create artefacts influencing the reproducibility, comparability and interpretation of DXA measurements. For this reason, regions of interest (ROI) containing larger bone areas [i.e., total-body, (TB)], which are less prone to these artefacts, are preferred for paediatric evaluations of bone health [9]. The skull region is generally excluded from TB-DXA scans as its relative contribution to bone mass is proportionally larger with respect to the rest of the body in children, and its inclusion has been shown to make diagnostic interpretation difficult [10]. However, from a locus discovery perspective, it may be advantageous to partition TB-DXA further into different regions, such as the upper and lower limbs and the skull. This is important if genetic heterogeneity exists in terms of loci differentially affecting BMD at different skeletal sites, or whose effect is greater at some locations than in others. Considering that environmental factors (i.e., mechanical loading) influence skeletal sites differently, analysis of skull-BMD may be particularly informative and

even provide greater power to identify genetic variants. This is the case given that the skull is less influenced by mechanical loading than appendicular and other axial sites. Further, the skull is frequently affected in monogenic conditions involving the skeleton. For example, craniofacial abnormalities such as thickening of the cranium and skull base are cardinal features of van Buchems disease, Sclerosteosis and other sclerosing bone dysplasias [11,12].

In the current study we examined whether genetic factors influence bone mass accrual in a site-specific manner, by performing regional analysis of TB-DXA scans, focusing on the total-body less head (TBLH), lower limb (LL), upper limb (UL), and skull (SK) regions. Using genome-wide complex trait analysis (GCTA) on participants from the Avon Longitudinal Study of Parents and their Children (ALSPAC), we assessed the proportion of BMD variance explained by common genetic variants, across each sub-region and additionally determined the shared genetic and residual correlation between each sub-region. Subsequently, we performed a genome-wide association (GWA) meta-analysis of BMD at each skeletal site in the ALSPAC and Generation R studies and went on to identify factors, which preferentially influence one or more skeletal regions.

RESULTS

Phenotypic correlation and genome-wide complex trait analysis of BMD at different regions

Univariate GCTA analysis revealed that common genotyped variants explained a greater proportion of the variance in SK-BMD (v_g = 0.51, SE = 0.07, $P = 2.0\times10^{-13}$) than LL- (v_g = 0.40, SE = 0.07, $P = 8.0\times10^{-9}$) or UL- (v_g = 0.39, SE = 0.07, $P = 2.0\times10^{-8}$) BMD. Higher *phenotypic* correlations were observed when comparing LL- and UL-BMD than with SK-BMD (Table 1). Simi-

Table 1. Bivariate GCTA estimates of the genetic and residual correlations for bone mineral density measurements at the total-body less head, lower limb, upper limb and skull for the ALSPAC cohort. TBLH-BMD = total body less head BMD, (LL-BMD) = lower limb BMD, (UL-BMD) = upper limb BMD, (SK-BMD) = skull BMD, r_g = genetic correlation between trait 1 and trait 2. r_e = residual correlation between trait 1 and trait 2. All traits, excluding SK-BMD were adjusted for age, gender and weight. SK-BMD was adjusted for age, gender and height. *P*-refers to the *P*-value for the likelihood ratio test of whether $r_g = 0$. Phenotypic correlations (r_p) were as follows: SK-BMD/TBLH-BMD (r_p = 0.40, SE = 0.013, P < 0.001), SK-BMD/LL-BMD (r_p = 0.31, SE = 0.013, P < 0.001), SK-BMD/UL-BMD (r_p = 0.40, SE = 0.013, P < 0.001) and LL-BMD/UL-BMD (r_p = 0.64, SE = 0.010, P < 0.001).

TRAIT 1	TRAIT2	SAMPLE SIZE	r_g	SE	r_e	SE	*P*
SK-BMD	**TBLH-BMD**	9732	0.52	0.088	0.29	0.086	4.1×10^{-6}
	LL-BMD	9732	0.44	0.099	0.20	0.088	1.2×10^{-3}
	UL-BMD	9732	0.58	0.090	0.24	0.085	9.1×10^{-7}
LL-BMD	**UL-BMD**	9782	0.78	0.067	0.55	0.055	1.4×10^{-7}

larly, bivariate GCTA analysis indicated that the strongest genetic correlation was between BMD at the two appendicular sites, whereas the *genetic* correlations involving SK-BMD were more moderate. The *residual* correlation between the different sites was in general smaller than the *genetic* correlation, and was higher for BMD between the appendicular sites than for comparisons involving the skull (Table 1). Highly similar magnitudes and patterns of residual correlations were obtained for a sensitivity analysis in which BMD at all skeletal sites was corrected for age, gender, weight and height (Online Resource Table S1).

Genome Wide Meta-Analysis of BMD across different skeletal regions in ALSPAC and Generation R

Genome-wide association meta-analyses were performed on TBLH-, LL-, UL- and SK-BMD, using regional BMD data derived from ~9,395 TB-DXA scans. Detailed population characteristics of the ALSPAC and Generation R cohorts are summarised in Tables S2 and S3. Summary statistics from each GWAS (after meta-analysis) indicated that negligible systematic inflation of test statistics was observed (META λ_{GC} = 1.01 – 1.03). In contrast, a marked deviation from the null was observed in the tail of the distribution amongst the lowest observed *P*-values of the meta-association analyses (Online Resource Figure S1). SNPs in thirteen published BMD-associated loci exceeded the genome-wide significance (GWS) threshold for association ($P \leq 5\times10^{-8}$, Table 2). They included variants which mapped close to, or within: *WNT4* (1p36.12), *WNT16/FAM3C/CPED1* (7q31.31) for all skeletal sites measured, *EYA4* (6q23.2), *COLEC10/TNFRSF11B* (8q24.12), *LIN7C/LGR4* (11p14.1), *PPP6R3/LRP5* (11q13.2) and *TNFRSF11A* (18q21.33) for SK-BMD, *CENPW/RSPO3* (6q22.32) for UL- and SK-BMD, *TNFSF11* (13q14.11) and *GALNT3* (2q24.3) for UL- and TBLH-BMD. In addition, variants proximal to or within *FUBP3* (9q34.11) and *KLHDC5/PTHLH* (12p11.22) were associated with TBLH- and LL-BMD. Furthermore, a novel signal (top SNP rs754388, 14q32.12), located within Ras and Rab interactor 3 (*RIN3*) achieved genome-wide significance after meta-analysis of LL-BMD (β = 0.13, SE = 0.02, $P = 1.4\times10^{-10}$, Figure 1-I, Table 2) and TBLH-BMD (β = 0.12, SE = 0.02, $P = 3.0\times10^{-9}$, Table 2 and Online Resource Figure S2). The full list of all genome-wide significant SNPs and regional association plots for each locus and skeletal site are presented in Online Resources Tables S4 – S7 and Figures S2 - S5.

A followup of 66 independent SNPs at 58 loci, previously associated with BMD [4,13], indicated that 31 loci showed nominal evidence of association ($P < 0.05$) with TBLH-BMD, 28 with LL-BMD, 26 with UL-BMD and 26 with SK-BMD (versus an expectation of 3.3. per phenotype) (Online Resource Table S8). A similar distribution of associations was also observed when a more conservative threshold considering multiple hypothesis testing was adopted that took into account the fact that 66 variants and four phenotypes had been tested (i.e., $\alpha < 1.9\times10^{-4}$). Using this threshold nine variants showed evidence of association with TBLH-BMD, seven with LL-BMD, six with UL-BMD and 10 with SK-BMD (versus an

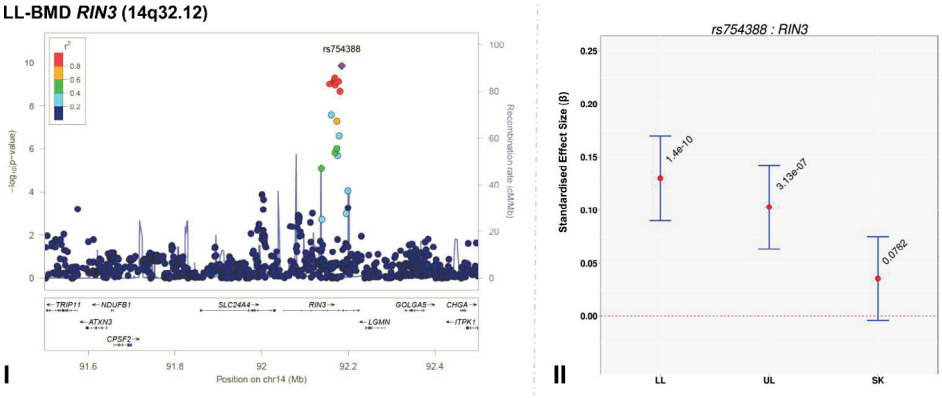


Figure 1: Regional association plot of the primary signal (rs754388) associate with lower limb-BMD at 14q32.12, in addition to a comparison of the effect of rs754388 on bone mineral density at three different skeletal sites. For I and II: Circles show GWA meta-analysis P-values and positions of SNPs found within the 14q32.12 locus. The top SNP, i.e., rs754388, is denoted by a diamond. Different colors indicate varying degrees of pair-wise linkage disequilibrium (HapMap 2 CEPH) between the top SNP and all other SNPs. For II: The per-allele effect in standard deviations (SD) (red dot) and the 95% confidence interval (error bar) of rs754388 for lower limb (LL), upper limb (UL) and skull (SK) BMD, plotted with the strength of evidence against the null hypothesis of no association.

expectation of 0.1 per phenotype). We note that in all cases where nominal significance was reached, the direction of effect was consistent with previous studies.

To ensure that our results were robust to the possible effects of population stratification and our choice of covariates, we performed sensitivity analyses where we either restricted our analysis to white European individuals only, or adjusted for the same set of covariates across all analyses (i.e., age, gender, height, and weight). Similar effect sizes and patterns of association were observed for the top SNPs when adjusting BMD measures of all four regions for age, gender, height and weight (Model 1a, Online Resource Table S9) and when limiting the GWAS meta-analysis to individuals of European ancestry (Model 1b, Online Resource Table S9). In both sensitivity analyses, no additional loci reached the threshold of genome-wide association (Online Resource Figure S6 and S7).

Identification of novel BMD-associated signals

We assessed the presence of novel secondary association signals at loci that contained genome-wide associated variants. Meta-analysis of conditional association analyses resulted in the attenuation of the majority of our top association signals (Online Resource Table S10, Figures S2 - S5), indicating that these loci were not independent from signals previously reported by other BMD GWAS. However, the top signal for SK-BMD (rs2130604, $\beta = 0.11$, SE = 0.02, $P = 3.3\times10^{-11}$), mapping near *RSPO3*, but closest to *CENPW* (previously known as *C6orf173*, 6q22.32, Figure 2A-I, Table 2) was only marginally attenuated after conditional

analysis (rs2130604, β = 0.10, SE = 0.02, $P = 7.1\times10^{-9}$, Figure 2A-II, Online Resource Table S10). This suggests that rs2130604 is largely independent from the previously reported signal at *RSPO3* (rs13204965, 6q22.32), which was identified in a GWAS of individuals with extremely high or low BMD at the hip [12] and later replicated in the second GEnetic Factors for OSteoporosis Consortium (GEFOS-II) BMD meta-analysis [4]. This observation is further supported by low estimates of LD ($r^2 = 0.14$) between rs2130604 and rs13204965. Furthermore, the secondary signal (after conditional analysis) reached the estimated significance threshold of association after multiple testing correction (i.e., $P \leq 7.2\times10^{-5}$).

Interestingly, after conditioning rs4418209 (another SNP in the same locus) on the published BMD-associated SNP rs13204965, we observed a marked increase in its evidence of association [(β = 0.07, SE = 0.02, $P = 1.1\times10^{-6}$) *before* and (β = 0.09, SE = 0.01, $P = 7.9\times10^{-10}$) *after*], (Figure 2A-II, Online Resource Table S10). The rs4418209 variant maps closest to *CENPW* (6q22.32) and is in moderate LD with the secondary independent signal (rs2130604, r^2 = 0.43) and in low LD with the published *RSPO3* SNP (rs13204965, r^2 = 0.12). Whilst no other

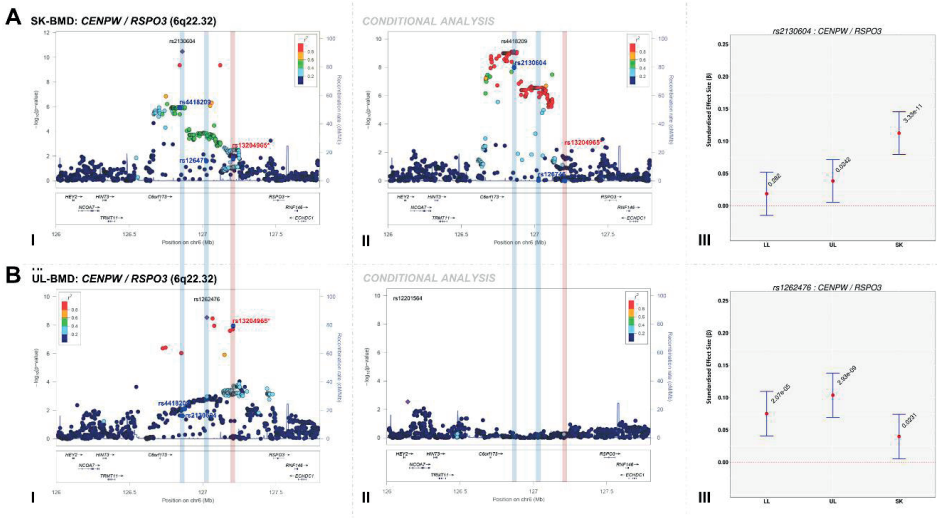


Figure 2: Regional association plots of the top skull- and upper limb-BMD associated SNPs at the 6q22.32 locus before and after conditioning on published SNP (rs13204965*) in addition to a comparison of the effect sizes of the top skull- (rs2130604) and upper limb-BMD (rs1262476) associated SNP (before conditional analysis) on BMD at three different skeletal sites. For I and II: Circles show GWA meta-analysis P-values and positions of SNPs found within each locus. Top SNPs are denoted by diamonds. Different colors indicate varying degrees of pairwise linkage disequilibrium (HapMap 2 CEPH) between the top SNP and all other SNPs. Blue vertical shaded areas indicate the position of rs2130604 (top SNP A-I) and rs1262476 (top SNP B-I) for each analysis. The red vertical shaded area represents the position of the published SNP (rs13204965*). Rsids of relevant SNPs (blue dots) have been provided. For III: The per allele effect in SD (red dot) and 95% confidence intervals (error bar) of each top SNP (before conditional analysis) for lower limb (LL), upper limb (UL) and skull (SK) BMD are plotted with their specific strength of evidence against the null hypothesis of no association. Please note: *RSPO3* is also found in the 6q.22.32 locus containing *CENPW*

Table 2. Top genome-wide significant SNPs associated with bone mineral density of the total-body less head, lower limb, upper limb and skull. (TBLH-BMD) = total-body less head BMD, (LL-BMD) = lower limb BMD, (UL-BMD) = upper limb BMD, (SK-BMD) = skull BMD, (GENE) = closest gene, (POS) = position in the genome based on hg18, (EAF) = effect allele frequency, (β) = estimates of effect size expressed as adjusted SD per copy of the effect allele (EA), (SE) = standard error of β, (P) = pvalue, (I^2) = Cochran's Q statistic evaluating heterogeneity, (P_{HET}) = evidence of heterogeneity and *Sample sizes used for SK-BMD genome-wide meta-analysis. **Please note that *PTHLH* is also located at the 12p11.22 locus containing *KLHDC5, RSPO3* is also located at the 6q.22.32 locus containing *CENPW, FAM3C and CPED1* are also located at the 7q.31.31 locus containing *WNT16*, *TNFRSF11B* is also located at the 8q.24.12 locus containing *COLEC10, LGR4* is also located at the 11p14.1 locus containing *LIN7C* and *LRP5* is also located at the 11q13.2 locus containing *PPP6R3*.

	RSID	LOCUS	POS	GENE	EA	ALSPAC (n = 5,330 / 5,299*)				Generation R (n = 4,086)				META-ANALYSIS (n = 9,416 / 9,385*)					
TRAIT						EAF	β	SE	P	EAF	β	SE	P	EAF	β	SE	P	I^2	P_{HET}
TBLH-BMD	rs3765350	1p36.12	22319903	*WNT4*	A	0.78	0.106	0.023	5.75×10^{-6}	0.78	0.109	0.026	2.92×10^{-5}	0.78	0.107	0.017	$\mathbf{7.04\times10^{-10}}$	0	9.32×10^{-1}
	rs6726821	2q24.3	166286360	*GALNT3*	T	0.51	0.094	0.019	1.32×10^{-6}	0.58	0.087	0.022	8.76×10^{-5}	0.54	0.091	0.015	$\mathbf{3.95\times10^{-10}}$	0	8.11×10^{-1}
	rs7776725	7q31.31	120820357	*FAM3C***	C	0.27	0.136	0.023	$\mathbf{3.65\times10^{-9}}$	0.26	0.188	0.026	$\mathbf{7.67\times10^{-13}}$	0.27	0.159	0.017	$\mathbf{5.67\times10^{-20}}$	54.7	1.38×10^{-1}
	rs7466269	9q34.11	132453905	*FUBP3*	A	0.64	0.094	0.020	3.72×10^{-6}	0.66	0.072	0.023	2.01×10^{-3}	0.65	0.084	0.015	$\mathbf{3.26\times10^{-8}}$	0	4.74×10^{-1}
	rs4420311	12p11.22	27875457	*KLHDC5***	G	0.47	0.080	0.020	7.84×10^{-5}	0.44	0.092	0.024	1.03×10^{-4}	0.46	0.085	0.016	$\mathbf{4.44\times10^{-8}}$	0	7.03×10^{-1}
	rs17536328	13q14.11	42041029	*TNFSF11*	T	0.43	0.079	0.020	6.14×10^{-5}	0.40	0.095	0.022	1.53×10^{-5}	0.42	0.086	0.015	$\mathbf{7.58\times10^{-9}}$	0	5.94×10^{-1}
	rs754388	14q32.12	92185163	*RIN3*	C	0.81	0.098	0.026	1.34×10^{-4}	0.83	0.149	0.031	1.41×10^{-6}	0.82	0.120	0.020	$\mathbf{2.96\times10^{-9}}$	36.0	2.11×10^{-1}
LL-BMD	rs3765350	1p36.12	22319903	*WNT4*	A	0.78	0.103	0.023	1.05×10^{-5}	0.78	0.090	0.026	5.74×10^{-4}	0.78	0.097	0.018	$\mathbf{2.89\times10^{-8}}$	0	7.12×10^{-1}
	rs2908004	7q31.31	120757005	*WNT16***	A	0.44	0.093	0.020	3.63×10^{-6}	0.50	0.108	0.022	1.25×10^{-6}	0.47	0.100	0.015	$\mathbf{3.01\times10^{-11}}$	0	6.19×10^{-1}
	rs7466269	9q34.11	132453905	*FUBP3*	A	0.64	0.097	0.020	1.85×10^{-6}	0.66	0.074	0.023	1.59×10^{-3}	0.65	0.087	0.015	$\mathbf{1.51\times10^{-8}}$	0	4.57×10^{-1}
	rs4420311	12p11.22	27875457	*KLHDC5***	G	0.47	0.086	0.020	2.06×10^{-5}	0.44	0.087	0.024	2.28×10^{-4}	0.46	0.086	0.016	$\mathbf{3.21\times10^{-8}}$	0	9.75×10^{-1}
	rs754388	14q32.12	92185163	*RIN3*	C	0.81	0.119	0.026	3.47×10^{-6}	0.83	0.145	0.031	2.51×10^{-6}	0.82	0.130	0.020	$\mathbf{1.40\times10^{-10}}$	0	5.26×10^{-1}
UL-BMD	rs2235529	1p36.12	22323074	*WNT4*	C	0.84	0.099	0.027	1.98×10^{-4}	0.85	0.14	0.031	5.99×10^{-6}	0.85	0.117	0.021	$\mathbf{1.21\times10^{-8}}$	0	3.22×10^{-1}
	rs6726821	2q24.3	166286360	*GALNT3*	T	0.51	0.078	0.019	6.44×10^{-5}	0.58	0.089	0.022	5.61×10^{-5}	0.54	0.083	0.015	$\mathbf{1.13\times10^{-8}}$	0	7.07×10^{-1}
	rs1262476	6q22.32	127028689	*CENPW***	G	0.76	0.130	0.022	$\mathbf{6.37\times10^{-9}}$	0.79	0.062	0.028	2.37×10^{-2}	0.77	0.104	0.018	$\mathbf{2.93\times10^{-9}}$	72.3	5.76×10^{-2}
	rs798943	7q31.31	120546135	*CPED1***	G	0.61	0.187	0.020	$\mathbf{8.84\times10^{-21}}$	0.62	0.205	0.023	$\mathbf{1.28\times10^{-19}}$	0.61	0.195	0.015	$\mathbf{1.47\times10^{-37}}$	0	5.57×10^{-1}
	rs9525638	13q14.11	42026577	*TNFSF11*	C	0.43	0.094	0.020	1.63×10^{-6}	0.41	0.083	0.022	1.52×10^{-4}	0.42	0.089	0.015	$\mathbf{2.47\times10^{-9}}$	0	7.13×10^{-1}
SK-BMD	rs3920498	1p36.12	22365474	*WNT4*	G	0.79	0.144	0.024	$\mathbf{4.56\times10^{-9}}$	0.82	0.118	0.030	8.40×10^{-5}	0.80	0.134	0.019	$\mathbf{1.56\times10^{-12}}$	0	5.01×10^{-1}

Table 2 (continued)

	RSID	LOCUS	POS	GENE	EA	ALSPAC (n = 5,330 / 5,299*)				Generation R (n = 4,086)				META-ANALYSIS (n = 9,416 / 9,385*)					
TRAIT						EAF	β	SE	*P*	EAF	β	SE	*P*	EAF	β	SE	*P*	I^2	P_{HET}
	rs2130604	6q22.32	126862254	*CENPW***	T	0.24	0.117	0.022	1.89×10^{-7}	0.23	0.106	0.026	6.47×10^{-5}	0.24	0.112	0.017	$\mathbf{3.33\times10^{-11}}$	0	7.48×10^{-1}
	rs3012465	6q23.2	133392629	*EYA4*	G	0.65	0.125	0.020	$\mathbf{7.00\times10^{-10}}$	0.69	0.129	0.023	$\mathbf{3.06\times10^{-8}}$	0.67	0.127	0.015	$\mathbf{8.29\times10^{-17}}$	0	8.96×10^{-1}
	rs13223036	7q31.31	120534544	*CPED1***	T	0.63	0.170	0.020	$\mathbf{3.09\times10^{-17}}$	0.65	0.167	0.023	$\mathbf{6.21\times10^{-13}}$	0.64	0.169	0.015	$\mathbf{1.53\times10^{-28}}$	0	9.22×10^{-1}
	rs2450083	8q24.12	120132723	*COLEC10***	T	0.48	0.105	0.020	1.66×10^{-7}	0.47	0.098	0.023	2.16×10^{-5}	0.47	0.102	0.015	$\mathbf{2.13\times10^{-11}}$	0	8.20×10^{-1}
	rs10835187	11p14.1	27462253	*LIN7C***	C	0.45	0.145	0.020	1.05×10^{-13}	0.50	0.106	0.022	1.63×10^{-6}	0.47	0.127	0.015	$\mathbf{1.63\times10^{-17}}$	41.1	1.93×10^{-1}
	rs12272917	11q13.2	68019946	*PPP6R3***	T	0.74	0.130	0.022	$\mathbf{4.01\times10^{-9}}$	0.76	0.080	0.026	2.52×10^{-3}	0.75	0.109	0.017	$\mathbf{1.34\times10^{-10}}$	53.0	1.45×10^{-1}
	rs884205	18q21.33	58205837	*TNFRSF11A*	C	0.72	0.092	0.023	5.38×10^{-5}	0.80	0.123	0.030	3.88×10^{-5}	0.75	0.104	0.018	$\mathbf{1.84\times10^{-8}}$	0	4.15×10^{-1}

SNPs reached the threshold for declaring genome-wide significance ($P < 5\times10^{-8}$), variants from three loci still yielded suggestive evidence for association ($P < 1\times10^{-5}$) after conditional analyses (Online Resource Table S10 and Figures S2 - S5). They included: i) *KLHDC5/PTHLH* (rs4420311, 12p11.22) associated with TBLH- (β = 0.08, SE = 0.016, $P = 7.6\times10^{-7}$) and LL-BMD (β = 0.08, SE = 0.016, $P = 1.9\times10^{-6}$), ii) *TNFSF11* (rs17536328 and rs2148072, 13q14.11) associated with TBLH- (β = 0.08, SE = 0.015, $P = 5.6\times10^{-7}$) and UL-BMD (β = 0.07, SE = 0.015, $P = 2.1\times10^{-6}$) respectively and iii) *LIN7C/LGR4* [rs10160456, 11p14.1, (β = 0.07, SE = 0.015, $P = 7.8\times10^{-6}$)] with SK-BMD. After conditional analysis, the secondary signal at *LIN7C/LGR4* (i.e., rs10160456) mapped closest to *CCDC34* and not to *LIN7C*, the gene closest to the primary signal. All these three loci might represent novel secondary signals as the residual signal reached the predicted locus specific threshold of association after multiple testing correction (Online Resource Table S10). However, we cannot exclude that both associations (i.e., the primary and secondary signals) could potentially arise from their association with one or more causal variants, which could occur, on the same haplotype background. For example, one such BMD-associated rare variant has recently been identified in *LGR4* in Icelandic populations although this mutation appears specific to this population and therefore is unlikely to account for the *LIN7C/LGR4* signal we observe [14].

Comparison of the magnitude of the effect sizes of genome-wide significant SNPs across skeletal-sites

The standardized per allele effect sizes (β) of all the top BMD-associated SNPs were compared across three (SK-, UL-, and LL) BMD regions to determine if they preferentially influenced one or more skeletal sites (Table 3, Online Resource Figure S8 – S11). Effect sizes of the following variants: rs2130604 (*CENPW/RSPO3,* 6q22.32), rs3012465 (*EYA4,* 6q23.2), rs2450083 (*COLEC10/TNFRS11B,* 8q24.12), rs10835187 (*LIN7C/LGR4,* 11p14.1) and rs884205 (*TNFRSF11A,* 18q21.33) appeared to be largest for SK-BMD when compared to UL- and LL-BMD (Online Resource Figure S8). Furthermore, differences in the magnitude of the effect were evident when comparing independent genetic variants that occurred in close proximity within a locus, as shown at the *CENPW/RSPO3* (6q22.32) and *WNT16/FAM3C/CPED1* (7q31.31) loci. Specifically, the independent signal (rs2130604, *CENPW/RSPO3,* 6q22.32) associated with SK-BMD [β = 0.11 (CI_{95}: 0.08, 0.15) $P = 3.3\times10^{-11}$], was not strongly related to LL-BMD [β = 0.02 (CI_{95}: –0.02, 0.05), $P = 0.28$], or UL-BMD [β = 0.04, (CI_{95}: 0.01, 0.07), $P = 0.02$] (Table 3, Figure 2A-III). In contrast, a neighbouring SNP (rs1262476) primarily associated with UL-BMD appeared to influence BMD across all skeletal sites (Table 3, Figure 2B-III). Differential patterns of association between SNPs at neighbouring positions were also observed at the *WNT16* locus (Table 3, Figure 3A-C). Effect sizes were largest for UL-BMD at rs2908004 (*WNT16,* 7q31.31, Table 3, Figure 3A-II) when compared to SK- and LL-BMD. Interestingly, as compared to LL-BMD, we observed consistently larger effect sizes for rs13223036 and rs798943 (*CPED1*, previously known as *C7orf58*) for SK- and UL-BMD, (Table 3, Figure 3B-II and 3C-II).

Table 3. Comparison of effect sizes and the strength of association of all variants that exceeded genome-wide significance at one or more skeletal sites. (LL-BMD) = lower limb BMD, (UL-BMD) = upper limb BMD, (SK-BMD) = skull BMD, (GENE) = closest gene, (POS) = position in the genome based on hg18, (EA) = effect allele, (β) = estimates of effect size expressed as adjusted SD per copy of the effect allele (EA), (CI-L) = lower limit of the 95% confidence interval for β, (CI-U) = upper limit of the 95% confidence interval for β, (*P*) = *P*-value. *Please note that *PTHLH* is also found in the 12p.11.22 locus containing *KLHDC5, RSPO3* is also found in the 6q.22.32 locus containing *CENPW, TNFRSF11B* is also located at the 8q.24.12 locus containing *COLEC10, LGR4* is also located at the 11p.14.1 locus containing *LIN7C* and *LRP5* is also located at the 11q13.2 locus containing *PPP6R3. FAM3C* and *CPED1* are also located at the 7q.31.31 locus containing *WNT16*.

LOCUS	RSID	GENE	EA	LL-BMD				UL-BMD				S-BMD				Fisher's Product P*	
				β	CI-L	CI-U	*P*	β	CI-L	CI-U	*P*	β	CI-L	CI-U	*P*	Chi	*P*
1p36.12	rs3765350	*WNT4*	A	0.097	0.06	0.13	$2.89x10^{-8}$	0.091	0.06	0.13	$1.82x10^{-7}$	0.115	0.08	0.15	$3.56x10^{-11}$	9.71	$4.55x10^{-2}$
1p36.12	rs2235529	*WNT4*	C	0.106	0.07	0.15	$3.27x10^{-7}$	0.117	0.08	0.16	$1.21x10^{-8}$	0.143	0.10	0.18	$2.99x10^{-12}$	8.34	$7.99x10^{-2}$
1p36.12	rs3920498	*WNT4*	G	0.075	0.04	0.11	$8.41x10^{-5}$	0.097	0.06	0.13	$2.85x10^{-7}$	0.134	0.10	0.17	$1.56x10^{-12}$	8.93	$6.29x10^{-2}$
2q24.3	rs6726821	*GALNT3*	T	0.078	0.05	0.11	$1.03x10^{-7}$	0.083	0.05	0.11	$1.13x10^{-8}$	0.031	0.00	0.06	$3.37x10^{-2}$	15.26	$4.19x10^{-3}$
6q22.32	rs1262476	*CENPW*	G	0.075	0.04	0.11	$2.07x10^{-5}$	0.104	0.07	0.14	$2.93x10^{-9}$	0.040	0.01	0.07	$2.31x10^{-2}$	27.26	$1.76x10^{-5}$
6q22.32	rs2130604	*CENPW*	T	0.018	-0.02	0.05	$2.82x10^{-1}$	0.038	0.01	0.07	$2.42x10^{-2}$	0.112	0.08	0.15	$3.33x10^{-11}$	15.09	$4.51x10^{-3}$
6q23.2	rs3012465	*EYA4*	G	0.019	-0.01	0.05	$2.09x10^{-1}$	0.051	0.02	0.08	$7.45x10^{-4}$	0.127	0.10	0.16	$8.29x10^{-17}$	35.66	$3.40x10^{-7}$
7q31.31	rs13223036	*CPED1*	T	0.020	-0.01	0.05	$2.02x10^{-1}$	0.187	0.16	0.22	$1.25x10^{-34}$	0.169	0.14	0.20	$1.53x10^{-28}$	178.00	$2.01 x10^{-37}$
7q31.31	rs798943	*CPED1*	G	0.030	0.00	0.06	$5.17x10^{-2}$	0.195	0.17	0.23	$1.47x10^{-37}$	0.166	0.14	0.20	$9.38x10^{-28}$	171.73	$4.44 x10^{-36}$
7q31.31	rs2908004	*WNT16*	A	0.100	0.07	0.13	$3.01x10^{-11}$	0.177	0.15	0.21	$1.41x10^{-32}$	0.088	0.06	0.12	$3.59x10^{-9}$	69.96	$2.31 x10^{-14}$
8q24.12	rs2450083	*COLEC10*	T	0.015	-0.02	0.05	$3.38x10^{-1}$	0.037	0.01	0.07	$1.42x10^{-2}$	0.102	0.07	0.13	$2.13x10^{-11}$	26.62	$2.37x10^{-5}$
9q34.11	rs7466269	*FUBP3*	A	0.087	0.06	0.12	$1.51x10^{-8}$	0.077	0.05	0.11	$3.67x10^{-7}$	0.052	0.02	0.08	$6.83x10^{-4}$	2.40	$6.63x10^{-1}$
11p14.1	rs10835187	*LIN7C*	C	0.045	0.02	0.07	$3.10x10^{-3}$	0.041	0.01	0.07	$5.53x10^{-3}$	0.127	0.10	0.16	$1.63x10^{-17}$	38.35	$9.47x10^{-8}$
11q13.2	rs12272917	*PPP6R3*	T	0.065	0.03	0.10	$1.38x10^{-4}$	0.067	0.03	0.10	$7.78x10^{-5}$	0.109	0.08	0.14	$1.34x10^{-10}$	13.50	$9.06x10^{-3}$
12p11.22	rs4420311	*KLHDC5*	G	0.086	0.06	0.12	$3.21x10^{-8}$	0.066	0.04	0.10	$2.25x10^{-5}$	0.037	0.01	0.07	$1.58x10^{-2}$	9.78	$4.43x10^{-2}$
13q14.11	rs9525638	*TNFSF11*	C	0.064	0.04	0.09	$2.09x10^{-5}$	0.089	0.06	0.12	$2.47x10^{-9}$	0.057	0.03	0.09	$1.43x10^{-4}$	9.62	$4.73x10^{-2}$
14q32.12	rs754388	*RIN3*	C	0.130	0.09	0.17	$1.40x10^{-10}$	0.103	0.06	0.14	$3.13x10^{-7}$	0.035	0.00	0.08	$7.82x10^{-2}$	11.10	$2.55x10^{-2}$
18q21.33	rs884205	*TNFRSF11A*	C	0.003	-0.03	0.04	$8.58x10^{-1}$	0.023	-0.01	0.06	$2.11x10^{-1}$	0.104	0.07	0.14	$1.84x10^{-8}$	25.87	$3.36x10^{-5}$

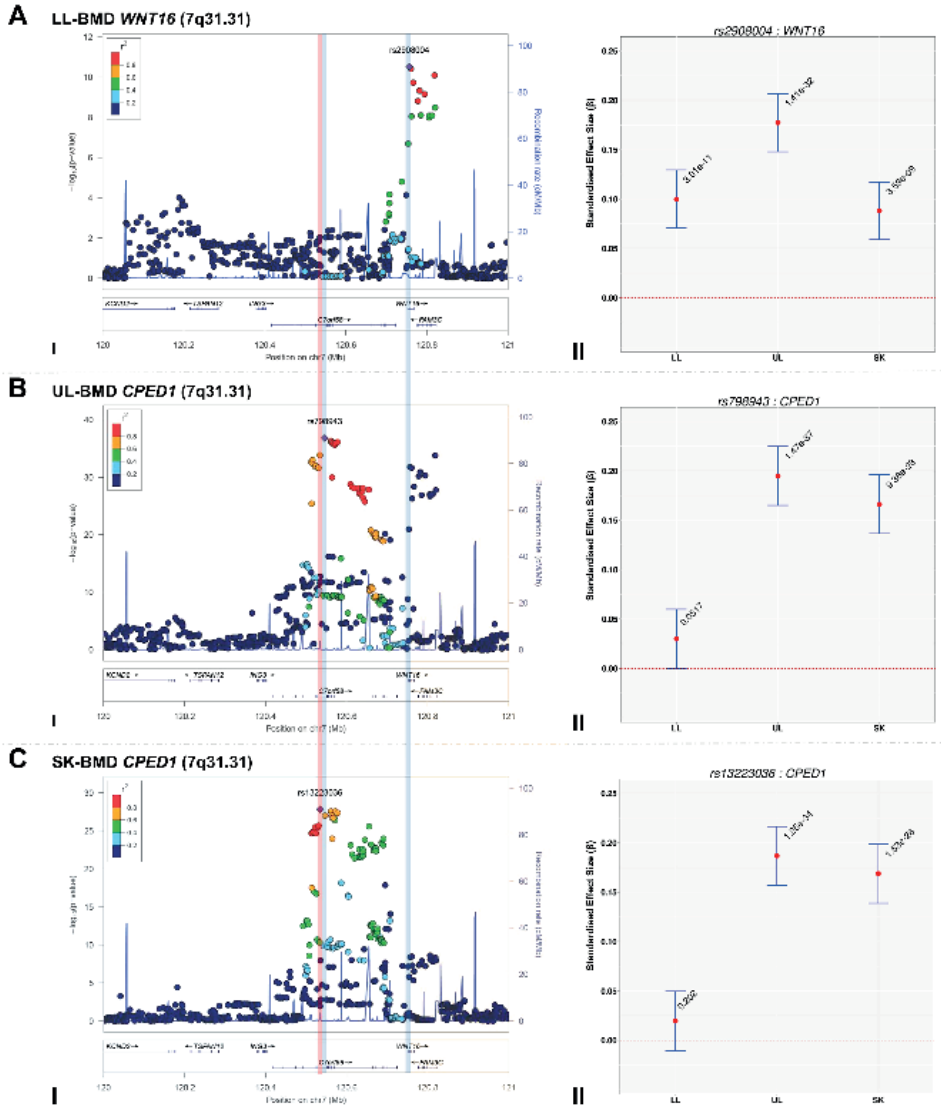


Figure 3: Regional association plots of the top SNPs associated with total-body less head-, lower limb-, upper limb- and skull-BMD at the 7q31.31 locus, in addition to a comparison of the effect size of the top site-specific SNP on BMD at the three different skeletal sites. For I: Circles show GWA meta-analysis P-values and positions of SNPs found within the 7q31.31 locus. Top SNPs are denoted by diamonds. Different colors indicate varying degrees of pair-wise linkage disequilibrium (HapMap 2 CEPH) between the top SNP and all other SNPs. Blue vertical shaded areas indicate the position of rs7776725 (top SNP A-I) and rs2908004 (top SNP B-I) and rs798943 (top SNP C-I) for each analysis. The red vertical shaded area represents the position of rs13223036 (top SNP D-I). For II: The per allele effect in SD (red dot) and the 95% confidence interval (error bar) of the top SNP for lower limb (LL), upper limb (UL) and skull (SK) are plotted with their specific strength of evidence against the null hypothesis of no association. Please note: *FAM3C* and *CPED1* are also located at the 7q.31.31 locus containing *WNT16*.

To formally determine whether the standardized regression coefficients of each of the above-mentioned variants truly differed across the skeletal sites, we fitted a multivariate normal likelihood model to the raw data in ALSPAC and Generation R (see Methods), and then meta-analysed the results using Fisher's method. Using a conservative threshold (i.e., $\alpha = 5\times10^{-8}$), we observed robust evidence indicating that i.e., rs13223036 and rs798943, located at *CPED1* exerted strong effects on UL and SK-BMD, when compared to LL-BMD [$P = 2.01\times10^{-37}$ and $P = 4.44\times10^{-36}$ (Table 3)], whereas the variant rs2908004 (*WNT16*) was strongly related to UL-BMD in comparison to BMD at the other sites ($P = 2.31\times10^{-14}$). Several variants at other loci were also suggestive of some degree of skeletal site specificity including *EYA4* and *LIN7C*, although they did not formally meet the criteria for statistical significance (Table 3, Online Resource Figures S8 – S11).

Association of novel variants with hip and spine BMD.

To elucidate if any of the novel primary and/or secondary signals, identified during the course of this study, were nominally associated with BMD in adults, we performed a lookup of these variants in the publicly available results of the GEFOS-II meta-analysis of hip and spine BMD (Online Resource Table S11) [4]. The novel *RIN3* variant (rs754388) was not associated with femoral neck ($P_{FN} = 0.87$) and lumbar spine BMD ($P_{LS} = 0.42$). The G allele of the *EYA4* variant (rs3012465), associated with increased SK-BMD ($\beta = 0.13$, SE = 0.02, $P = 8.3\times10^{-17}$), but suprisingly showed nominal association with decreased BMD at both the hip ($P = 7.1\times10^{-3}$) and spine ($P = 0.04$). A followup of this variant in a recent published GWAS of 4061 premenopausal women aged 20 to 45 revealed no evidence of association with FN-BMD ($P = 0.73$) [15].

A lookup of the secondary independent SNPs revealed no evidence of a relationship between the TBLH- and LL-BMD-associated *KLHDC5/PTHLH* variant (rs4420311) and associations with hip or spine BMD ($P_{FN} = 0.33$ and $P_{LS} = 0.45$) in GEFOS-II. Similarly no evidence of association was detected for the SK-BMD-associated variant at *CENPW/RSPO3* (rs2130604: $P_{FN} = 0.98$ and $P_{LS} = 0.40$). Interestingly, the T allele of the *CENPW/RSPO3* variant (rs4418209), which was associated with increased SK-BMD ($\beta = 0.07$, SE = 0.02, $P = 1.1\times10^{-6}$), appeared to be nominally associated with decreased hip BMD ($P = 5.0\times10^{-3}$) but not spine BMD ($P = 0.34$). Further inspection revealed that the T allele of rs4418209 was nominally associated with decreased BMD at the TBLH ($\beta = -0.03$, SE = 0.02, $P = 1.7\times10^{-2}$), LL ($\beta = -0.04$, SE = 0.02, $P = 6.5\times10^{-3}$) and UL ($\beta = -0.04$, SE = 0.02, $P = 8.2\times10^{-3}$). The T allele of rs17536328 located within *TNFSF11,* associated with increased TBLH-BMD, showed nominal evidence of association with increased hip ($P = 0.04$) but not spine BMD ($P = 0.87$). In contrast, the G allele of an independent *TNFSF11* variant (rs2148072) associated with increased UL-BMD was associated with decreased spine BMD ($P = 0.05$). In addition, the C allele *LIN7C/LGR4*

variant (rs10160456) associated with increased SK-BMD showed weak evidence of association with increased hip ($P = 0.06$) and spine ($P = 0.03$) BMD.

Bioinformatic analysis of *RIN3*

We fine mapped the *RIN3* region by imputing common and rare variants using a reference panel from the 1000 Genomes Project and identified a missense variant (rs117068593) that was in strong linkage disequilibrium ($r^2 = 0.96$) with the top LL-and TBLH-BMD associated *RIN3* variant (rs754388). The C allele of rs117068593 (EAF = 0.82) was associated with increased BMD of the lower limbs ($\beta = 0.13$, SE = 0.020, $P = 5.97\times10^{-11}$) and total-body less head ($\beta = 0.12$, SE = 0.020, $P = 1.87\times10^{-9}$). A search of the SIFT database [16] revealed that the missense variant could negatively affect RIN3 functioning. This prediction was further supported by a search of the Regulome database [17], which suggested that the missense variant alters the binding of the following transcription factors: EBF1, EGR1, SP1, NFKB1 and POLR2A, in lymphoblastic cell lines.

***RIN3* expression profiling**

Evaluation of cis-expression quantitative trait loci (eQTLs) from primary human osteoblasts using array-based gene expression suggested that variants located within 1MB of *RIN3* (i.e., including variants tagging *SLC24A4, LGMN, GOLGA5, CHGA* and *ITPK1*) were nominally associated with *ITPK1* expression ($P = 0.04$). This observation failed to meet the level of significance after correction for multiple testing. Examination of the temporal pattern of gene expression across osteoblastogenesis, using mouse calvarial derived cells, starting with the pre-osteoblast stage, through to mature osteoblasts revealed that *Rin3, Golga5* and *Lgmn,* and *Iptk1* were expressed in this cell type (Online Resource Figure S12). In contrast, *Slc24a4* and *Chga* were not expressed at all in the pre- or mature osteoblast, as determined by RNAseq. A further investigation of the expression profiles of the aforementioned genes in human mesenchymal stem cells [(hMSCs), differentiated into adipocytes and osteoblasts] and peripheral blood monocytes [(PBMCs) differentiated into osteoclasts] indicated the following: *SLC24A4* was not expressed in any of these cell lines when differentiated, *GOLGA5* had an intermediate expression level in both differentiating hMSCs and PBMCs and *LGMN* was immediately upregulated upon differentiation into adipocytes (8 fold), osteoblasts (5 fold) and osteoclasts [(5 fold), Online Resource Figure S13 and S14]. Moreover, we found that the expression of *RIN3* was 2-fold downregulated during the proliferative phase of differentiating PBMCs into osteoclasts (Online Resource Figure S13). Finally a comparison of expression profiles across the *RIN3* region of illiac bone biopsies derived from 39 osteoporotic and 27 healthy postmenopausal donors revealed one transcript (i.e., 220439_at, originating from *RIN3*), that demonstrated reduced expression in the osteoporotic group relative to the control group [$P = 2.7\times10^{-3}$, (Online Resource Table S13)].

DISCUSSION

This study assessed whether regional analysis of skeletal sites from TB-DXA could be used to estimate the extent to which genetic and environmental factors influence bone mass accrual of differentially loaded skeletal sites (skull, lower limbs, and upper limbs). Point estimates indicated that common SNPs on a commercially available genotyping array, explained a larger proportion of the overall variance of SK-BMD, when compared to BMD measured at the appendicular sites (i.e., lower and upper limbs). These differences potentially reflect differential exposure of each skeletal site to varying environmental stimuli that influence BMD. Specifically the skull, as opposed to appendicular sites, is less influenced by environmental factors, particularly those acting through mechanical loading. To explore this result further, we estimated the residual correlation (i.e., the proportion of environmental and other sources of variation not tagged by SNPs on the Illumina platform) across the different skeletal sites and found that whilst the environmental (and other residual) factors influencing the appendicular sites were moderately similar to each other, they appeared to be appreciably different from the factors influencing SK-BMD. Taken together, lower v_g estimates, coupled with a high residual correlation between the two appendicular sites, may reflect the greater exposure of these sites to loading and muscular stimulation, when compared to the skull.

Likewise, estimates of the genetic correlations indicated that the appendicular limbs shared a more similar genetic architecture when compared to the skull, possibly reflecting the composition of bone at each skeletal site and the biological processes that govern their growth and maintenance. For example, appendicular sites consist of broadly equivalent proportions of cortical and trabecular bone. The skull on the other hand is mainly comprised of flat bones, which consist primarily of cortical bone [18]. The developmental processes also differ between long and flat bones, with dermal bones such as the skull vault arising exclusively through intramembranous bone formation, in contrast to long bones, which form through endochondral bone formation involving intermediary formation of cartilage [19].

To further explore the basis for the above-mentioned differences in underlying genetic architecture, GWA meta-analyses of sub-regional TB-DXA data were performed. These analyses helped identifying genetic signals that were associated with one or more skeletal region(s). When comparing the evidence of association for all SNPs (identified in this effort) across each skeletal site, our GWA meta-analyses echoed the findings of our GCTA results, supporting the notion that although the underlying genetic architecture influencing BMD appears to be largely similar, it does vary according to skeletal site. The majority of the top SNPs were nominally associated ($P \leq 0.05$) with BMD across all skeletal sites (i.e.,

SNPs at *WNT4*, *WNT16, FAM3C, GALNT3*, *FUBP3*, *KLHDC5/PTHLH*, *TNSF11*, *LIN7C/LGR4* and *PPP6R3/LRP5*). In contrast, variants near or within *CPED1*, *COLEC10/TNFRSF11B* and *EYA4* were strongly associated with UL- and SK-BMD, but not LL-BMD. A further variant was identified within *TNFRSF11A* that appeared to be solely related to SK-BMD. Most notably we observed a novel association between rs754388 (located within *RIN3*) and LL-/UL-BMD, but not SK-BMD. To the best of our knowledge this is the first GWAS to report an association between *RIN3* and BMD. It seems likely that this association reflects a true relationship with BMD as the same *RIN3* signal (as determined by conditional analysis) has previously been associated with an increased risk of Paget's Disease [i.e., rs10498635-C OR: 1.44, 95%-CI (1.29-1.60) $P = 3.x10^{-11}$] [20].

In an attempt to further understand how the genetic variation surrounding *RIN3* may influence BMD, we fine mapped *RIN3* and identified a missense variant (rs117068593) that was in high LD with our LL-BMD associated SNP. Data mining of SIFT and ENCODE databases suggested a functional role of the missense variant that putatively affects binding of several transcription factors in lymphoblastic cell lines. We further evaluated expression quantitative trait locus (eQTL) data from primary human osteoblasts using SNP data from HapMap (i.e., not including rs117068593) and found no substantial evidence that our LL-BMD associated SNPs located at 14q32.12 regulated the expression of *RIN3* or any of the genes located nearby. However, differential patterns gene expression were detected when comparing *RIN3* expression profiles of osteoporotic and healthy individuals. Further, we also observed differential expression during osteoclast differentiation that was not present in osteoblast and adipocyte differentiation processes. Collectively, the aforementioned observations appear to be in line with previous findings that suggest that RIN3 could influence osteoclast activity, especially when considering the prior association of *RIN3* with Paget's Disease, a disease driven by osteoclast dysfunction and molecular studies that indicate that RIN3 is involved in vesicular trafficking, a process critical for bone resorption [20,21]. Further study is however needed to elucidate the precise role of role of RIN3 in bone metabolism.

To further understand the preferential associations of some variants with different skeletal sites, we compared the standardized effect sizes of all the genome-wide significant BMD-associated variants, across each skeletal site using a formal multivariate normal likelihood model. Variants at the *CPED1* locus were strongly associated with BMD at the skull and upper limb sites, but not with LL-BMD. Similarly variants at *WNT16* were more strongly related to UL-BMD, than to BMD at the other sites. Several other SNPs showed evidence for site specificity including variants at the *EYA4* and *LIN7C* loci that were very strongly related to SK-BMD, although these variants did not surpass our conservative criterion for declaring significant heterogeneity, corroboration is needed from independent studies.

Conceivably, differences in the pattern of results across SNPs may have arisen from an artefact of the measurement (i.e., where sub-regional-specific associations reflect how accurately BMD is measured at each skeletal site). However, if the latter were the case, one would expect to observe a consistent pattern of results across all loci (i.e., the strength of association should be greatest at those sites measured more accurately). From our results, this is clearly not the case as evidence of association is sometimes greatest for the skull, whilst for other SNPs evidence is greatest for lower and/or upper limbs. In terms of biological explanations, larger effect sizes of genetic variants that influence SK-BMD possibly reflect their preferential involvement in cortical as opposed to trabecular bone metabolism and/or the involvement of intramembranous ossification vs. endochondral ossification [13]. Certain genetic factors also appeared to influence UL-BMD more strongly than LL-BMD, or vice versa. Since the composition and developmental origin of these two sites is broadly similar, presumably, other explanations are responsible. It is reasonable to think that genetic factors, which we identified, could be acting to alter responses to stimuli that are themselves site-specific. For example, adipose tissue has previously been reported to influence cortical bone of the tibia in preference to the radius [22].

Quantitative SK-BMD measurements have traditionally been ignored by genetic and epidemiological studies as they are thought to be prone to errors such as dental augmentation. Despite these concerns, a study conducted in premenopausal woman found a high correlation between the upper half of the skull (i.e., cranial vault) and total skull-BMD ($r^2 = 0.991$, n = 91, Age range 19-30 years), with a mean difference of -0.004 g/cm2, suggesting that these two measurements of bone mass are similar [23]. We found that paediatric SK-BMD measures are well suited to GWAS, as indicated by the very low *P*-values obtained at some of the known BMD associated loci (10^{-17} to 10^{-28}) despite our relatively small sample size. This observation may reflect the fact that SK-BMD is considerably less subject to environmental influences, such as those acting through mechanical loading. In addition, genetic variants associated with SK-BMD identified in this study may primarily reflect molecular pathways involved in bone mass accrual and growth, in contrast to variants identified from previous adult scans which may be more strongly related to mechanisms involved in bone maintenance and/or loss.

Almost all the loci we have identified in this study (i.e., with the exception of SNPs in *RIN3* and *EYA4*) have been associated with BMD at either the hip or the lumbar spine previously. Variants mapping to *RIN3* have been implicated in Paget's disease but this is the first time the locus is associated with BMD, and interestingly, the alleles associated with increased BMD are associated with increased risk for the condition. This shows that performing GWAS of BMD at sites other than at the femoral neck (FN-BMD) or lumbar spine (LS-BMD) can be used to identify loci that exert pleiotropic effects on bone. Potential advantages of examin-

ing these additional sites from a locus discovery perspective are that (i) genetic variants may exert stronger effects at these sites than at FN-BMD/LS-BMD, and/or (ii) the genetic effects may be more apparent at these sites because the effect of environmental noise is minimized. For example, the *P*-values for skull BMD at several loci (e.g., variants around *CPED1*, *EYA4* and *LIN7C*) are many orders of magnitude stronger than the corresponding *P*-values for TBLH-BMD (see Online Resource Table S8). Likewise variants in *LIN7C* were first discovered using a GWAS meta-analysis of lumbar spine that was over five times the size of the present study, and even then only just exceeded the threshold for genome-wide significance [4], whereas in our study a variant at this locus has $P < 1 \times 10^{-16}$ with SK-BMD. Hence, GWAS of BMD at sites such as the skull could be used to efficiently detect clinically relevant loci that might be more difficult to discover in GWAS of the femoral neck and/or lumbar spine.

To further illustrate the value of SK-BMD, we draw attention to rs3012465, a variant proximal to the eyes absent (*EYA4*) gene and associated with increased SK-BMD. We show that the signal is analogous to that previously associated with increased volumetric cortical BMD of the tibia (i.e., C allele of rs271170: $\beta = 0.11$, $P = 2.7 \times 10^{-12}$), based on a GWAS in ALSPAC and other young adult cohorts [13], suggesting that both findings reflect the relationship of the *EYA4* locus with cortical bone. However, a look-up in a separate cortical bone site (i.e., the femoral neck of the hip), from a GWAS in older adults, revealed that the BMD-increasing allele at the *EYA4* locus was in fact associated with lower BMD for both rs3012465 and rs271170 [4]. Taken together, these findings may reflect an age dependent effect of *EYA4* whereby *EYA4* contributes to bone accrual in early life, yet maybe influences bone loss in older adults. To test this hypothesis, we followed up these *EYA4* variants in a recent GWAS meta-analysis of FN-BMD in 4061 pre-menopausal women aged 20 – 45 (as described in Koller and colleagues [15]) and failed to find any evidence of association with FN-BMD ($P = 0.73$). These results suggest that the discrepancy in results between GEFOS and the present study is unlikely to be solely due to age, but rather is likely to represent a real difference between skeletal sites.

In summary, our strategy of analysing regional paediatric DXA measures of TB-BMD represents a novel approach to dissecting the genetic architecture influencing bone mass accrual and growth at different skeletal sites. Specifically, variants at 13 loci reached genome-wide significance with BMD and several displayed different degrees of association according to skeletal site. Furthermore, we report a novel association between a variant within *RIN3* and LL-BMD and note its previous association with risk of Paget's disease. We additionally provide suggestive evidence of allelic heterogeneity at the *CENPW/RSPO3*, *KLHDC5/PTHLH* and *LIN7C/LGR4* loci. In conclusion our results provide evidence that different skeletal sites as measured by TB-DXA are to a certain extent under distinct environmental and genetic

influences. Allowing for these differences may help to uncover new genetic influences on BMD, particularly those examined in children as involved in bone growth and accrual.

MATERIALS AND METHODS

Subjects

ALSPAC

ALSPAC is a longitudinal population-based birth cohort that recruited pregnant women residing in the former county of Avon, UK, with an expected delivery date between 1st April 1991 and 31st December 1992. This cohort has been described in detail on the website (http://www.alspac.bris.ac.uk) and elsewhere [24]. DXA, height and weight measurements were performed on children who attended the 9 year old focus group clinic [mean age of participant 9 (± 0.32 years)]. Ethical approval was obtained from the ALSPAC Law and Ethics committee and relevant local ethics committees, and all parents provided written informed consent.

Generation R Study

The Generation R Study is a prospective cohort study enrolling 9,778 pregnant women living in Rotterdam with a delivery date from April 2002 until January 2006. Details of study design and data collection have been described elsewhere [25]. DXA, height and weight measurements were performed on children who visited the research centre whilst being accompanied by their mothers at a mean age of 6 (± 0.5 years). All research aims and specific measurements taken during the course of the Generation R Study have been approved by the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam. All parents provided written informed consent.

Phenotypes

ALSPAC

TB-DXA scans were performed on all participants, using a Lunar Prodigy scanner (Lunar Radiation Corp, Madison, WI) with paediatric scanning software (GE Healthcare Bio-Sciences Corp., Piscataway, NJ). DXA measures of BMD were derived for the following regions of interest: TBLH-, UL-, LL- and S. All DXA scans were subsequently reviewed by a trained researcher, and re-analysed as necessary, to ensure that borders between adjacent ROI's were placed correctly by the automated software. The coefficient of variation for TBLH-BMD measures was 0.8%, based on the analysis of 122 children who had two scans performed on the same day. Height was measured to the nearest 0.1 cm using a Harpenden

stadiometer (Holtain Ltd., Crymych, UK) and weight was measured to the nearest 50 g using Tanita weighing scales (Tanita UK Ltd, Uxbridge).

Generation R Study

TB-BMD was measured in all participants using a GE-Lunar iDXA scanner. Well-trained research assistants obtained the DXA scans using the same device and software (enCORE) following standard manufacturer protocols. The same regions of interest as described for ALSPAC were derived from TB-DXA scan. To ensure that the lines between adjacent ROI's were placed correctly by the automated software, scans were evaluated twice, directly after the scanning and at a later time point by a second well-trained research assistant. The coefficient of variation for total TBLH-BMD measures was 0.23%, based on duplicate scans of children that were performed on the same day.

Genotyping and Imputation

ALSPAC

A total of 9,912 subjects were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina Inc., San Diego, CA, USA) by 23andMe subcontracting the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America (LabCorp Holdings., Burlington, NC, USA). PLINK software (v1.07) was used to carry out quality control measures [26]. Individuals were excluded from further analysis on the basis of having incorrect gender assignments, minimal or excessive heterozygosity (< 0.320 and > 0.345 for the Sanger data and < 0.310 and > 0.330 for the LabCorp data), disproportionate levels of individual missingness (> 3%), evidence of cryptic relatedness (> 10% IBD) and being of non-European ancestry (as detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). EIGENSTRAT analysis revealed no additional obvious population stratification and genome-wide analyses with other phenotypes indicate a low lambda) [27]. SNPs with a minor allele frequency of < 1% and call rate of < 95% were removed. Furthermore, only SNPs that passed an exact test of Hardy–Weinberg equilibrium ($P > 5\times10^{-7}$) were considered for analysis. After quality control, 8,365 unrelated individuals who were genotyped at 500,527 SNPs were available for analysis. Known autosomal variants were imputed with Markov Chain Haplotyping software (MACH 1.0.16) [28,29], using CEPH individuals from phase II of the HapMap project (hg18) as a reference set (release 22) [30]. The BMD associated *RIN3* locus was further imputed using the complete reference panel from the third phase of the 1000 Genomes Project (i.e., March 2012) [31].

Generation R Study

Genotyping was performed using the Illumina HumanHap 610 QUAD microarray using standard manufacturer protocols. Stringent quality control of the genotype and imputa-

tion process was performed in this study as previously described [7]. Samples with gender discrepancy, excess of heterozygosity, low genotype quality and sample replicates were excluded from the analysis. A reference panel for imputation, consisting of CEPH, YRI and CHB/JPT haplotypes was constructed using data from phase 2 of the HapMap project (hg18, release 22) [30]. A two-step imputation process was performed using MACH for haplotype phasing and Minimac for imputation [28,29]. A similar 1000 Genomes imputation strategy (as described above for ALSPAC) was used to fine map the *RIN3* locus.

Statistical Methods

Choice of covariates

BMD as measured by DXA is strongly influenced by weight, in part because weight is related to skeletal size. BMD as assessed by DXA, does not correct for the thickness (depth) of bone, therefore true (volumetric) bone mineral density is often underestimated in smaller individuals and overestimated in larger subjects. Weight is also thought to affect BMD by other pathways such as increased skeletal loading, and possibly by other metabolic influences. We reasoned that SK-BMD is likely to be relatively unaffected by these other pathways, and so whereas TBLH-, UL- and LL-BMD measures were adjusted for weight, SK-BMD was adjusted for height.

Genome-wide complex trait analysis

Univariate restricted maximum likelihood (REML) genome-wide complex trait analyses (GCTA) [32] were performed on ≥ 4,866 ALSPAC subjects to estimate the proportion of additive genetic variance in BMD at each site, explained by directly genotyped variants that had a minor allele frequency ≥ 1%. Bivariate REML GCTA analysis [33] was further used to estimate the pair-wise genetic and residual correlations between BMD at each skeletal site. A cryptic relatedness cut-off of 0.025 was applied in order to ensure that distantly related individuals (i.e., n ~ 444) were removed prior to the analysis, thereby reducing the potential for bias (Online Resource Figure S15). GCTA analysis was not performed in the Generation R Study given its multi-ethnic composition. Pearson Product Moment Correlation was used to estimate the linear relationship between standardised residuals of BMD after adjusting for age, gender, and weight or height using the STATA statistical package [34].

Genome-wide association meta-analysis of BMD in ALSPAC and Generation R

To identify genetic loci influencing variation in TBLH-, LL-, UL- and SK-BMD, we performed GWAS meta-analyses combining 5,330 children (5,299 for SK-BMD) from the ALSPAC cohort and 4,086 children from the Generation R Study, who had DXA BMD measurements and imputed GWAS data. Cohort specific GWAS analyses were conducted in ALSPAC and Generation R using standardised residuals derived from BMD measures after adjustment

for age, gender and weight for all skeletal sites except the skull, where weight was substituted for height. The first 20 ancestry informative principal components were additionally incorporated into the Generation R model to control for population stratification, due to the multi-ethnic nature of this cohort as described previously [7]. Genome-wide association analyses were performed using MACH2QTL [29] as implemented in GRIMP [35], using linear regression models based on an expected allelic dosage for SNPs, adjusting for the above mentioned covariates where necessary. We combined association data for ~2.5 million imputed autosomal SNPs into an inverse variance fixed-effects meta-analysis, using METAL and controlled for genomic inflation in each cohort [36]. *P*-values less than 5×10^{-8} were considered genome-wide significant. Heterogeneity was evaluated using Cochran's Q statistic and was quantified by the I^2 metric. Regional association plots from our genome-wide association scans were generated using Locuszoom (v1.1) [37], using linkage-disequilibrium (LD) information estimated from the HapMap 2 (hg18) CEPH reference dataset [38]. All pair-wise LD estimates were obtained using SNAP software in conjunction with the HapMap2Phase II (hg18) CEPH reference dataset [39]. All remaining plots were generated in *R* [40] using the ggplot2 software package [41].

Sensitivity Analysis

In order to test the robustness of our results to the choice of covariates at each site, we performed a sensitivity analysis adjusting each region-specific BMD measure for: age, gender, height and weight (i.e., Model 1a) and performing GCTA and GWAS meta-analysis using the residuals. In order to confirm that our results were not being driven by underlying population substructure, we performed further GWAS meta-analyses using the same residuals derived for Model 1a, except that the analyses were restricted to individuals of European ancestry (i.e., Model 1b).

Conditional meta-analysis

To identify novel secondary BMD association signals (i.e., independent from those published), at loci which reached genome-wide significance for each BMD meta-analysis, we carried out conditional meta-analyses by conditioning on all previously published BMD-associated variants that mapped to between 1 - 2 Mb of the top locus specific SNP [depending on the extend to LD, (Online Resource Table S12)]. In the case of rs7851693 (*FUBP3*), not present in the Generation R dataset, conditional analysis was performed using a proxy SNP (i.e., rs7030440), which was in high LD ($r^2 > 0.96$) with the missing BMD associated variant. For *RIN3*, there were no BMD associated loci previously published, thus for that locus we conditioned on the top SNP. After conditional analysis, locus specific significance correction thresholds (for multiple testing of SNPs which are in linkage disequilibrium with each other) were calculated using the Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) software package [42]. Locus specific regions used for SNPSpD were defined as

the region proximal (1 – 2 Mb) to each locus specific signal. The presence of independent secondary association signals was confirmed in situations where the residual signal (after conditional meta-analysis) reached the locus specific threshold corrected for multiple testing.

Comparison of the magnitude of the effect sizes of genome-wide significant SNPs across skeletal-sites

As we were interested in whether the genetic variants exerted greater influence on BMD at a particular site, it would be misleading to directly compare standardized regression coefficients from the meta-analysis by e.g., simple Z test, since the summary statistics were derived from correlated measures on the same individuals and are therefore not independent. In addition, because we are only testing variants that have already met the criterion for genome-wide significance (and were therefore selected on the basis of having an extreme regression coefficient at a particular site), it is not appropriate to compare the regression coefficients from different sites using an uncorrected type I error level of $\alpha = 0.05$. To address both these considerations, we fitted a multivariate normal model to the standardized BMD scores at each site by maximum likelihood using the software package Mx [43]. We fitted a model where the standardized regression coefficients for each site (SK-BMD, LL-BMD, UL-BMD) were constrained to be equal, and then another model in which the regression coefficient most different from the other two was allowed to vary). Twice the difference in log-likelihood between these models is distributed as a chi-square statistic with one degree of freedom. We analysed each cohort separately using this method and then combined the results using Fisher's product of *P*-values evaluating statistical significance against a conservative threshold of $\alpha = 5\times10^{-8}$ (i.e., as if we had performed the comparison genome-wide, not just post-hoc on the significant sites).

Functional analysis of *RIN3*

In an attempt to identify a potential functional or regulatory mechanism underlying the association between *RIN3* and BMD, a range of bio-informatic and functional analyses were performed. These included: fine mapping the *RIN3* locus, data mining Regulome [17] and SIFT [16] databases and performing eQTL analysis on primary human osteoblasts. The expression profiles of *RIN3* and neighboring genes: *SLC2484, LGMN, GOLGA5, CHGA* and *ITPK1* were also investigated in bone biopsies of healthy and osteoporotic women, in addition to murine and human cell lines that were differentiated into osteoblasts and/or osteoclasts. Methods specific to each analysis are described below.

Human primary osteoblasts

Expression profiling of untreated primary human osteoblasts, obtained from 113 (51 female and 62 male) unrelated Swedish donors, was performed using the Illumina HumRef-8

BeadChips in accordance with the manufacturer's instructions. Up to 3 biological replicates were analysed per sample. Genotyping for genotype-expression association was performed using the Illumina HapMap 550k Duo chip. Individuals with low genotyping rate and SNPs showing significant deviation from Hardy-Weinberg equilibrium ($P < 0.05$) were excluded. Similarly, low frequency (MAF < 0.05) SNPs and SNPs with high rates of missing data were excluded. Genotypes from samples that passed quality control (n = 103) were imputed for all SNPs (n = 478,805) oriented to the positive strand from phased autosomal chromosomes of HapMap Phase 2 CEPH panel (release 22, build 36) using MACH 1.0. A RSQR cut-off of < 0.3 was used to remove poorly imputed markers. Association of imputed genotypes using estimated genotype probabilities with nearby expression traits (defined as ±1Mb window flanking *RIN3*) was performed using a linear regression model implemented in the MACH2QTL software with sex and age as covariates. Detailed methods pertaining to the data generation and analysis are described elsewhere [44,45].

Human illiac bone biopsies

Gene expression profiles were generated from iliac bone biopsies donated by healthy control (n =27), osteopenic (n = 18) and osteoporotic (n = 39) postmenopausal Norwegian women. The affection status of each individual was determined by BMD measurements of the total hip or lumbar spine (L1–L4 verterbrae). Individuals with a T-score less than -2.5 and with at least one low trauma fracture were deemed osteoporotic, whilst individuals with a T-score > -1 were deemed healthy. Expression profiling was performed using an Affymetrix HG U133 2.0 plus array. The Affymetrix Cel files were imported into Partek Genomics Suite (Partek Inc., St Louis, MO, USA), and normalized using the RMA (Robust Multichip Average) algorithm. Gene expression patterns were further adjusted, as reported by Jemtland and colleages [46], for batch effects and differing synthesis times. The gene expression profiles of all transcripts located ±250 kb of the top LL-BMD associated *RIN3* SNP (i.e., rs754388) were compared between the osteoperotic and control group. Note: the intermediate osteopenic group was excluded from this analysis.

Murine pre-osteoblasts

All procedures and use of mice for the neonatal osteoblast expression studies were approved by the Jackson Laboratory Animal Care and Use Committee (ACUC), in accordance with NIH guidelines for the care and use of laboratory animals. Pre-osteoblast-like cells were isolated from neonatal calvaria from C57BL/6J mice expressing cyan florescent protein (CFP) under the control of the Col3.6 promoter (pOBCol3.6CFP), using standard techniques [47]. Cells were cultured for 4 days in growth media [DMEM containing 10% fetal bovine serum (FBS) and 1X penicillin/streptomycin], and thereafter removed from culture and subjected to fluorescence-activated cell sorting (FACS) based on the presence/absence of CFP expression. Cells expressing CFP, and therefore considered pre-osteoblasts,

were plated at a density of 1×10^4 cells per cm^2, differentiated into osteoblasts using standard methods (αMEM containing 50 μg/ml Ascorbic Acid, 4 mM β-glycerol phosphate, 10% FBS and 1X penicillin/streptomycin). RNA was collected at 9 time points post differentiation. mRNA profiles for triplicate samples for each time point were generated by Next Generation High throughput RNA sequencing (RNAseq), using an Illumina HiSeq 2000. The alignments for abundance estimation of transcripts was conducted using Bowtie version 0.12.9 [48] using the NCBIm37 transcriptome as the reference genome. Expression level per gene was calculated using RSEM version 1.2.0 using the following parameters: --fragment-length-mean 280 and --fragment-length-sd 50 [49] and expression level for each sample was normalized relative to the per sample upper quartile [50]. This data has been submitted to the gene expression omnibus (Accession Number: GSE54461).

Human mesenchymal stem cells and peripheral blood mononuclear cells

Human bone marrow derived mesenchymal stem cells [(hMSC), Lonza Group Ltd., Basel, Switzerland] were seeded in 12-well plates (5×10^3 cells per cm^2) and differentiated into osteoblasts (using α-Mem pH7.5, 10% heat inactivated foetal calf serum (FCS), 100 nM Dexamethasone and 10 mM β-glycerophosphate) or adipocytes (using α-MEM pH 7.5, 10% heat inactivated FCS, 100 nM dexamethasone, 500 μM IBMX and 60 μM Indomethacin). Total RNA was isolated (triplicates) using Trizol (Life Technologies, Carlsbad, CA, USA) twice a week during differentiation until the 25 day of culture. Human peripheral blood mononuclear cells (PBMCs) were retrieved from buffy coats using Ficoll and seeded in 12-well plates (1.3×10^5 cells per cm^2). Monocytes were allowed to attach for 4 hours and non-adherent cells were removed by careful washing. In the next three days, cells were grown in α-Mem pH7.5, containing 15% heat-inactivated serum and 25 ng/ml macrophage colony stimulating factor [(M-CSF), R&D Systems Inc., Minneapolis, MN, USA] to stimulate proliferation of the monocytes. After 3 days, media was replaced with α-Mem pH7.5 containing 15% heat inactivated serum, 25 ng/ml M-CSF and 30 ng/ml RANKL (Peprotech Inc., Rocky Hill, NJ, USA) to initiate osteoclastogenesis. Total-RNA was isolated twice a week using Trizol until the 21st day of culture. Amplification of total-RNA, Illumina microarray hybridization, data extraction and normalization were performed as previously described [51].

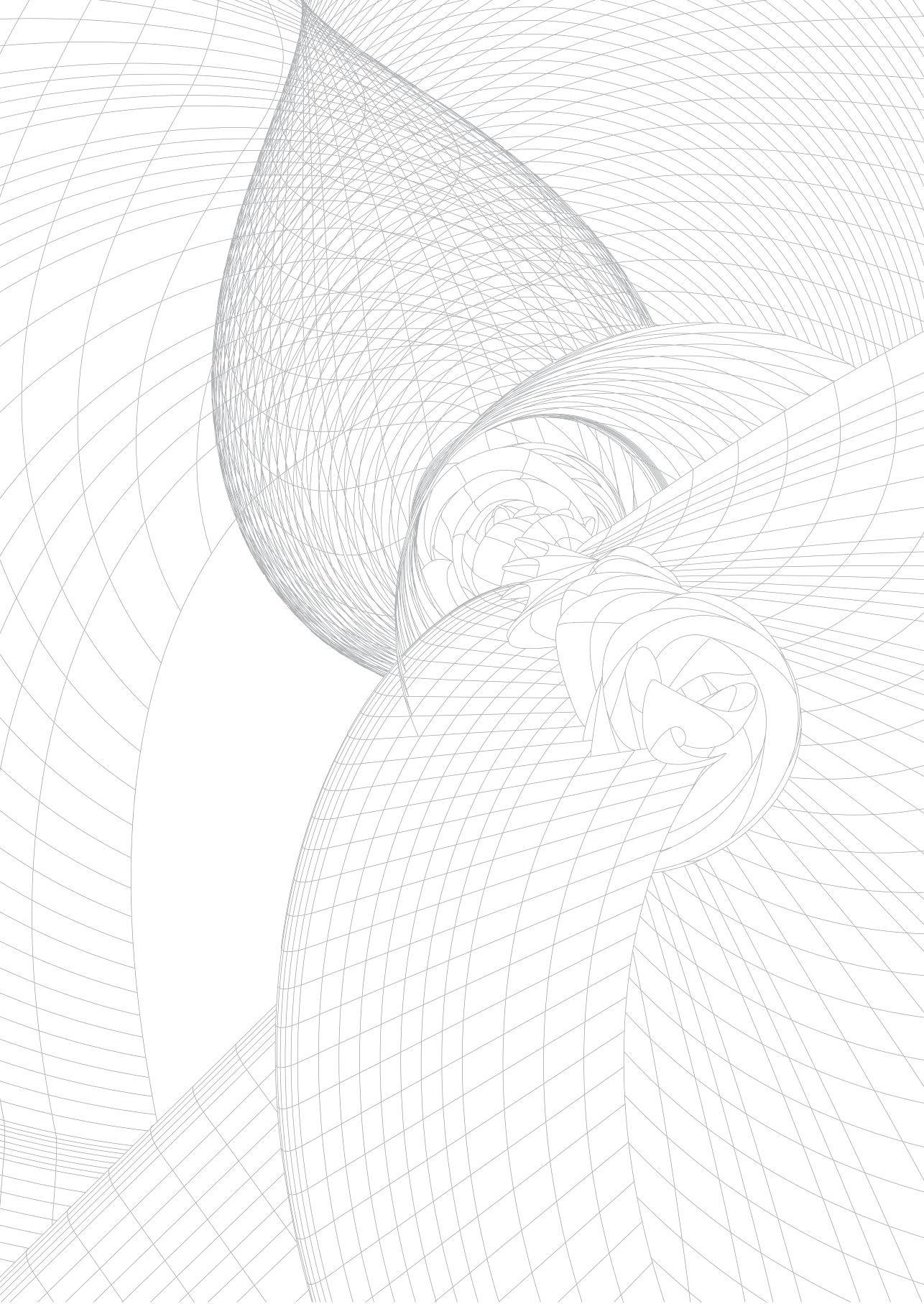
Detailed acknowledgements and online resources can be found in the published article online: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004423

REFERENCES

1. Cummings SR, Black DM, Nevitt MC, et al. Bone density at various sites for prediction of hip fractures. **The Lancet**. 1993;341:72-5.
2. Ferrari S, Rizzoli R, Slosman D, Bonjour JP. Familial resemblance for bone mineral mass is expressed before puberty. **The Journal of clinical endocrinology and metabolism**. 1998;83(2):358-61.
3. Cooper C, Westlake S, Harvey N, Javaid K, Dennison E, Hanson M. Review: developmental origins of osteoporotic fracture. **Osteoporos Int**. 2006;17(3):337-47.
4. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nature genetics**. 2012;44(5):491-501.
5. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. **N Engl J Med**. 2008;358(22):2355-65.
6. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. **Nat Genet**. 2009;41(11):1199-206.
7. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-Analysis of Genome-Wide Scans for Total Body BMD in Children and Adults Reveals Allelic Heterogeneity and Age-Specific Effects at the WNT16 Locus. **PLoS genetics**. 2012;8(7):e1002718.
8. Timpson NJ, Tobias JH, Richards JB, et al. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. **Hum Mol Genet**. 2009;18(8):1510-7.
9. Clark EM, Ness AR, Bishop NJ, Tobias JH. Association between bone mass and fractures in children: a prospective cohort study. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2006;21(9):1489-95.
10. Lewiecki EM. Managing osteoporosis: challenges and strategies. **Cleveland Clinic journal of medicine**. 2009;76(8):457-66.
11. Truswell AS. Osteopetrosis with syndactyly; a morphological variant of Albers-Schonberg's disease. **The Journal of bone and joint surgery British volume**. 1958;40-B(2):209-18.
12. Beighton P. Sclerosteosis. **Journal of medical genetics**. 1988;25(3):200-3.
13. Paternoster L, Lorentzon M, Lehtimaki T, et al. Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure. **PLoS genetics**. 2013;9(2):e1003247.
14. Styrkarsdottir U, Thorleifsson G, Sulem P, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. **Nature**. 2013;497(7450):517-20.
15. Koller DL, Zheng HF, Karasik D, et al. Meta-analysis of genome-wide studies identifies WNT16 and ESR1 SNPs associated with bone mineral density in premenopausal women. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2013;28(3):547-58.
16. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic acids research**. 2003;31(13):3812-4.
17. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. **Genome research**. 2012;22(9):1790-7.
18. Morgan EF, Barnes GL, Einhorn TA. The bone organ system: Form and Function. 4th ed. Marcus R, Feldam D, Dempster DW, Luckey M, Cauley JA, editors: Academic Press; 2013.
19. Gilbert SF. Osteogenesis: The Development of Bones. Developmental Biology. 6th ed. Sunderland (MA): Sinauer Associates; 2000.
20. Albagha OM, Wani SE, Visconti MR, et al. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. **Nature genetics**. 2011;43(7):685-9.
21. Dunford JE, Rogers MJ, Ebetino FH, Phipps RJ, Coxon FP. Inhibition of protein prenylation by bisphosphonates causes sustained activation of Rac, Cdc42, and Rho GTPases. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2006;21(5):684-94.

22. Lorentzon M, Landin K, Mellstrom D, Ohlsson C. Leptin is a negative independent predictor of areal BMD and cortical bone size in young adult Swedish men. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2006;21(12):1871-8.
23. Turner AS, Maillet JM, Mallinckrodt C, Cordain L. Bone mineral density of the skull in premenopausal women. **Calcified tissue international**. 1997;61(2):110-3.
24. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. **International journal of epidemiology**. 2012.
25. Jaddoe VW, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. **Eur J Epidemiol**. 2012;27(9):739-56.
26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American journal of human genetics**. 2007;81(3):559-75.
27. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature genetics**. 2006;38(8):904-9.
28. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. **Annu Rev Genomics Hum Genet**. 2009;10: 387-406.
29. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. **Genetic epidemiology**. 2010;34(8):816-34.
30. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. **Nature**. 2007;449(7164):851-U3.
31. Pennisi E. Genomics. 1000 Genomes Project gives new map of genetic diversity. **Science**. 2010; 330(6004):574-5.
32. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. **American journal of human genetics**. 2011;88(1):76-82.
33. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. **Bioinformatics**. 2012;28(19):2540-2.
34. Boston RC, Sumner AE. STATA: a statistical analysis system for examining biomedical data. **Advances in experimental medicine and biology**. 2003;537:353-69.
35. Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. **Bioinformatics**. 2009;25(20):2750-2.
36. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. **Bioinformatics**. 2010;26(17):2190-1.
37. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. **Bioinformatics**. 2010;26(18):2336-7.
38. International HapMap C. The International HapMap Project. **Nature**. 2003;426(6968):789-96.
39. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. **Bioinformatics**. 2008;24(24):2938-9.
40. Team RDC. R: A language and environment for statistical computing. ISBN 3-900051-07-02010.
41. Wickham H. ggplot2: Elegant graphics for data analysis.: Springer New York; 2009.
42. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. **American journal of human genetics**. 2004;74(4):765-9.
43. Neale MC, Psychiatry MCoVDo, Psychiatric VIf, Genetics B. MX: Statistical Modeling: Department of Psychiatry, Medical College of Virginia; 1997.
44. Grundberg E, Adoue V, Kwan T, et al. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. **PLoS genetics**. 2011;7(1):e1001279.
45. Grundberg E, Kwan T, Ge B, et al. Population genomics in a disease targeted primary cell model. **Genome research**. 2009;19(11):1942-52.

46. Jemtland R, Holden M, Reppe S, et al. Molecular disease map of bone characterizing the postmenopausal osteoporosis phenotype. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2011;26(8):1793-801.
47. Kalajzic I, Kalajzic Z, Kaliterna M, et al. Use of type I collagen green fluorescent protein transgenes to identify subpopulations of cells at different stages of the osteoblast lineage. **J Bone Miner Res**. 2002; 17(1):15-25.
48. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome biology**. 2009;10(3):R25.
49. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**. 2011;12:323.
50. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. **BMC Bioinformatics**. 2010;11:94.
51. Drabek K, van de Peppel J, Eijken M, van Leeuwen JP. GPM6B regulates osteoblast function and induction of mineralization by controlling cytoskeleton and matrix vesicle release. **Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research**. 2011; 26(9):2045-51.



Chapter 4.5

Bivariate genome-wide association analysis unveil the pleiotropic effects of *SREBF1* on BMD and lean mass in children

C. Medina-Gomez, J.P Kemp, N. L. Dimou, E. Kreiner-Møller, A. Chesi, D. H. M. Heppe, B. S Zemel, K. Bønnelykke, C. Boer, H. Bisggard, , E. Evangelou, Lynda F. Bonewald, Jeffrey P. Gorski, Matthew T. Maurano, Bram van der Eerden, Cheryl L Ackert-Bicknell,Jeroen van de Peppel, V. W. V. Jaddoe, A. G. Uitterlinden, J. H Tobias, G. Duque, S.F.A Grant,P.G. Bagos, D.M. Evans, F. Rivadeneira

In preparation

ABSTRACT

Lean and bone mass are heritable traits with high phenotypic correlation, likely reflecting the underlying mechanical and biochemical interactions between these tissues. Bone mineral density (BMD) is known to be under strong genetic control. Nonetheless, genetic variants contributing to lean mass variation remain largely unknown, yet a shared genetic component influencing both traits is expected. We aimed to estimate the shared SNP-heritability (genetic correlation) of both traits in children and to identify genetic determinants displaying pleiotropic effects on lean mass and bone mass accrual, by using a bivariate GWAS approach. We estimated the heritability of both traits in children to be between 30-45% with a genetic correlation of ~ 0.30. Bivariate GWAS meta-analysis in ~10,000 children identified seven established BMD loci as having an effect on both traits: *WNT4* ($P=4.7x10^{-9}$), *GALNT3* ($P=3.6x10^{-8}$), *MEPE* ($P=3.3x10^{-8}$), *CPED1/WNT16* ($P=3.1x10^{-20}$), *TNFSF11* ($P=1.3x10^{-8}$), *RIN3* ($P=3.3x10^{-8}$) and *PPP6R3/LRP5* ($P=2.2x10^{-9}$). Another signal ($P=1.4x10^{-10}$) with stronger effect on lean mass than on BMD mapped to the *TOM1L2/SREBF1* locus. Using publicly available databases for functional annotation, we identified active enhancers and other regulatory elements in relevant cells underlying this signal: We describe here the use of bivariate GWAS meta-analysis that powerfully replicate known BMD loci and unveil loci with pleiotropic effects on lean mass and BMD. SREBP-1, the product of *SREBF1,* an adipocyte differentiation factor, drives muscle protein synthesis by regulating *MYOD1*, *MYOG* and *MEF2C* factor and henceforth muscle cell differentiation. In addition, overexpression of this protein in osteoblastic cells increases the number of mineralized foci thus regulating mineralization.

INTRODUCTION

The interaction of skeletal muscle and bone has long been perceived as a mechanical one, where bone provides an attachment site for muscles and where muscles apply load to bone. Such muscle-derived forces drive an adaptive response in bone effecting its morphology, structure and strength as described by mechanostat theory [1]. Thus, a high phenotypic correlation between these traits is expected, as muscle mass constitutes a key contributor to bone mineral density (BMD) variation [2-4]. Whenever mechanical challenges exceed an acceptable level, bone tissue is added at the location where it is mechanically necessary [5]. Furthermore, in recent years, we have begun to appreciate that the tight coupling of these two tissues is much more complex, exceeding the mechanical interaction, and begins already during embryonic development, when bones and muscle cells share a common mesenchymal precursor [6]. Throughout life, there is a constant paracrine crosstalk across bone and muscle tissues, as they are both able to respond to the influence of common hormones and to produce and secrete factors that signal on nearby tissues and on each other [7]. Under this perspective of coupled development, growth, and physiological relationships, it is highly likely that common genetic determinants exert pleiotropic effects on each component of the musculoskeletal system.

The most studied bone phenotype at a genetic level is bone mineral density (BMD), as it is commonly used in clinical practice to diagnose osteoporosis and assess fracture risk. Heritability studies have demonstrated that between 50-85% of BMD variation can be explained by genetic factors [8]. Heritability estimates for muscle traits, as lean mass, grip strength, arm flexion and leg strength range between 30-65% [9,10]. Genetic association Studies (GWAS), which allow the simultaneous examination of several genetic variants (Single Nucleotide Polymorphisms (SNPs)) to identify their association with a trait of interest, has been successful in unveiling the genetic architecture of BMD with close to 60 different loci robustly associated with the trait at different body sites [11-14]. On the contrary, no robust associations have been reported today by GWAS of lean mass. Thus, the potential molecular genetic mechanisms regulating lean mass remain largely unknown. Nonetheless, twin studies have calculated the additive genetic correlation of bone and muscle traits to be above 30% [15,16].

Traditionally, GWAS test one trait at the time, although new methods have emerged which enable multivariate analysis to be performed. Multivariate methodologies have advantages such as increased power in the presence of genetic correlation, reduction of the multiple testing burden and biological insight in the case of pleiotropy [17]. Bivariate GWAS analysis of appendicular lean mass with femoral neck bone geometry [18] have suggested that the *GLYAT* gene underlies variation of bone size and body lean mass, although this result is yet to be replicated by other studies or proven in a functional setting.

Here we applied a bivariate GWAS approach for Total-body lean mass (TB-LM) and BMD (TB-BMD) in pediatric cohorts, to identify genes exerting pleiotropic effects on both traits and to elucidate possible biological pathways, which underlie the crosstalk between these two tissues during skeletal development.

METHODS

Study populations

Total body bone mineral density and lean mass were measured in four pediatric population-based studies: the Generation R Study, the Avon Longitudinal Study of Parents and their Children (ALSPAC), the Bone Mineral Density in Childhood Study (BMDCS) and the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) cohort. All participants underwent DXA scans; measurements were conducted by well-trained research assistants and daily quality control assurance was performed. Prior to the scan procedure, participants were asked to take off their shoes, heavy clothes, and metal accessories. As recommended by the International Society for Clinical Densitometry [19], total body less head BMD was the measurement used in the analysis in all four studies, henceforth TB-BMD. Also, all individuals included in this study had genome-wide array data imputed to the HapMap Phase II reference panel (build 36 release 22). All research aims and the specific measurements used in this study have been approved by the Medical Ethical Committee of each of the involved institutions and written informed consent was provided by all parents or custodians.

Statistical Methods

Total-body BMD less head and Total-Body Lean Mass measurements were derived from DEXA scans and adjusted by age, gender, height, fat percent and study specific covariates (genetic principal components and measurement center when applicable). Standardized residuals were then generated and used as input for all analyzes described here. Phenotypic correlation between these variables was assessed by Pearson correlation tests in each study independently.

SNP-Heritability and genetic correlation

To characterize the extent to which common genetic variants determine pediatric BMD and lean mass, and the shared genetic etiology of these traits, we applied a recently proposed approach for estimating SNP-heritability and genetic correlation based on genome-wide sharing between distantly related individuals, as implemented in the GCTA software [20,21]. These assessments were performed in independent genotyped markers data from the ALSPAC and Generation R studies. In the latter, given its multiethnic nature,

relatedness coefficients used in the genetic relationship matrix (GRM) were estimated from REAP (Relatedness Estimation in Admixed Populations) [22]. This modified matrix was then used as input for the GCTA analysis (Supplementary Note). No pairs exceeded the standard cutoff coefficient of 0.025 for genetic relatedness, confirming that no two individuals in the analysis were closer than third-degree cousins in either of the two studies. In total 431 and 1,043 individuals were excluded from ALSPAC and Generation R for this calculation.

Univariate and Bivariate GWAS

Analysis of the two quantitative traits was performed in each study individually based on best guess data using the "qt-command" available in PLINK [23] to obtain the means and standard deviations of both lean mass and BMD stratified by genotype. Meta-analysis of 2,276,811 SNPs present in at least two of the studies was performed following an inverse variance methodology. In the meta-analysis approach, we used a newly proposed method for bivariate meta-analysis, which is based on calculating the within-studies covariances of the outcome specific estimates [24]. The algorithm uses a general approach and has been proposed for both, discrete and continuous outcomes, but in the case of continuous outcomes, such as the one encountered here, a direct extension of the method proposed by Wei and Higgins [25] is applied. We used the additive model of inheritance (per-allele mean difference of the quantitative phenotype) for both outcomes and the method of moments for estimation with the *mvmeta* command in Stata [26]. Source code and details are given in (http://www.compgen.org/tools/multiple-outcomes). The overall bivariate association test was obtained by a standard Wald-type statistic (chi-square on 2 d.f.) that tests the null hypothesis that the particular SNP is not associated with either one of the outcomes. Genome-wide significance (GWS) was defined as a P value less than $5x10^{-8}$. Results were summarized in Manhattan plots generated in Easystrata [27] and QQplots from a modification of EPACTS [28] scripts.

Bioinformatic annotations and eQTL analysis

We identified variants in data from the 1000 Genomes Project that Phase 1 that are in strong linkage disequilibrium (LD) [CEU $r^2 > 0.8$] with the lead SNPs in the new susceptibility loci, using HaploReg 3 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php), for functional annotation. These SNPs were first searched in the GWAS catalog (https://www.genome.gov/26525384) to ascertain possible association with other traits. Identified nonsynonymous variants were interrogated for likely downstream functional consequences using SIFT (http://sift.jcvi.org/) and POLYPHEN2 (http://genetics.bwh.harvard.edu/pph2/) databases. Additionally, we interrogated two public databases for micro-RNAs, namely miRbase (http://www.mirbase.org/) [29] and miRDB (http://mirdb.org/miRDB/) [30]. Variants were also assessed for overlap with regions of predicted regulatory function generated by the Encyclopedia of DNA Elements (ENCODE) Project and the ROADMAP epigenomic

project including: regions of enhancer activity, DNase I hypersensitivity, local histone modifications or proteins bound to these regulatory sites in cell lines of potential interest as HSMM (Skeletal Muscle Myoblast), HSMMtube (Skeletal Muscle myotubes differentiated from the HSMM), SKMC (skeletal muscle cells) and when present in the different assays Osteobl (Ostoblasts). In addition, evidence of chromatin interactions was further explorer by Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) data from the ENCODE project in both MCF-7 and K562 oncologic line cells. The UCSC Genome Browser (https://genome.ucsc.edu/) was used for visualizing ENCODE data tracks indicative of regulatory function. Topologically associated domains (TADs) were also retrieved from a recent web-based tool for this end (http://promoter.bx.psu.edu/hi-c/download.html) for K562.

The eQTL analysis was completed using recently published eQTL data sets: a meta-analysis of the transcriptional profiles from the peripheral blood cells of 5,311 Europeans (http://genenetwork.nl/bloodeqtlbrowser) [31] and the GTEx database based on 1,641 samples across 43 tissues from 175 individuals (http://www.gtexportal.org/home/) [32].

RESULTS

Characteristics of the 10,414 children, from four different cohorts, included in this study are shown in **Table 1**. The phenotypic correlation between BMD and lean mass (after adjustment for sex, age, height, fat percent, and genomic principal components) was significant in all four cohorts and of similar magnitude: Generation R (ρ=0.437), ALSPAC (ρ=0.449), BMD-CS (ρ=0.486) and COPSAC (ρ=0.420).

Table 1. Anthropometric characteristics of study participants.

	Generation R		ALSPAC		BMD-CS		COPSAC	
	n=4,071		n=5,251		n=821		n=273	
Age, years	6.21	0.32	9.94	0.32	8.74	1.91	6.89	0.72
Women, %	2,035	49.98%	2,673	50.90%	431	61.80%	144	53.14%
CEU ancestry, %	2,171	53.32%	5,251	100.00%	634	77.22%	271	100.00%
Height, m	1.19	0.63	1.4	0.64	1.32	0.12	1.24	0.61
Weight, kg	23.08	4.08	34.7	7.41	30.78	8.76	24.6	4.25
Total body LH BMD, g/cm2	0.555	0.05	0.777	0.053	0.67	0.10	0.583	0.05
Total body LH BMC, g	528.5	104.21	891.92	181.85	726.12	205.92	2,932	591.04
Total body Lean Mass, g	16,393	2,284	24,553	3,184	22,360	5,830	17,460	2,600
Total body Fat Mass, g	5,862	2,328	8,561	5,108	7,590	3,580	7,010	2,490

SNP-Heritability and genetic correlation

SNP-Heritability and Genetic correlation for adjusted TB-BMD and TB-LM were calculated for the Generation R (N=3,027) and ALSPAC (N=4,820) studies and are displayed in **Table 2**. SNP-Heritability estimates were between 30 and 45% for both traits. All calculated estimates except the genetic correlation in Generation R were statistically significant (P<0.05).

Table 2. Trait SNP-heritability and genetic correlation TB-BMD/TB-LM in pediatric populations. Only unrelated individuals (e.g., no two individuals in the analysis were closer than third degree cousins in either of the two studies) were included in the analysis.

Study	N	TB-BMD Heritability			TB-LM Heritability			Genetic correlation TB-BMD/TB-LM		
		h^2	SE	P	h^2	SE	P	ρ	SE	P
Generation R	3,028	0.311	0.12	0.004	0.4	0.12	0.0003	0.299	0.21	0.126
ALSPAC	4,820	0.437	0.07	6x10-10	0.325	0.07	3x10-6	0.323	0.12	0.016

Univariate and Bivariate GWAS

Results from the univariate GWAS meta-analysis identified variants associated at GWS level with BMD in four different loci: *WNT16/CPED1* locus, leading SNP rs917727-T (beta=0.129SD, $P=1.28x10^{-16}$); *LRP5/PPP6R3* locus, leading SNP rs12272917-C (beta=-0.097SD, $P=1.42x10^{-9}$); *WNT4* locus, leading SNP rs3765350-G (beta=-0.094SD, $P=9.75x10^{-9}$); and the *GALNT3* locus, leading SNP rs6726821-G (beta=-0.077SD, $P=2.95x10^{-8}$). A Manhattan plot showing the results of this GWAS is displayed in **Figure 1**. The univariate GWAS meta-analysis of lean mass yield no GWS associations. QQ-Plots for these two analyses showed no evidence for early inflation as consequence of bias (i.e., possible stratification issues or genotyping errors) (**Supplementary Figure 1**).

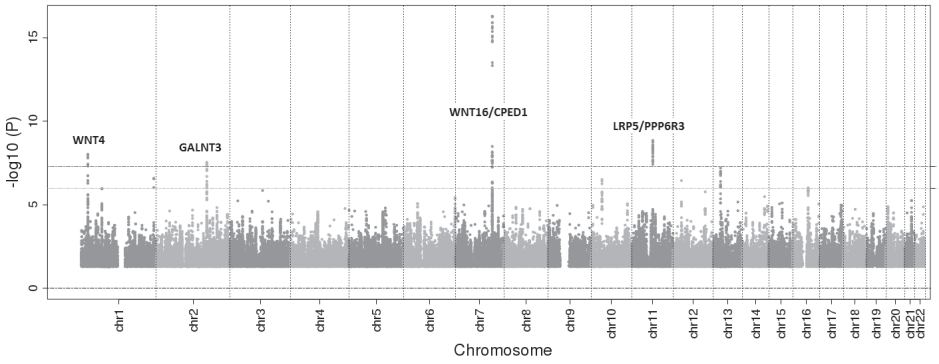


Figure 1. Manhattan plot of association statistics (-log10(P values)) for TB-BMD. Dashed red and yellow lines mark the GWS threshold ($P<5x10^{-8}$) and suggestive threshold ($P<1x10^{-6}$), respectively.

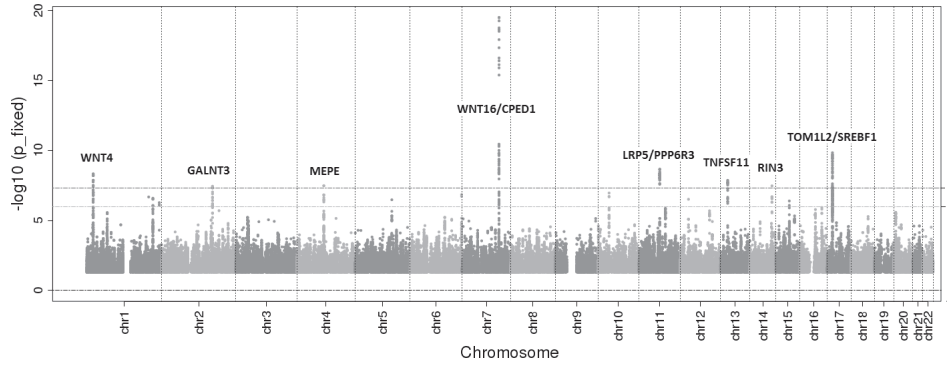


Figure 2. Manhattan Plot of the fixed effects bivariate meta-analysis of the association analysis of TB-BMD and TB-LM. Dashed red and yellow lines mark the GWS threshold ($P<5x10^{-8}$) and suggestive threshold ($P<1x10^{-6}$), respectively.

Based on the adjusted TB-BMD and TB-LM phenotypes, the genomic inflation factor λ was <1.08 (**Supplementary Figure 1**). Bivariate meta-analysis displayed eight different GWS peaks (**Figure 2**): 1p36.12, 2q24.3, 4q22.1, 7q31.31, 11q13.2, 13q14.11, 14q.2.12, 17p11.2, most of these associations driven by the TB-BMD analysis, except for signals in 2q24.3, 11q13.2 and 17p11.2 where there is evidence of association (P<0.05) with both traits (**Table 3**). In 1p36.12, the signal GWS SNPs extends for 264 Kb, with variants in *WNT4* and in close vicinity from the 5'region of *ZBTB40* (leading SNP rs6684375, $P=2.8x10^{-9}$). The low LD between the SNPs located at the opposite ends of this region are indicative of at least two independent association signals mapping to this locus (**Supplementary Figure 2a**). GWS variants harbored in 2q24.3 (**Supplementary Figure 2b**) are in high LD and located in the vicinity of 3'region of *CSRNP3* and *GALNT3* (leading SNP rs6726821, $P=2.8x10^{-8}$). The signal in 4q22.1 (**Supplementary Figure 2c**) is located in the vicinity of the 3'regions of *MEPE* (leading SNP rs11733405, $P=3.3x10^{-8}$). The signal in 7q31.31 spans for 307 Kb in a region harboring *CPED1*, *WNT16,* and *FAM3C* (**Supplementary Figure 2d**)*,* the LD pattern of this region is compatible with the presence of different signals underlying the captured association (leading SNP rs917727, $P=3.1x10^{-20}$). SNPs in 11q13.2 (**Supplementary Figure 2e**) are in high LD and spans throughout the *PPP6R3* gene, moreover, they are in moderate correlation with markers mapping to the 3'region of *LRP5* (leading SNP rs12271290, $P=4.4x10^{-9}$). Variants underlying the signal in 13q14.11 (**Supplementary Figure 2f**) are in high LD and located in *TNFSF11* or its vicinity (leading SNP rs9525638, $P=1.4x10^{-8}$). The signal on 14q.2.12 (**Supplementary Figure 2g**) arises from variants mapping to *RIN3* (leading SNP rs754388, $P=3.3x10^{-8}$). The signal on 17p11.2 (leading SNP rs7501812, $P=1.4x10^{-10}$) is the only of the GWS signals where associations are stronger with TB-LM than with TB-BMD, although all GWS SNPs show nominally significant opposite effect in both traits. The region underlying this signal (**Figure 3**) extends along an LD-block harboring several genes including *RAI1, LRC48, MIR33B, C17orf39, DRG2, MYO15A, SREBF1, TOM1L2, ATPAF2*, the latter four all shown to be expressed in skeletal muscle.

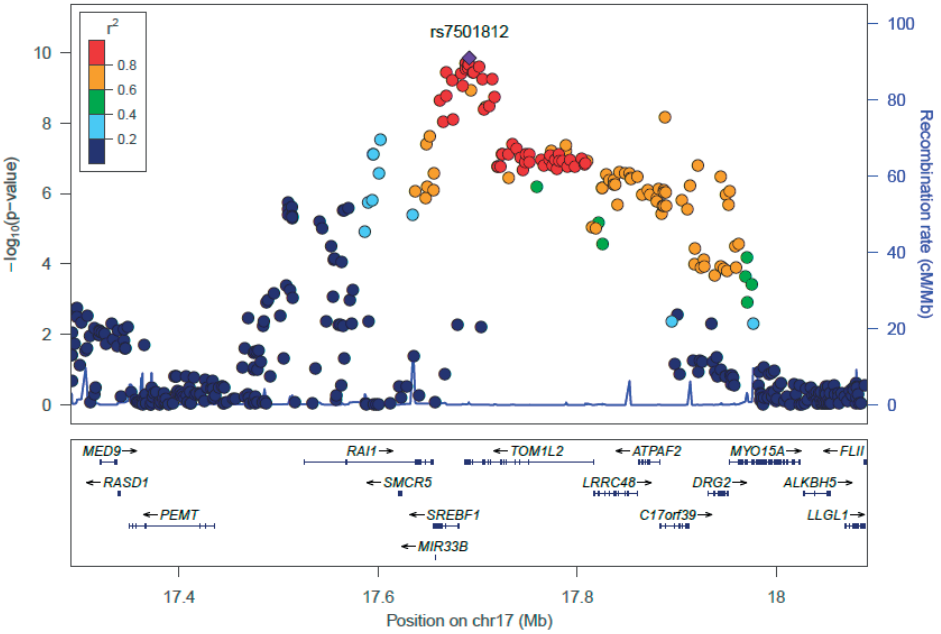


Figure 3. Regional association plot for the bivariate meta-analysis of TB-BMD and TB-LM displaying the 17p11.2 locus. Genetic coordinates are as per Hapmap phase II-CEU.

Bioinformatic annotations and eQTL analysis

Using Haploreg we identified proxies (r^2>0.8, 1000 Genomes CEU population) of our GWS SNPs in the 17p11.2 region to include them in an integrative approach to identify the possible association mechanism underlying this signal; this was made based on a literature review and several sources of bioinformatics' annotation. A total of 163 proxy SNPs were found from which 78 were present in our analysis (**Supplementary Table 1**). Only one SNP in the associated region was previously reported in the GWAS catalog, rs11868035. The G-allele of this SNP was reported to have a protective effect for Parkinson's disease in a meta-analysis of ~30,000 individuals ($P=5.6x10^{-8}$) [33]. The same SNP approaches significance in our bivariate meta-analysis ($P_{bivariate}=8x10^{-7}$) with the G-allele being nominally positively associated with lean mass (beta=0.05SD; $P=3x10^{-4}$). Yet, in our analysis, another SNP in high LD with rs11868035 (rs11654081, r^2>0.8) surpassed the GWS threshold (**Table 3**). In the proximity of these variants and also in high LD (r^2>0.8) with rs11654081, we identified two missense variants within the *RAI* gene (**Supplementary Table 1**). One of them (rs11649804, $P_{bivariate}=8.8x10^{-7}$) Pro165Thr, is predicted to be tolerated by SIFT and as probably damaging by Polyphen2. The other missense variant (rs3803763) Gly90Ala was not present in our analysis.

We further retrieved information on micro-RNA databases for the associated region and found it harbors two micro-RNAs, mir6777 and mir33b. This last microRNA is co-transcribed with *SREBF1*. Additionally, as most of the association signal in 17p11.2 is arising from non-coding variants, we reviewed the possible regulatory annotation of these SNPs by using the information from ENCODE and ROADMAP projects through the UCSC-browser. The visual summary of this information can be found in **Supplementary Figure 3**. Shadowed areas correspond to CTCF-associated areas of chromatin interaction by looping, as assessed by ChIA-PET in the MCF-7 and K562 cell-lines. As these ChiA-PET experiments are not available for musculoskeletal cells, we have chosen interactions, in MCF-7 and K562, for which the CTCF binding sites are also predicted to exist in the musculoskeletal cells. Three selected regions are shown, the first one brings together an intronic region of *TOM1L2* and the 3'region of *SREBF1*. The second one, two regions within *TOM1L2* and the third one, intronic regions from *TOM1L2* and *ATPAF2*. TADs were in line with these results evidencing complex chromatin structures (**Supplementary Figure 4**). Moreover, as the region described by the association signal is broad there are marks of histone modifications that are signatures of transcribed regions, transcription factor binding, active transcription sites, enhancers or repressed regions, and CPG island both methylated and unmethylated.

The SNP with the strongest bivariate association with TB-BMD and TB-LM in this locus, rs7501812, is also a *cis*-eQTL that was found to regulate the expression of *SREBF1, C17orf39, TOM1L2* and *ATAPF2* (FDR < 0.05) on whole blood. Nonetheless, *SREBF1* expression, represented by two probes in the dataset, showed the highest correlation with rs7501812. The G-allele, which was associated with higher TB-BMD (beta=0.043SD; $P=2x10^{-3}$) and lower TB-LM (beta=-0.056SD; $P=5.5x10^{-5}$) in our study, decreased the expression of the two *SREBF1* probes in whole blood. By interrogating the dataset per gene, rather than per SNP, we found that in general alleles from the GWS-bivariate SNPs associated with higher TB-BMD and lower TB-LM in the region are associated with decreased expression of *SREBF1*, *TOM1L2,* and *C17orf39* but increased expression of *ATPAF2* in whole blood (**Supplementary Table 1**). GTEx collection also supported the significant association of rs7501812 with the expression of *SREBF1* in whole blood. Additionally, the rs4925114 A-allele, nominally associated with higher TB-BMD (beta=0.036SD; P=0.01) and lower TB-LM (beta=-0.051SD; $P=3.6x10^{-4}$), significantly downregulated (FDR<0.01) the expression of *SREBF1* in skeletal muscle. Several other SNPs in the 17p11.2 locus were also reported as being associated with *SREBF1* expression in other tissues (data not shown).

Table 3. Genome-wide significant SNPs for the bivariate meta-analysis of total body bone mineral density (TB-BMD) and total body lean mass (TB-LM). Estimates were derived from 10,414 children participants of 4 different pediatric studies worldwide.

SNP	CHR	Position	A1	A2	EAF	Beta TB-BMD	P TB-BMD	Beta TBLM	P TBLM	P-bivariate	LD
rs3765350	1	22319903	G	A	0.23	-0.0939	9.75E-09	-0.0083	0.64	5.92E-09	0.005
rs2235529	1	22323074	T	C	0.15	-0.1064	1.60E-08	-0.0134	0.51	1.72E-08	0.011
rs3820282	1	22340802	T	C	0.15	-0.1077	1.13E-08	-0.0162	0.40	1.50E-08	0.002
rs12741884	1	22467282	A	G	0.22	-0.0912	3.73E-08	-0.0124	0.46	3.79E-08	0
rs7524102	1	22571034	G	A	0.19	0.0795	8.18E-06	-0.0290	0.11	1.35E-08	1
rs12568930	1	22574818	C	T	0.19	0.0790	9.65E-06	-0.0268	0.15	3.12E-08	1
rs10493013	1	22575622	C	T	0.19	0.0840	2.63E-06	-0.0270	0.14	5.61E-09	1
rs6684375	1	22579021	T	C	0.19	0.0842	2.47E-06	-0.0273	0.14	4.69E-09	top
rs6426749	1	22584060	C	G	0.19	0.0823	4.24E-06	-0.0291	0.11	6.26E-09	1
rs6426749	1	22584060	C	G	0.19	0.0823	4.24E-06	-0.0291	0.11	6.26E-09	1
rs12185748	2	166285202	T	C	0.45	-0.0762	3.98E-08	-0.0102	0.49	4.90E-08	0.966
rs7586085	2	166285735	G	A	0.45	-0.0764	3.68E-08	-0.0098	0.51	4.08E-08	0.966
rs6726821	2	166286360	G	T	0.45	-0.0769	2.95E-08	-0.0104	0.48	3.56E-08	top
rs6710388	2	166291387	T	C	0.45	-0.0769	2.95E-08	-0.0118	0.42	4.22E-08	0.966
rs6710518	2	166291490	T	C	0.45	-0.0767	3.18E-08	-0.0112	0.44	4.27E-08	0.966
rs7672749	4	89017308	A	G	0.09	0.0822	4.35E-04	-0.0591	0.01	3.29E-08	1
rs11733405	4	89017674	G	T	0.09	0.0822	4.35E-04	-0.0591	0.01	3.29E-08	1
rs2110281	7	120513045	A	G	0.35	-0.0859	3.24E-09	-0.0032	0.81	4.57E-10	0.003
rs2968345	7	120516108	G	A	0.35	-0.0838	6.90E-09	-0.0015	0.91	8.80E-10	0.003
rs6466767	7	120518813	G	C	0.34	-0.0827	1.22E-08	-0.0049	0.73	3.61E-09	0.003
rs1917113	7	120522298	A	G	0.34	-0.0840	7.88E-09	-0.0046	0.74	1.82E-09	0.003
rs12673968	7	120524294	A	G	0.34	-0.0825	1.35E-08	-0.0047	0.74	3.93E-09	0.003
rs6466769	7	120529339	G	A	0.34	-0.0831	1.14E-08	-0.0037	0.78	2.38E-09	0.003
rs6954757	7	120530418	A	G	0.34	-0.0831	1.12E-08	-0.0039	0.77	2.41E-09	0.003
rs13223036	7	120534544	G	T	0.35	-0.0825	1.29E-08	0.0007	0.98	1.14E-09	0.003
rs798943	7	120546135	A	G	0.38	-0.0789	3.36E-08	0.0148	0.30	9.86E-11	0.006
rs7801723	7	120561396	T	C	0.37	-0.0793	3.21E-08	0.0166	0.25	5.15E-11	0.006
rs12706318	7	120562177	G	A	0.37	-0.0798	2.46E-08	0.0119	0.41	1.53E-10	0.003
rs13232048	7	120563517	T	G	0.37	-0.0797	2.57E-08	0.0119	0.41	1.63E-10	0.003
rs6947453	7	120564419	T	G	0.37	-0.0720	4.71E-07	0.0141	0.33	3.87E-09	0.003
rs6952113	7	120564855	A	G	0.37	-0.0799	2.37E-08	0.0115	0.42	1.64E-10	0.003
rs872007	7	120567185	T	C	0.37	-0.0801	2.14E-08	0.0111	0.44	1.61E-10	0.003
rs10275439	7	120570661	A	G	0.37	-0.0777	5.18E-08	0.0143	0.31	2.06E-10	0.003
rs10261671	7	120570787	T	C	0.37	-0.0774	5.68E-08	0.0147	0.30	2.27E-10	0.003
rs13245690	7	120572300	G	A	0.37	-0.0800	2.37E-08	0.0164	0.25	3.51E-11	0.003
rs6950680	7	120577523	G	A	0.37	-0.0801	2.28E-08	0.0157	0.27	4.31E-11	0.003
rs7798060	7	120609622	T	C	0.40	-0.0715	4.34E-07	0.0219	0.12	4.16E-10	0.001
rs1554634	7	120613474	C	T	0.40	-0.0686	1.14E-06	0.0208	0.14	2.30E-09	0.001
rs10085590	7	120619259	G	A	0.41	-0.0665	2.32E-06	0.0220	0.12	4.21E-09	0.001
rs7797976	7	120630752	T	C	0.40	-0.0709	5.21E-07	0.0215	0.13	6.21E-10	0.001
rs6947494	7	120630944	T	C	0.41	-0.0670	1.98E-06	0.0209	0.14	4.76E-09	0
rs1917118	7	120631560	T	C	0.40	-0.0680	1.45E-06	0.0213	0.13	2.82E-09	0.004

Table 3 (continued)

SNP	CHR	Position	A1	A2	EAF	Beta TB-BMD	P TB-BMD	Beta-TBLM	P-TBLM	P-bivariate	LD
rs6954210	7	120635621	A	G	0.40	-0.0675	1.72E-06	0.0217	0.12	3.02E-09	0.001
rs6970762	7	120639270	T	A	0.40	-0.0692	9.64E-07	0.0242	0.09	6.30E-10	0
rs1357756	7	120639429	T	C	0.41	-0.0662	2.54E-06	0.0223	0.11	4.23E-09	0.001
rs1534015	7	120640301	A	G	0.41	-0.0663	2.49E-06	0.0224	0.11	4.08E-09	0.001
rs7786203	7	120640843	A	G	0.40	-0.0688	1.11E-06	0.0208	0.14	2.16E-09	0.001
rs1404268	7	120644983	A	G	0.41	-0.0648	4.16E-06	0.0209	0.14	1.08E-08	0.001
rs1524503	7	120655239	C	A	0.39	-0.0713	4.71E-07	0.0210	0.14	5.92E-10	0.001
rs3779381	7	120754026	G	A	0.26	0.1188	4.77E-14	-0.0054	0.72	1.25E-16	0.874
rs2908004	7	120757005	A	G	0.49	0.1127	2.83E-16	0.0012	0.98	1.19E-18	0.478
rs2536189	7	120760857	G	C	0.49	0.1109	8.03E-16	0.0012	0.99	4.55E-18	0.461
rs3801387	7	120762001	G	A	0.28	0.1281	1.29E-16	-0.0021	0.87	2.36E-19	1
rs2707466	7	120766325	T	C	0.48	0.1046	3.16E-14	0.0008	1.00	4.06E-16	0.505
rs2536182	7	120778073	G	C	0.48	0.1103	1.49E-15	0.0086	0.57	7.38E-17	0.462
rs2536180	7	120781909	C	T	0.49	0.1109	9.52E-16	0.0060	0.70	2.44E-17	0.471
rs3801382	7	120785513	G	T	0.28	0.1282	1.29E-16	-0.0029	0.83	1.72E-19	1
rs2254595	7	120794485	C	T	0.49	0.1096	1.80E-15	0.0049	0.76	3.88E-17	0.471
rs917727	7	120805815	T	C	0.29	0.1289	5.29E-17	-0.0042	0.77	3.07E-20	top
rs917726	7	120806093	T	A	0.28	0.1291	5.83E-17	-0.0045	0.75	3.26E-20	1
rs718766	7	120812738	C	T	0.28	0.1271	2.20E-16	-0.0041	0.77	2.23E-19	1
rs4727924	7	120819115	T	C	0.48	0.1125	4.33E-16	-0.0045	0.70	3.06E-19	0.471
rs7776725	7	120820357	C	T	0.27	0.1299	5.46E-17	-0.0031	0.81	5.45E-20	1
rs11228258	11	68010904	A	C	0.26	-0.0877	2.39E-08	-0.0661	2.53E-05	2.30E-08	0.955
rs7116994	11	68011203	T	C	0.26	-0.0893	1.33E-08	-0.0669	2.04E-05	1.29E-08	0.908
rs12294029	11	68013635	C	A	0.26	-0.0893	1.33E-08	-0.0669	2.04E-05	1.29E-08	0.908
rs7126340	11	68013869	T	C	0.26	-0.0874	2.53E-08	-0.0654	3.03E-05	2.62E-08	0.955
rs12272917	11	68019946	C	T	0.24	-0.0971	1.42E-09	-0.0639	7.07E-05	3.27E-09	1
rs11228262	11	68031483	T	G	0.25	-0.0948	2.65E-09	-0.0657	3.95E-05	4.55E-09	1
rs10896334	11	68038070	T	C	0.26	-0.0908	8.31E-09	-0.0687	1.25E-05	6.71E-09	0.955
rs10896337	11	68040812	C	T	0.26	-0.0895	1.30E-08	-0.0678	1.60E-05	1.08E-08	0.955
rs11228269	11	68046372	G	A	0.26	-0.0907	8.40E-09	-0.0706	7.00E-06	5.09E-09	0.955
rs7925275	11	68047009	C	T	0.25	-0.0939	4.19E-09	-0.0662	3.52E-05	6.17E-09	1
rs6591340	11	68054212	G	A	0.25	-0.0940	4.00E-09	-0.0657	3.94E-05	6.25E-09	1
rs7106259	11	68055088	G	T	0.26	-0.0909	7.95E-09	-0.0709	6.46E-06	4.67E-09	0.955
rs10896339	11	68059422	C	G	0.27	-0.0872	2.02E-08	-0.0713	4.65E-06	8.77E-09	0.818
rs12281742	11	68060205	C	T	0.28	-0.0852	3.81E-08	-0.0737	2.01E-06	8.92E-09	0.955
rs4316515	11	68065416	T	A	0.25	-0.0939	4.11E-09	-0.0658	3.87E-05	6.34E-09	1
rs7944870	11	68065684	G	C	0.26	-0.0926	4.34E-09	-0.0727	3.94E-06	2.21E-09	0.955
rs948315	11	68065716	C	T	0.26	-0.0875	2.60E-08	-0.0666	2.24E-05	2.31E-08	0.955
rs948316	11	68066246	T	G	0.25	-0.0939	4.11E-09	-0.0658	3.87E-05	6.34E-09	1
rs10896341	11	68069756	A	G	0.26	-0.0928	4.08E-09	-0.0721	4.74E-06	2.31E-09	0.955
rs7104345	11	68073348	A	G	0.25	-0.0939	4.11E-09	-0.0658	3.87E-05	6.34E-09	1
rs12284933	11	68076065	A	G	0.24	-0.0969	1.85E-09	-0.0699	1.49E-05	2.19E-09	top
rs11228284	11	68079343	T	A	0.25	-0.0939	4.11E-09	-0.0658	3.87E-05	6.34E-09	0.954
rs12271290	11	68082812	T	C	0.26	-0.0914	6.67E-09	-0.0705	7.47E-06	4.40E-09	0.955

Table 3 (continued)

SNP	CHR	Position	A1	A2	EAF	Beta TB-BMD	P TB-BMD	Beta-TBLM	P-TBLM	P-bivariate	LD
rs7102898	11	68085446	A	G	0.25	-0.0946	3.25E-09	-0.0660	3.58E-05	5.04E-09	1
rs2155730	11	68086050	C	T	0.26	-0.0876	2.32E-08	-0.0713	4.98E-06	9.51E-09	0.955
rs7109294	11	68088669	C	T	0.25	-0.0946	3.25E-09	-0.0660	3.58E-05	5.04E-09	1
rs2282563	11	68089776	T	C	0.25	-0.0949	2.81E-09	-0.0662	3.38E-05	4.37E-09	1
rs6591341	11	68092054	T	C	0.28	-0.0862	2.73E-08	-0.0741	1.74E-06	6.51E-09	0.955
rs3740631	11	68098298	G	T	0.25	-0.0949	2.81E-09	-0.0662	3.38E-05	4.37E-09	1
rs11228287	11	68106330	G	A	0.25	-0.0951	2.72E-09	-0.0664	3.22E-05	4.15E-09	1
rs10896347	11	68109703	A	G	0.25	-0.0949	2.81E-09	-0.0662	3.38E-05	4.37E-09	1
rs10896348	11	68113944	C	T	0.28	-0.0866	2.36E-08	-0.0739	1.84E-06	5.99E-09	0.955
rs7118897	11	68117256	A	G	0.25	-0.0921	6.78E-09	-0.0656	3.74E-05	1.01E-08	1
rs6591344	11	68117449	A	G	0.25	-0.0944	3.57E-09	-0.0660	3.59E-05	5.49E-09	1
rs7123564	11	68118740	T	C	0.25	-0.0963	1.69E-09	-0.0686	1.78E-05	2.23E-09	1
rs7127948	11	68119638	A	G	0.25	-0.0940	4.08E-09	-0.0657	3.98E-05	6.46E-09	1
rs3758643	11	68123769	T	C	0.25	-0.0929	6.23E-09	-0.0646	5.36E-05	1.05E-08	1
rs12283755	11	68128234	G	A	0.25	-0.0942	3.90E-09	-0.0655	4.19E-05	6.37E-09	1
rs7113287	11	68128762	T	A	0.28	-0.0866	2.36E-08	-0.0739	1.84E-06	5.99E-09	0.955
rs7104877	11	68134178	C	A	0.26	-0.0904	8.27E-09	-0.0717	4.54E-06	4.14E-09	0.955
rs2236708	11	68134621	A	G	0.25	-0.0932	5.39E-09	-0.0650	4.90E-05	8.92E-09	1
rs11228292	11	68136348	A	G	0.25	-0.0933	5.21E-09	-0.0652	4.67E-05	8.47E-09	1
rs4988291	11	68138183	A	G	0.25	-0.0933	5.21E-09	-0.0652	4.67E-05	8.47E-09	1
rs11228293	11	68141222	C	T	0.25	-0.0932	5.39E-09	-0.0650	4.90E-05	8.92E-09	1
rs7102273	11	68142155	C	T	0.26	-0.0895	1.26E-08	-0.0685	1.32E-05	9.69E-09	0.955
rs749934	13	42006245	A	G	0.46	0.0693	6.25E-07	-0.0110	0.43	1.56E-08	0.87
rs17458078	13	42009354	C	A	0.46	0.0684	8.48E-07	-0.0116	0.41	2.19E-08	0.875
rs9533143	13	42009405	C	G	0.46	0.0687	7.40E-07	-0.0112	0.42	2.05E-08	0.875
rs9525638	13	42026577	C	T	0.42	0.0756	6.42E-08	-0.0006	0.97	1.35E-08	top
rs1325798	13	42037049	T	C	0.42	0.0734	1.46E-07	-0.0016	0.89	2.98E-08	0.901
rs9533154	13	42038102	T	C	0.42	0.0744	1.01E-07	-0.0013	0.92	1.98E-08	1
rs17536328	13	42041029	T	C	0.42	0.0735	1.42E-07	-0.0023	0.85	2.47E-08	1
rs7325635	13	42043319	A	G	0.42	0.0741	1.18E-07	-0.0021	0.87	2.07E-08	1
rs9525643	13	42057516	C	T	0.42	0.0744	1.06E-07	-0.0021	0.87	1.83E-08	1
rs754388	14	92185163	G	C	0.18	-0.0862	3.26E-06	0.0210	0.24	3.36E-08	top
rs4925109	17	17602527	A	G	0.35	0.0202	0.163	-0.0641	1.06E-05	3.09E-08	0.445
rs11654081	17	17648854	T	C	0.41	0.0318	0.025	-0.0531	1.84E-04	3.90E-08	0.765
rs4925114	17	17651995	A	G	0.41	0.0362	0.011	-0.0508	3.64E-04	2.39E-08	0.765
rs4925115	17	17662182	A	G	0.39	0.0388	0.007	-0.0550	1.18E-04	2.37E-09	0.806
rs11656665	17	17665514	G	A	0.43	0.0361	0.009	-0.0523	1.77E-04	9.50E-09	0.881
rs9899634	17	17668668	T	A	0.41	0.0415	0.003	-0.0528	1.93E-04	1.69E-09	0.92
rs8066560	17	17668768	A	G	0.38	0.0422	0.003	-0.0559	8.84E-05	3.82E-10	0.881
rs9902941	17	17674485	C	T	0.40	0.0398	0.005	-0.0566	6.26E-05	6.22E-10	0.92
rs1889018	17	17675465	G	A	0.42	0.0387	0.005	-0.0506	2.78E-04	8.20E-09	0.92
rs4925119	17	17683629	A	G	0.39	0.0419	0.003	-0.0561	7.98E-05	3.92E-10	0.92
rs9891957	17	17685164	G	A	0.40	0.0431	0.002	-0.0524	1.82E-04	8.79E-10	0.92
rs3183702	17	17688014	A	G	0.39	0.0433	0.002	-0.0559	8.05E-05	2.05E-10	0.96

Table 3 (continued)

SNP	CHR	Position	A1	A2	EAF	Beta TB-BMD	P TB-BMD	Beta-TBLM	P-TBLM	P-bivariate	LD
rs2236513	17	17688091	A	C	0.39	0.0433	0.002	-0.0559	8.05E-05	2.05E-10	0.96
rs9915248	17	17688239	C	T	0.40	0.0417	0.003	-0.0561	6.81E-05	3.07E-10	0.96
rs3744115	17	17690547	G	A	0.40	0.0421	0.003	-0.0558	7.05E-05	2.86E-10	0.96
rs1052299	17	17691144	A	G	0.40	0.0440	0.002	-0.0549	9.09E-05	1.96E-10	0.96
rs1108648	17	17691283	G	A	0.41	0.0424	0.002	-0.0560	6.11E-05	2.21E-10	0.96
rs7501812	17	17691632	G	A	0.41	0.0431	0.002	-0.0563	5.53E-05	1.44E-10	top
rs12951376	17	17693534	T	C	0.31	0.0369	0.014	-0.0617	3.41E-05	1.17E-09	0.672
rs4925120	17	17695358	C	T	0.40	0.0458	0.001	-0.0525	2.22E-04	3.61E-10	0.96
rs16960744	17	17695984	A	G	0.40	0.0453	0.001	-0.0528	2.03E-04	3.77E-10	0.96
rs8079321	17	17701514	T	C	0.40	0.0446	0.002	-0.0541	1.32E-04	2.64E-10	0.92
rs4924823	17	17705227	A	G	0.39	0.0404	0.004	-0.0567	6.51E-05	5.87E-10	0.92
rs12941039	17	17706380	C	T	0.43	0.0396	0.004	-0.0515	2.22E-04	4.29E-09	0.92
rs1889014	17	17707890	T	C	0.53	-0.0409	0.003	0.0499	3.33E-04	3.46E-09	0.781
rs9907246	17	17711690	T	C	0.42	0.0392	0.005	-0.0528	1.58E-04	3.33E-09	0.92
rs9907287	17	17714843	C	T	0.39	0.0403	0.005	-0.0569	6.15E-05	5.71E-10	0.92
rs8080061	17	17717114	C	T	0.40	0.0416	0.003	-0.0530	1.90E-04	1.82E-09	0.92
rs4925125	17	17735169	T	C	0.41	0.0386	0.006	-0.0481	6.26E-04	4.09E-08	0.813
rs6502625	17	17789011	G	C	0.53	-0.0385	0.005	0.0461	8.95E-04	4.46E-08	0.74
rs2955382	17	17888435	T	C	0.54	-0.0363	0.008	0.0528	1.48E-04	6.73E-09	0.75

DISCUSSION

We have confirmed the important role of genetic factors in both TB-BMD and TB-LM at an early age. Heritability estimates, calculated in more than 9,000 children, are in the range of 30-45% percent for both traits. Moreover, we found evidence for the presence of share genetic influences across the traits, estimating a genetic correlation of about 30%. The bivariate GWAS meta-analysis we conducted identified eight loci associated with both TB-BMD and TB-LM. All loci but 17p11.2 -harboring genes as *RAI1, LRRC48, MIR33B, SREBF1, TOM1L2, ATPAF2*- have been previously reported as associated with BMD in large genetic studies. The 17p11.2 locus is marked by a long stretch of linkage disequilibrium and several lines of evidence point to an enrichment of DNA looping interactions in this region, making it difficult to unveil the gene responsible for the association signal.

SNP-heritability in two of the participant cohorts reached similar results for both TB-BMD and TB-LM, consistent with genetic factors influencing each of them. We found that about one-third of the additive genetic effect is shared between TB-BMD and TB-LM. These estimates are lower than those derived from a previous twin study that estimated the heritability for TB-BMD to be between 80-90%, for TB-LM between 70-88% and a genetic correlation between these traits of 46% [(34)]. Similar to other complex traits, SNP-heritability estimates

for TB-BMD and TB-LM were lower than heritability estimates from twin studies [35,36]. One explanation could be that the SNP-heritability is attributed to the common SNPs while that estimated with the related individuals is attributed to the entire genome. Also, we have used only genotyped independent variants thus, if the selected markers are not sufficiently correlated with the true genetic variants explaining the trait-variance heritability would certainly be underestimated [37].

We used a bivariate approach for meta-analysis of GWAS. This methodology proved to be more powerful than the univariate analysis, particularly for those variants, exerting an opposite effect on both traits under study. The univariate approach for TB-BMD identified only four GWS signals, all of them reported before in BMD site-specific GWAS [12,13,38]. Up to date, four loci have been identified to be associated with TB-LM in adults, but none of them overlap with the loci identified in this research (submitted manuscript). Further implementation of the bivariate methodology resulted in eight loci identified. Results showed that except for signals in 2q24.3, 11q13.2 and 17p11.2, all other GWS signals were pulled to significance by their association with TB-BMD, rather than by an association ($P<0.05$) with both traits.

A previous study for TB-BMD including only Generation R and ALSPAC GWAS identified as many as seven different loci associated with TB-BMD [13]. The difference in the number of discovered loci (7 vs. 4) could have derived from the model used to adjust TB-BMD. In Kemp et al. [13] we adjusted the outcome for age, sex, weight, height and principal components while in this new effort we included age, sex, height, fat percent and principal components on analysis. As recently demonstrated by Aschard et al. [39], adjusting for heritable covariates, as weight, height, and fat percent can affect the association results. In the present study then, we replaced weight by fat percent. Fat percent is a heritable variable but its correlation with TB-BMD and TB-LM in Generation R is 0.3 and 0.1 which would represent a rather small bias in the effect size of variants correlated with this trait and TB-BMD/TB-LM, and thus this would not likely resulting in a false discovery. Moreover, for protein products of genes as *LRP5* [40], *TNSFI1*[41], *WNT16* [42] and *WNT4* [43] identified in our study and previous GWAS meta-analyses [11-13,38,44-46] –regardless of the adjustment variables used in the association model- their role in bone metabolism is long-established.

We focused our effort on the region of 17p11.2 as it is new in the GWAS of musculoskeletal traits and was the only region in which the association signal was stronger in TB-LM than in TB-BMD. This region was described recently as a QTL in heterogeneous stock rat [47]. In the same publication, they attribute the possible association to *Rai1* as previous studies have shown knockout mice for this gene are growth retarded and display malformations in both the craniofacial and the axial skeleton [48]. This gene in humans is responsible for

the Smith-Magenis syndrome, which as in mice is presented with craniofacial and skeletal abnormalities [49]. Nonetheless, the two exonic variants in LD with the bivariate GWS SNPs have not been reported as responsible for the disease. We retrieved different lines of evidence to disentangle the gene(s) underlying the association signal although results were not conclusive as, besides a long stretch of LD in the region there is evidence of complex chromatin interactions that could result in regulation by the associated variants of more than one gene. These chromatin interactions assessed by ChIA-PET experiments were retrieved for two oncogenic cell lines. Nonetheless, the CTCFs described are also present in the muscular cells and osteoblast, and TADs have been shown to be largely invariant across cell types, and physiological conditions [50]. Therefore, we believe similar interactions might take place in osteoblasts and muscular cells, suggesting that associated SNPs can be influencing distant genes within the TAD. For example, The SNP leading the signal (rs7501812) has a strong correlation with the expression of various genes in the region in whole blood (**Supplementary Table 1**), in line with the predicted complex looping in this locus. However, its correlation with *SREBF1* is the strongest. In addition, another bivariate-GWS SNP acts as *cis*-eQTL of this gene in skeletal muscle. Hence, it is reasonable to believe that *SREBF1* is the gene underlying the signal.

SREBF1, an adipocyte differentiation factor, produces SREBP-1, a ubiquitously expressed transcription factor, strongly expressed in lipogenic tissues [51]. In skeletal muscle SREBP-1 proteins regulate the expression of hundreds of genes, but particularly its role in down-regulating *MYOD1*, *MYOG* and *MEF2C* suggests it acts as a key regulator of myogenesis. In-keeping with this view, overexpression of SREBP-1 inhibits myoblast–to-myotube differentiation [52], reduces cell size and leads to loss of muscle-specific proteins in differentiated myotubes [51]. Contrary to its inhibitory role in myogenesis, SREBP-1, in its active form, has been shown to be important for the mineralization of osteoblastic cultures in vitro and bone formation in vivo as its overexpression increases the number of mineralized foci [53].

Thus, there is compelling evidence that *SREBF1* has opposite effects on the differentiation of myocytes and osteoblasts. This opposite effect is shown in the SNPs associated with TB-BMD/TB-LM in our bivariate approach, and therefore, it is conceivable to think that *SREBF1* is underlying the association signal. Nonetheless, the eQTL analysis seems to contradict the expected effect direction. The G-allele from rs7501812, shown to downregulate the expression of *SREBF1* in whole blood, is actually associated with higher TB-BMD and lower TB-LM, which is opposite to that predicted by its function bone and muscle metabolism. This different behavior could be due to tissue-specific effects. Additionally, SREBP-1 need to be activated by proteolytic cleavage in the endoplasmic reticulum, nuclear transport and intra-nuclear activation/degradation within the nucleus, actions that cannot be captured by expression analysis.

In summary, in this first large-scale bivariate GWAS meta-analysis of TB-BMD and TB-LM in pediatric cohorts, eight GWAS loci (in/near *WNT4, GALNT3, MEPE, CPED1/WNT16, TNFSF11, RIN3, TOM1L2/SREBF1,* and *PPP6R3/LRP5*) have been identified. Although for most of these loci the association with TB-BMD is the dominant association responsible for their significance. Only the 17p11.2 locus containing among other genes *TOM1L2* and *SREBF1*, showed a stronger association with TB-LM than with TB-BMD. The genomic region described by the association signal in this locus spans ~290Kb in a long stretch of LD and chromatin interactions. Nonetheless, *SREBF1* arises as a strong candidate to underlie the association as this transcription factor is known to play an opposite role in the differentiation of osteoblasts and myoblasts. To disentangle the mechanism by which associated variants in this study influence SREBF1 action in musculoskeletal metabolism further functional studies are required. This study demonstrates a considerable increase in power can be achieved when using a bivariate methodology as compared to testing each component of a multivariate phenotype individually, particularly in those cases where the effect of the associated variant goes in opposite directions in the involved phenotypes. Moreover, we show that this methodology may help discover previously unknown mechanisms of pleiotropic effects of genetic variants.

Additional detailed methods, results and acknowledgements are to be provided in the Online Supplement of the journal containing this publication.

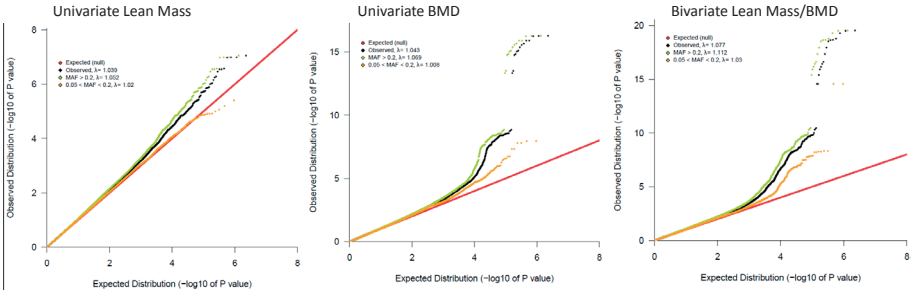
REFERENCES

1. Robling AG. Is bone's response to mechanical signals dominated by muscle forces? **Medicine and science in sports and exercise**. 2009;41(11):2044-9.
2. Li SL, Wagner R, Holm K, Lehotsky J, Zinaman MJ. Relationship between soft tissue body composition and bone mass in perimenopausal women. **Maturitas**. 2004;47(2):99-105.
3. Torres-Costoso A, Gracia-Marco L, Sanchez-Lopez M, et al. Lean mass as a total mediator of the influence of muscular fitness on bone health in schoolchildren: a mediation analysis. **J Sports Sci**. 2015; 33(8):817-30.
4. Van Langendonck L, Claessens AL, Lefevre J, et al. Association between bone mineral density (DXA), body structure, and body composition in middle-aged men. **Am J Hum Biol**. 2002;14(6):735-42.
5. Rauch F, Schoenau E. The developing bone: slave or master of its cells and molecules? **Pediatric research**. 2001;50(3):309-14.
6. Brotto M, Johnson ML. Endocrine Crosstalk Between Muscle and Bone. **Curr Osteoporos Rep**. 2014; 12(2):135-41.
7. Cianferotti L, Brandi ML. Muscle-bone interactions: basic and clinical aspects. **Endocrine**. 2014;45(2): 165-77.
8. Ralston SH, Uitterlinden AG. Genetics of Osteoporosis. **Endocr Rev**. 2010;31(5):629-62.
9. Silventoinen K, Magnusson PKE, Tynelius P, Kaprio J, Rasmussen F. Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. **Genet Epidemiol**. 2008;32(4): 341-9.
10. Arden NK, Spector TD. Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. **J Bone Miner Res**. 1997;12(12):2076-81.
11. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **PLoS Genet**. 2012;8(7):e1002718.
12. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
13. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **PLoS Genet**. 2014;10(6):e1004423.
14. Duncan EL, Danoy P, Kemp JP, et al. Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. **PLoS Genet**. 2011;7(4): e1001372.
15. Park JH, Song YM, Sung J, et al. The association between fat and lean mass and bone mineral density: The Healthy Twin Study. **Bone**. 2012;50(4):1006-11.
16. Bogl LH, Latvala A, Kaprio J, Sovijarvi O, Rissanen A, Pietilainen KH. An Investigation into the Relationship Between Soft Tissue Body Composition and Bone Mineral Density in a Young Adult Twin Sample. **Journal of Bone and Mineral Research**. 2011;26(1):79-87.
17. Galesloot TE, van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH. A Comparison of Multivariate Genome-Wide Association Methods. **Plos One**. 2014;9(4).
18. Guo YF, Zhang LS, Liu YJ, et al. Suggestion of GLYAT gene underlying variation of bone size and body lean mass as revealed by a bivariate genome-wide association study. **Hum Genet**. 2013;132(2):189-99.
19. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
20. Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. **Annu Rev Genet**. 2013;47:75-95.

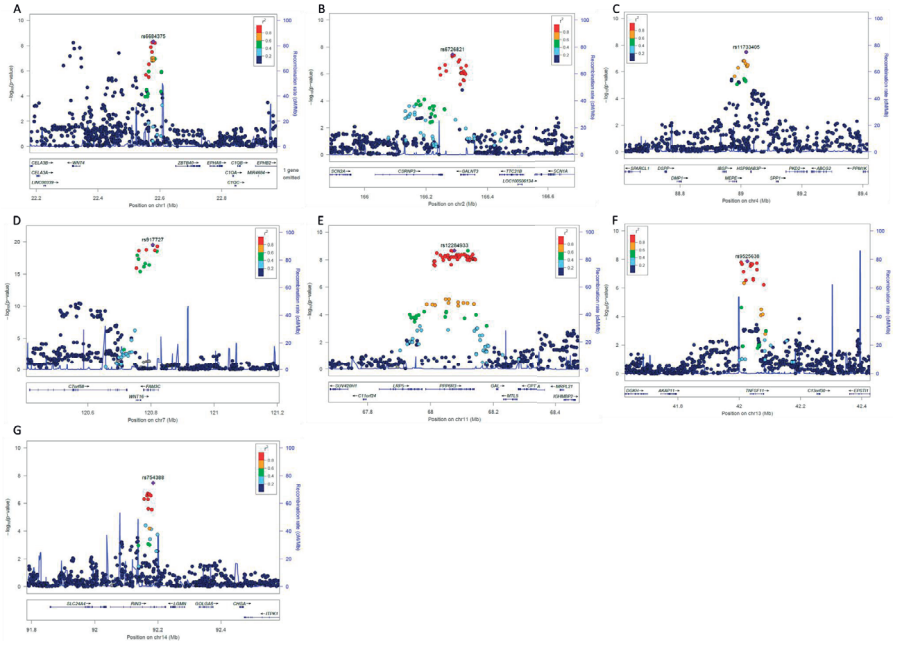
21. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. **Bioinformatics**. 2012;28(19):2540-2.
22. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. **Am J Hum Genet**. 2012;91(1):122-38.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**. 2007;81(3):559-75.
24. Dimou NL, Bagos PG. A Multivariate Method for Meta-Analysis of Multiple Outcomes in Genetic Association Studies. 35th Annual Conference of the International Society for Clinical Biostatistics; Vienna, Austria2014. p. 26.
25. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. **Stat Med**. 2012.
26. White IR. Multivariate random-effects meta-analysis. **Stata Journal**. 2009;9:40-56.
27. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. **Bioinformatics**. 2015;31(2):259-61.
28. Lab CfSA. Efficient and parallelizable association container toolbox (EPACTS). University of Michigan School of Public Health; 2013.
29. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. **Nucleic Acids Res**. 2014;42(D1):D68-D73.
30. Wong N, Wang XW. miRDB: an online resource for microRNA target prediction and functional annotations. **Nucleic Acids Res**. 2015;43(D1):D146-D52.
31. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. **Nature Genetics**. 2013;45(10):1238-U195.
32. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. **Science**. 2015;348(6235):648-60.
33. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. **PLoS Genet**. 2011;7(6):e1002141.
34. Bogl LH, Latvala A, Kaprio J, Sovijarvi O, Rissanen A, Pietilainen KH. An investigation into the relationship between soft tissue body composition and bone mineral density in a young adult twin sample. **J Bone Miner Res**. 2011;26(1):79-87.
35. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. **Nature Genetics**. 2015;47(10):1114.
36. Arpegard J, Viktorin A, Chang Z, de Faire U, Magnusson PKE, Svensson P. Comparison of Heritability of Cystatin C- and Creatinine-Based Estimates of Kidney Function and Their Relation to Heritability of Cardiovascular Disease. **J Am Heart Assoc**. 2015;4(1).
37. Visscher PM, Goddard ME. A General Unified Framework to Assess the Sampling Variance of Heritability Estimates Using Pedigree or Marker-Based Relationships. **Genetics**. 2015;199(1):223-32.
38. Koller DL, Zheng HF, Karasik D, et al. Meta-analysis of genome-wide studies identifies WNT16 and ESR1 SNPs associated with bone mineral density in premenopausal women. **J Bone Miner Res**. 2013; 28(3):547-58.
39. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for Heritable Covariates Can Bias Effect Estimates in Genome-Wide Association Studies. **American Journal of Human Genetics**. 2015; 96(2):329-39.
40. Gong Y, Slee RB, Fukai N, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. **Cell**. 2001;107(4):513-23.
41. Kartsogiannis V, Zhou H, Horwood NJ, et al. Localization of RANKL (receptor activator of NF kappa B ligand) mRNA and protein in skeletal and extraskeletal tissues. **Bone**. 1999;25(5):525-34.
42. Moverare-Skrtic S, Henning P, Liu X, et al. Osteoblast-derived WNT16 represses osteoclastogenesis and prevents cortical bone fragility fractures. **Nat Med**. 2014;20(11):1279-88.

43. Yu B, Chang J, Liu Y, et al. Wnt4 signaling prevents skeletal aging and inflammation by inhibiting nuclear factor-kappaB. **Nat Med**. 2014;20(9):1009-17.
44. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. **Nat Genet**. 2009;41(11):1199-206.
45. Richards JB, Rivadeneira F, Inouye M, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. **Lancet**. 2008;371(9623):1505-12.
46. Richards JB, Zheng HF, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges (vol 13, pg 576, 2012). **Nat Rev Genet**. 2012;13(9):672-.
47. Alam I, Koller DL, Canete T, et al. High-resolution genome screen for bone mineral density in heterogeneous stock rat. **J Bone Miner Res**. 2014;29(7):1619-26.
48. Bi W, Ohyama T, Nakamura H, et al. Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith-Magenis syndrome. **Hum Mol Genet**. 2005;14(8): 983-95.
49. Elsea SH, Girirajan S. Smith-Magenis syndrome. **Eur J Hum Genet**. 2008;16(4):412-21.
50. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. **Nature**. 2012;485(7398):376-80.
51. Dessalle K, Euthine V, Chanon S, et al. SREBP-1 transcription factors regulate skeletal muscle cell size by controlling protein synthesis through myogenic regulatory factors. **Plos One**. 2012;7(11):e50878.
52. Lecomte V, Meugnier E, Euthine V, et al. A new role for sterol regulatory element binding protein 1 transcription factors in the regulation of muscle mass and muscle cell differentiation. **Mol Cell Biol**. 2010;30(5):1182-98.
53. Gorski JP, Huffman NT, Chittur S, et al. Inhibition of proprotein convertase SKI-1 blocks transcription of key extracellular matrix genes regulating osteoblastic mineralization. **J Biol Chem**. 2011;286(3): 1836-49.

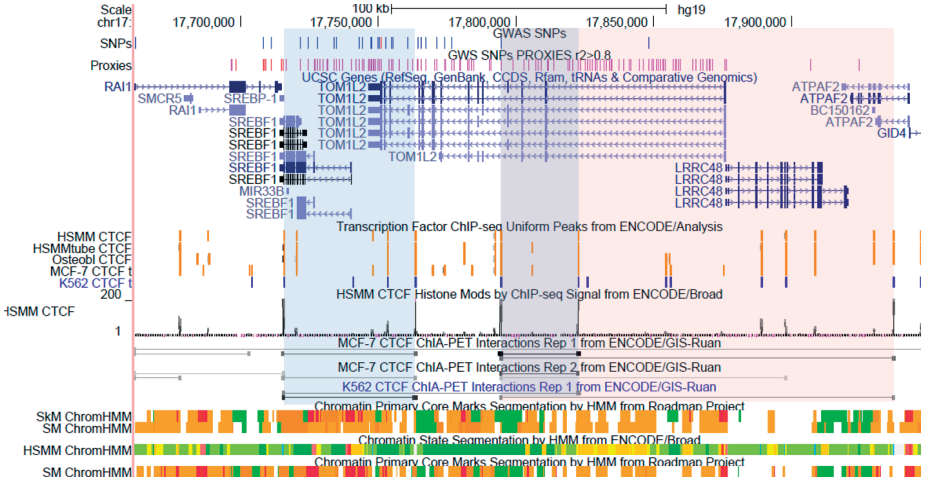
SUPPLEMENTARY MATERIAL



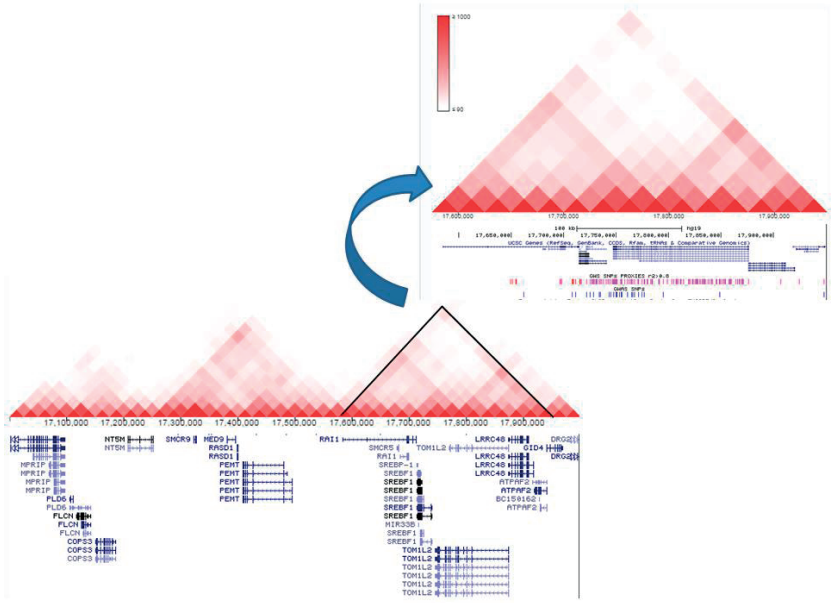
Supplementary Figure 1. QQPlots for the genome-wide meta-analyses. Quantile-Quantile plots showing P-Values for association, deviating from the null hypothesis of no association (identity line), per minor allele frequency category.



Supplementary Figure 2. Regional plots. Regional association plot for each of the genome-wide significant loci. (**A. 1p36.12, B. 2q24.3, C. 4q22.1, D. 7q31.31, E. 11q13.2, F. 13q14.11, G.14q.2.12**). SNPs are plotted by position in a 1Mb window against association with the bivariate LM/BMD (–log10 P). Plot highlighting the most significant SNP in the meta-analysis . Blue peaks indicate recombination rates. The SNPs surrounding the most significant SNP are color coded to reflect their LD with this SNP (from pairwise r^2 values from the HapMap CEU). Genes, exons and the direction of transcription from the UCSC genome browser are noted.



Supplementary Figure 3. 17p11.2 locus displaying binding peaks and interaction tracks by CTCF ChIA-PET data together with histone marks based on chromatin state characterization. Predicted strong interactions shared in MCF-7 or K562 cell lines are shadowed in blue while a weaker interaction in both cell lines is shadowed in pink. These interactions overlap with predicted CTCF binding sites in osteoblast, myoblast and myotubes, hypothetically localizing SNPs in *TOM1L2* close to transcribed regions in *SREBF1* and A*TPAF2*. Histone markers in muscular cells display high representation of associated SNPs in active enhancer predicted regions. Chromatin states are defined as follows, Bright Red: Active Promote, Orange: Strong enhancer, Yellow: Weak/poised enhancer, Blue: Insulator, Dark Green: Transcriptional elongation, Light Green: Weak transcribed



Supplementary Figure 4. A topological associated domain (TAD) includes associated variants and *RAI1, SREBF1, TOM1L2*, *LRRC8* and *ATPAF2*. In line with the chromatin interactions, TADs in K562 cell line suggest a complex regulation structure in the 17p11.2 locus.

Supplementary Table 1. Blood cis-eQTL (FDR<0.05) for TB-BMD and TB-LM bivaritely associated SNPs and proxies in 17p11.2. Quantities under the gene headers represent z-scores for the cis-expression effect of allele A1 on that particular gene. Positive values denote an increase in expression and negative values a decrease in gene expression. SNPs present in the analysis are shadowed. Proxies are chosen to be in high LD (r^2>0.8) with analysed SNPs (GWS). Effect sizes and P-values for the univariate and bivariate associations are also provided. Missense variants are in bold.

CHR	BP	SNP	A1	GWSass.	r^2	SREBF1	TOM1L2	ATPAF2	C17orf39	β-TBLM	P-TBLM	β-TBBMD	P-BMD	P-bivariate
17	17807622	rs10048206	G	rs9899634	0.81	-17.63	-5.34	5.14	-8.53	-0.046	1.27E-03	0.038	0.007	1.49E-07
17	17691144	rs1052299	A	rs1052299	1.00	-20.05	-4.55	4.22	-8.62	-0.055	9.09E-05	0.044	0.002	1.96E-10
17	17781548	rs11078406	T	rs9899634	0.81	-17.62	-5.44	5.32	-8.49	-0.046	9.58E-04	0.039	0.006	7.60E-08
17	17691283	rs1108648	G	rs1108648	1.00	-20.05	-4.57	4.21	-8.62	-0.056	6.11E-05	0.042	0.002	2.21E-10
17	17637480	**rs11649804**	A	rs11654081	0.88	-15.52	.	.	.	-0.051	6.64E-04	0.029	0.048	8.84E-07
17	17731044	rs11650649	C	rs9899634	0.81	-17.68	-5.45	5.33	-8.53	-0.044	1.62E-03	0.040	0.003	7.53E-08
17	17648854	rs11654081	T	rs11654081	1.00	-17.33	.	3.97	.	-0.053	1.84E-04	0.032	0.025	3.90E-08
17	17665514	rs11656665	G	rs11656665	1.00	-19.13	-4.14	4.11	-8.48	-0.052	1.77E-04	0.036	0.009	9.50E-09
17	17595044	rs11656775	A	rs4925109	1.00	-13.16	.	.	.	-0.065	8.01E-06	0.016	0.271	7.64E-08
17	17778079	rs11656840	T	rs9899634	0.81	-17.62	-5.44	5.32	-8.49	-0.043	2.11E-03	0.041	0.004	1.13E-07
17	17655826	rs11868035	A	rs11654081	0.92	-16.2	.	.	.	-0.053	3.59E-04	0.027	0.069	8.05E-07
17	17808339	rs12600546	G	rs3183702	0.82	-17.65	-5.35	5.14	-8.55	-0.046	1.25E-03	0.038	0.007	1.44E-07
17	17706380	rs12941039	C	rs12941039	1.00	-19.27	-4.97	4.71	-8.93	-0.051	2.22E-04	0.040	0.004	4.29E-09
17	17731313	rs12943500	C	rs1889014	0.89	-17.89	-5.45	.	-8.18	-0.042	2.55E-03	0.038	0.006	3.51E-07
17	17921396	rs12943914	G	rs1889014	0.82	.	-4.81	.	-8.23	-0.046	1.06E-03	0.036	0.010	1.66E-07
17	17693534	rs12951376	T	rs12951376	1.00	-16.96	-3.95	.	-7.36	-0.062	3.41E-05	0.037	0.014	1.17E-09
17	17695984	rs16960744	A	rs16960744	1.00	-19.9	-4.59	4.41	-8.64	-0.053	2.03E-04	0.045	0.001	3.77E-10
17	17707890	rs1889014	C	rs1889014	1.00	-19.1	-4.99	.	-8.6	-0.050	3.33E-04	0.041	0.003	3.46E-09
17	17675465	rs1889018	G	rs1889018	1.00	-19.92	-4.35	4.22	-8.52	-0.051	2.78E-04	0.039	0.005	8.20E-09
17	17688091	rs2236513	A	rs2236513	1.00	-20.03	-4.49	4.14	-8.51	-0.056	8.05E-05	0.043	0.002	2.05E-10
17	17656042	rs2297508	C	rs11654081	1.00	-17.66	-3.82	3.95	.	-0.050	5.11E-04	0.032	0.027	2.70E-07

Supplementary Table 1 (continued)

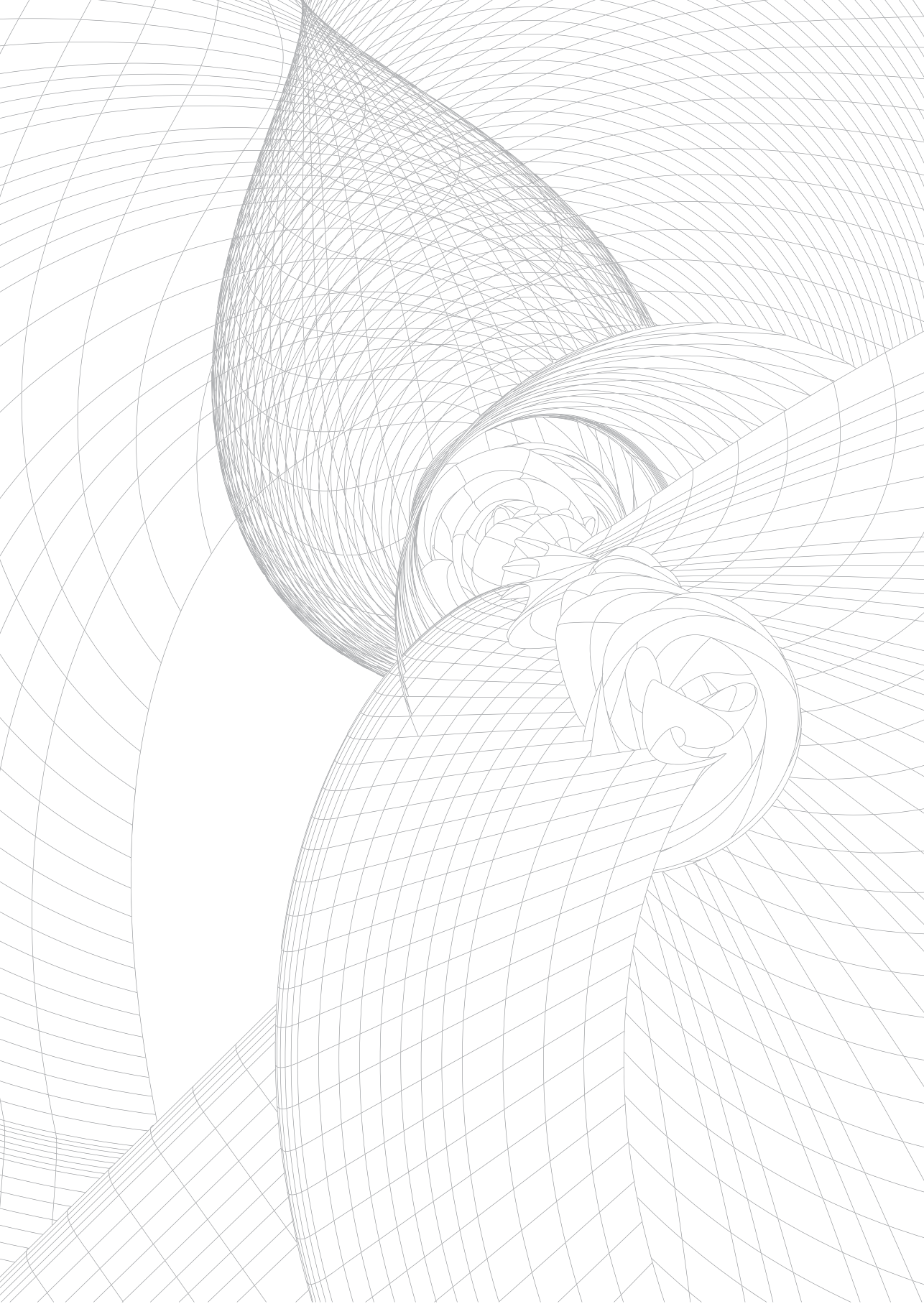
CHR	BP	SNP	A1	GWSass.	r^2	SREBF1	TOM1L2	ATPAF2	C17orf39	β-TBLM	P-TBLM	β-TBBMD	P-BMD	P-bivariate
17	17649254	rs2350976	A	rs11654081	0.96	-17.17	.	4.11	.	-0.046	9.70E-04	0.031	0.024	6.60E-07
17	17743504	rs2350977	C	rs9899634	0.81	-17.64	-5.45	5.35	-8.53	-0.044	1.95E-03	0.041	0.004	9.88E-08
17	17888435	rs2955382	C	rs2955382	1.00	-18.82	-4.75	.	-9.12	-0.053	1.48E-04	0.036	0.008	6.73E-09
17	17688014	rs3183702	A	rs3183702	1.00	-20.2	-4.49	4.14	-8.36	-0.056	8.05E-05	0.043	0.002	2.05E-10
17	17690547	rs3744115	G	rs3744115	1.00	-20.06	-4.53	4.2	-8.57	-0.056	7.05E-05	0.042	0.003	2.86E-10
17	17637256	**rs3803763**	G	rs11654081	0.88	.	.	.	.	.	.	.	.	.
17	17774369	rs4077828	T	rs8079321	0.81	-17.51	-5.42	5.27	-8.55	-0.045	1.50E-03	0.041	0.004	6.46E-08
17	17782181	rs4244602	G	rs9899634	0.81	-17.63	-5.44	5.31	-8.49	-0.043	1.91E-03	0.040	0.004	1.18E-07
17	17797728	rs4257260	G	rs9899634	0.81	-17.63	-5.42	5.3	-8.48	-0.045	1.51E-03	0.038	0.006	1.70E-07
17	17779749	rs4341796	A	rs9899634	0.81	-17.62	-5.44	5.32	-8.49	-0.043	1.91E-03	0.039	0.005	2.03E-07
17	17785595	rs4470201	A	rs11656665	0.85	-17.63	-5.43	5.31	-8.49	-0.046	1.20E-03	0.038	0.007	1.25E-07
17	17705227	rs4924823	A	rs4924823	1.00	-19.3	-4.99	4.7	-8.99	-0.057	6.51E-05	0.040	0.004	5.87E-10
17	17602527	rs4925109	A	rs4925109	1.00	-13.08	.	.	.	-0.064	1.06E-05	0.020	0.163	3.09E-08
17	17651995	rs4925114	A	rs4925114	1.00	-17.18	-3.82	3.94	.	-0.051	3.64E-04	0.036	0.011	2.39E-08
17	17662182	rs4925115	A	rs4925115	1.00	-19.06	-4.11	4.08	-8.46	-0.055	1.18E-04	0.039	0.007	2.37E-09
17	17683629	rs4925119	A	rs4925119	1.00	-19.83	-4.32	4.32	-8.54	-0.056	7.98E-05	0.042	0.003	3.92E-10
17	17695358	rs4925120	C	rs4925120	1.00	-19.98	-4.59	4.33	-8.61	-0.053	2.22E-04	0.046	0.001	3.61E-10
17	17725099	rs4925123	T	rs9899634	0.81	-17.69	-5.45	5.33	-8.53	-0.044	1.62E-03	0.040	0.003	7.53E-08
17	17735169	rs4925125	T	rs4925125	1.00	-17.66	-5.44	5.23	-8.62	-0.048	6.26E-04	0.039	0.006	4.09E-08
17	17739358	rs4925126	A	rs9899634	0.81	-17.64	-5.45	5.35	-8.54	-0.047	7.76E-04	0.039	0.006	5.47E-08
17	17747911	rs4925129	G	rs9899634	0.81	-17.64	-5.45	5.35	-8.53	-0.044	1.45E-03	0.039	0.005	1.34E-07
17	17748583	rs4925130	A	rs9899634	0.81	-17.64	-5.45	5.35	-8.53	-0.044	1.64E-03	0.040	0.003	7.51E-08
17	17773185	rs5002487	G	rs9899634	0.81	-17.61	-5.44	5.33	-8.48	-0.043	1.75E-03	0.040	0.003	8.55E-08

Supplementary Table 1 (continued)

CHR	BP	SNP	A1	GWSass.	r^2	SREBF1	TOM1L2	ATPAF2	C17orf39	β-TBLM	P-TBLM	β-TBBMD	P-BMD	P-bivariate
17	17720429	rs6502619	C	rs9899634	0.81	-17.73	-5.44	5.29	-8.54	-0.044	1.54E-03	0.038	0.006	1.79E-07
17	17775416	rs6502622	G	rs9899634	0.81	-17.63	-5.42	5.33	-8.46	-0.045	1.39E-03	0.038	0.007	1.64E-07
17	17788661	rs6502624	A	rs11656665	0.81	-17.53	-5.41	5.27	-8.55	-0.046	1.13E-03	0.040	0.005	6.26E-08
17	17789011	rs6502625	C	rs6502625	1.00	-17.73	-5.39	.	-8.22	-0.046	8.95E-04	0.039	0.005	4.46E-08
17	17790396	rs6502627	T	rs9899634	0.81	-17.64	-5.44	5.3	-8.49	-0.045	1.48E-03	0.038	0.007	1.78E-07
17	17810367	rs6502629	G	rs1889014	0.89	-17.93	-5.32	.	-8.32	-0.043	1.75E-03	0.039	0.005	1.24E-07
17	17792829	rs7214002	C	rs9899634	0.81	-17.63	-5.43	5.3	-8.49	-0.048	7.10E-04	0.037	0.009	1.15E-07
17	17759542	rs7222480	T	rs9899634	0.81	-16.91	-5.25	4.53	-8.35	-0.043	1.75E-03	0.034	0.012	6.64E-07
17	17786525	rs7224815	T	rs1889014	0.89	-17.88	-5.42	.	-8.11	-0.044	1.41E-03	0.037	0.007	1.48E-07
17	17691632	rs7501812	G	rs7501812	1.00	-20.11	-4.5	4.34	-8.52	-0.056	5.53E-05	0.043	0.002	1.44E-10
17	17720183	rs7503334	C	rs9899634	0.81	-17.75	-5.44	5.28	-8.55	-0.044	1.54E-03	0.038	0.006	1.79E-07
17	17723129	rs8065563	G	rs9899634	0.81	-17.7	-5.43	5.29	-8.47	-0.044	1.54E-03	0.038	0.006	1.79E-07
17	17668768	rs8066560	A	rs8066560	1.00	-19.9	-4.33	4.23	-8.52	-0.056	8.84E-05	0.042	0.003	3.82E-10
17	17745450	rs8070128	C	rs9899634	0.81	-17.34	-5.28	5.39	-8.34	-0.043	1.91E-03	0.039	0.005	2.09E-07
17	17772441	rs8073001	T	rs9899634	0.81	-17.61	-5.44	5.34	-8.48	-0.043	2.11E-03	0.041	0.004	1.13E-07
17	17751976	rs8078138	T	rs9899634	0.81	-17.64	-5.45	5.35	-8.53	-0.044	1.64E-03	0.040	0.003	7.51E-08
17	17752190	rs8078583	T	rs9899634	0.81	-17.64	-5.45	5.35	-8.53	-0.044	1.45E-03	0.039	0.005	1.34E-07
17	17701514	rs8079321	T	rs8079321	1.00	-19.35	-4.97	4.68	-9	-0.054	1.32E-04	0.045	0.002	2.64E-10
17	17717114	rs8080061	C	rs8080061	1.00	-19.15	-5.02	4.92	-9.01	-0.053	1.90E-04	0.042	0.003	1.82E-09
17	17804773	rs8080823	C	rs9899634	0.81	-17.64	-5.42	5.28	-8.5	-0.044	1.59E-03	0.039	0.004	1.07E-07
17	17949827	rs854762	G	rs1889014	0.86	.	.	.	-8.17	-0.044	1.71E-03	0.033	0.017	1.05E-06
17	17952475	rs854764	T	rs1889014	0.86	.	.	.	-8.15	-0.042	2.38E-03	0.032	0.019	2.04E-06
17	17953455	rs854765	T	rs1889014	0.82	.	.	.	-8.16	-0.044	1.45E-03	0.033	0.018	8.80E-07

Supplementary Table 1 (continued)

CHR	BP	SNP	A1	GWSass.	r^2	SREBF1	TOM1L2	ATPAF2	C17orf39	β-TBLM	P-TBLM	β-TBBMD	P-BMD	P-bivariate
17	17944570	rs854813	C	rs1889014	0.86	.	.	.	-8.14	-0.046	9.31E-04	0.033	0.015	3.43E-07
17	17595267	rs941448	C	rs4925109	1.00	-13.17	.	.	.	-0.065	8.01E-06	0.016	0.271	7.64E-08
17	17724473	rs950966	A	rs9899634	0.81	-17.68	-5.43	5.31	-8.48	-0.044	1.40E-03	0.040	0.004	7.53E-08
17	17763900	rs9890341	C	rs9899634	0.81	-17.63	-5.45	5.34	-8.52	-0.043	2.06E-03	0.041	0.004	1.13E-07
17	17685164	rs9891957	G	rs9891957	1.00	-19.98	-4.33	4.2	-8.56	-0.052	1.82E-04	0.043	0.002	8.79E-10
17	17668668	rs9899634	T	rs9899634	1.00	-19.9	-4.33	4.24	-8.52	-0.053	1.93E-04	0.041	0.003	1.69E-09
17	17674485	rs9902941	C	rs9902941	1.00	-19.92	-4.37	4.17	-8.51	-0.057	6.26E-05	0.040	0.005	6.22E-10
17	17711690	rs9907246	T	rs9907246	1.00	-19.12	-5.01	4.89	-8.95	-0.053	1.58E-04	0.039	0.005	3.33E-09
17	17714843	rs9907287	C	rs9907287	1.00	-19.16	-5.02	4.91	-9.01	-0.057	6.15E-05	0.040	0.005	5.71E-10
17	17766745	rs9912895	G	rs9899634	0.81	-17.62	-5.45	5.34	-8.5	-0.045	1.26E-03	0.038	0.008	1.60E-07
17	17688239	rs9915248	C	rs9915248	1.00	-19.98	-4.51	4.16	-8.56	-0.056	6.81E-05	0.042	0.003	3.07E-10



Chapter 5

Evolutionary perspective of bone fragility

BMD Loci Contribute to Ethnic And Developmental Differences in Skeletal Fragility Across Populations: Assessment of Evolutionary Selection Pressures

Carolina Medina-Gómez, Alessandra Chesi, Denise H.M. Heppe, Babette S. Zemel, Jia-Lian Yin, Heidi J. Kalkwarf, Albert Hofman, Joan M. Lappe, Andrea Kelly, Manfred Kayser, Sharon E. Oberfield, Vicente Gilsanz, André G. Uitterlinden, John A. Shepherd, Vincent W.V. Jaddoe, Struan F.A. Grant, Oscar Lao, and Fernando Rivadeneira.

Mol Biol Evol. 2015 Nov;32(11):2961-72. doi: 10.1093/molbev/msv170.

ABSTRACT

Bone mineral density (BMD) is a highly heritable trait used both for the diagnosis of osteoporosis in adults and to assess bone health in children. Ethnic differences in BMD have been documented, with markedly higher levels in individuals of African descent, which partially explain disparity in osteoporosis risk across populations. To date, 63 independent genetic variants have been associated with BMD in adults of Northern-European ancestry. Here, we demonstrate that at least 61 of these variants are predictive of BMD early in life by studying their compound effect within two multiethnic pediatric cohorts. Furthermore, we show that within these cohorts and across populations worldwide the frequency of those alleles associated with increased BMD is systematically elevated in individuals of Sub-Saharan African ancestry. The amount of differentiation in the BMD genetic scores among Sub-Saharan and non-Sub-Saharan populations together with neutrality tests, suggest that these allelic differences are compatible with the hypothesis of selective pressures acting on the genetic determinants of BMD. These findings constitute an explorative contribution to the role of selection on ethnic BMD differences and likely a new example of polygenic adaptation acting on a human trait.

INTRODUCTION

Enormous progress in mapping complex traits has been achieved in the last decade with thousands of loci identified by the implementation of genome-wide association studies (GWAS)[1]. While individually, variants at these loci each make modest contributions, collectively, the effect of large sets of variants increase significantly the amount of explained trait variance. These discoveries have also enabled studying how selective pressures influence the genetic architecture of complex traits across populations. For some phenotypes such as skin pigmentation [2], type 2 diabetes [3], height [4], biliary liver cirrhosis and ulcerative colitis [5], there is increasing evidence that the genetic basis of trait differences across human populations has been shaped by evolutionary selective pressures [6].

Just as in most other medical areas, the field of genetics of osteoporosis has also been revolutionized by the advent of the GWAS approach, with up to 63 independent genetic variants identified as robustly associated with BMD in adults of Northern-European ancestry [7]. BMD measured by Dual-energy X-ray Absorptiometry (DXA) is a highly heritable trait used to diagnose osteoporosis and assess the risk of fracture [8]. BMD levels are determined by the processes of bone accrual (throughout childhood until young adulthood) and bone loss (from peak bone mass acquisition to senescence). Ethnic differences in BMD are well documented and partially explain differences in osteoporosis and fracture risk across populations. Individuals of Sub-Saharan African ancestry tend to have higher BMD levels and lower fracture risk as compared to other populations [9,10].

It is likely that the skeletal system as a whole (and its derived bone strength) has been subjected to selective pressures throughout human evolution [11]. This includes changes brought about through human history ranging from early nomadism to the domestication of animals, crops and even more recently, a tendency to ever increasing levels of sedentariness [12]. In order to understand how selective pressures have shaped bone strength variation across human populations we 1) studied the collective effect of BMD associated variants within two independent multiethnic pediatric cohorts; 2) examined the allele frequency distribution of these variants across diverse ethnic populations using two different catalogues of genetic variation; and 3) assessed the presence of selective pressures by testing for deviation from genetic drift.

RESULTS

Total Body BMD differences across ethnic ancestries

To characterize pediatric ethnic differences in BMD, we selected a sample of 3,994 children (mean age of 6 years) of multi-ethnic background taking part in the Generation R Study [13], who had GWAS data and DXA measurements. First, we determined genetic population substructure from genotyped data, children were assigned to one of the three main genetic ancestry groups (Figure 1): Sub-Saharan African (n=336), European (n=3,499) and East Asian (n=159) descent. Table 1 summarizes the characteristics of the Generation R participants under study. Children of European and East Asian ancestry had respectively, 4% (0.552 g/cm^2, P<2x10^{-16}) and 2.8% (0.559 g/cm^2, P=7x10^{-6}) lower BMD than children who clustered in the Sub-Saharan African group (0.575 g/cm^2), after controlling for age, sex, height, fat and lean mass. These differences remained statistically significant after correction for lifestyle factors affecting mothers and children. Furthermore, we observed that, in European children, the proportion of genetic Sub-Saharan African ancestry in their genomes (range 0.00-0.49) was positively associated with BMD levels, representing a BMD increase of 0.0003 g/cm^2 per increase in genomic Sub-Saharan African component (beta= 0.72, P=1.87x10^{-5}). Thus, the BMD difference between an individual with 0% Sub-Saharan

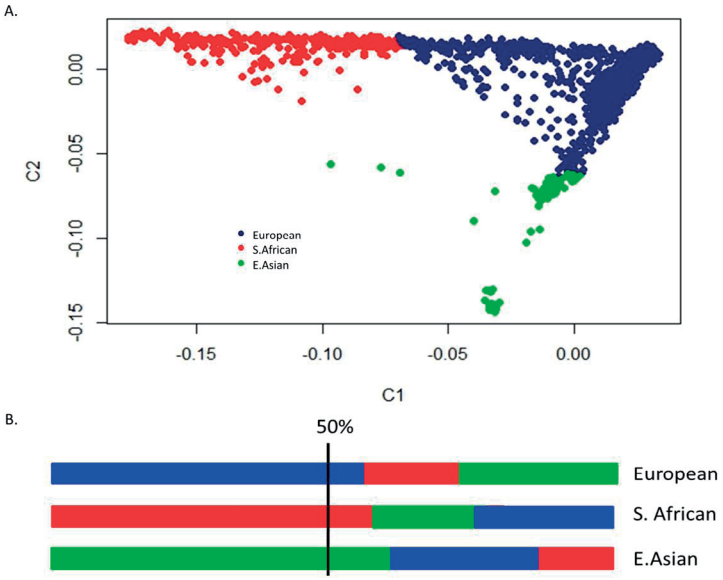


Figure 1. Genetic substructure of the Generation R Study. A: Two-dimensional plots from multidimensional scaling analyses of the Generation R Study based on the first two genomic principal components**. B.:** Clustering rule based on ancestral proportions. Children were assigned to one of the three main genetic ancestry groups, based on their highest fraction of estimated ancestry (i.e., >0.50) proportion: European (blue), Sub-Saharan African (red) and East Asian (green). Individuals with no ancestry proportion >0.50 were excluded.

Table 1. Participant characteristics in the Generation R Study and BMD-Childhood Study by defined genetic ancestry

	Generation R Study (The Netherlands)								BMD Childhood Study (U.S.A.)							
	S. African		European		East Asian		All		S. African		European		East Asian		All	
	n=336		n=3,499		n=159		n=3,994		n=315		n=1,277		n=78		n=1,670	
Age, years	6.41	0.68	6.15	0.46	6.20	0.42	6.18	0.50	11.45	4.46	11.4	4.41	12.27	4.68	11.46	4.43
Women (n, %)	171	50.9	1,747	50.1	72	54.7	2001	50.1	165	52.4	666	52.2	35	44.9	866	51.9
Height, m	1.219	0.07	1.195	0.06	1.177	0.06	1.196	0.06	1.452	0.21	1.446	0.22	1.455	0.21	1.447	0.22
Weight, kg	24.48	4.99	22.94	3.87	21.64	4.50	23.06	4.05	43.31	18.56	42.24	19.02	43.6	18.28	42.5	18.9
Fat Mass, kg	6.397	2.90	5.787	2.21	5.947	2.71	5.844	2.30	9.88	6.00	10.11	5.87	10.69	6.20	10.10	5.90
Lean Mass, kg	17.67	2.80	16.33	2.20	14.92	2.38	16.39	2.28	32.05	13.90	31.06	14.40	30.99	13.50	31.24	14.30
BMD, g/cm^2	0.595	0.06	0.551	0.046	0.541	0.055	0.555	0.05	0.828	0.198	0.775	0.192	0.798	0.199	0.786	0.195

Data are means and SD (except for women number and percentage)

S.African: Sub-Saharan African

African ancestry and one with 49.99% Sub-Saharan African ancestry in their genomes will be equivalent to 0.016 g/cm^2, suggesting that genetic factors play a role in the observed ethnic BMD differences.

We confirmed the existence of these ethnic differences in BMD in an independent multi-ethnic set of 1,670 U.S. children from the Bone Mineral Density in Childhood Study (BMDCS), using the same ethnic definition and analytical approach. Characteristics of the BMDCS participants under study are presented in Table 1. Consistent with our findings, after adjustments children of European (n=1,277) and East Asian (n=78) ancestry had 4.7% (0.778 g/cm^2, P=0.0002) and 2.5% (0.794 g/cm^2, P<2x10^{-16}) lower BMD levels than children of Sub-Saharan African ancestry (n= 315) with mean BMD 0.816 g/cm^2, respectively.

Bone Mineral Density Genetic Score (BMD-GS) association with pediatric BMD

We constructed a genetic score –composed of BMD-increasing alleles- for each child using single nucleotide polymorphisms (SNPs) known to be associated with BMD in adults of Northern-European ancestry [7] (Online Resource Table S1). In the Generation R Study, the calculated BMD-GS explained 4.5% of the BMD variance in all the sampled children (P<2x10^{-16}). Additional correction by ten genomic principal components, addressing potential residual population stratification, indicated that up to 2.6% of BMD variation is explained by the BMD-GS (P<2x10^{-16}). In children from the BMDCS, the BMD-GS explained 5.7% and 2.2% of the BMD variance, before and after correction by ten principal components (Online Resource Table S2). The BMD-GS was positively associated with BMD in children of predominantly European and African descent of both studies, albeit the latter reached nominal significance only in the BMDCS cohort. In children of predominantly East Asian ancestry (with the lowest sample size) the effect of the score on BMD did not reach statistical significance in either study. Joint analysis of the Generation R and BMDCS resulted in significant associations of BMD with the BMD-GS in children of European (beta= 4.43, P<2x10^{-16}) and Sub-Saharan African (beta=2.41, P=0.016) background, but not in those of East Asian (beta=1.18, P=0.472) descent.

We then examined BMD levels across quintiles of the calculated BMD-GS in the two pediatric cohorts. There was a positive correlation between the score and the adjusted BMD in both studies. In the Generation R Study, as compared to children in the middle quintile (55.8% of the population with the mean BMD; n=2,232), children in the highest quintile (2.2% of the population; n=88) had 0.72 SDs higher BMD (P<1x10^{-6}), while those in the lowest quintile (1.3% of the population; n=53) had 0.53 SDs (P=4x10^{-6}) lower BMD. Similar results were obtained in the BMDCS study where children in the highest quintile (1.1% of the population; n=19) had 0.42 SDs higher BMD (P=0.06) and those in the lowest quintile (2.9%

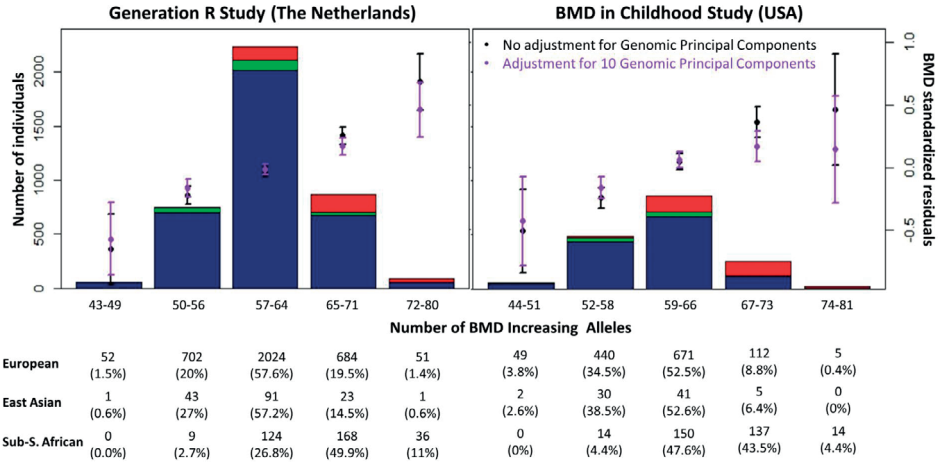


Figure 2. Distribution of the BMD-increasing allele score by genetic ancestry in the Generation R Study and BMD in Childhood Study. The increasing-BMD GS has been divided in 5 category bins. Colors in the stacked bars represent ethnic background: European (blue), East Asian (green) and Sub-Saharan African (red). Black dots represent the mean BMD adjusted for age, gender, height, fat and lean mass. Magenta dots represent the mean BMD per bin, adjusted for all former variables plus the first 10 principal components.

of the population; n=49) 0.55 SDs lower BMD (P=0.001) as compared to the children in the middle quintile (51.3% of the population, n=860).

As depicted in Figure 2, the overrepresentation of children of Sub-Saharan African descent across the higher quintiles of the BMD-GS was highly significant in both studies (Generation R: $P=2.7x10^{-71}$ and BMDCS, $P=2.3x10^{-70}$). In the Generation R Study 60.7% of children of Sub-Saharan African ancestry are found in the two highest GS quintiles, as compared to children of European (20.8%) and East Asian (15.1%) ancestry. For the BMDCS, 47.9% of children of Sub-Saharan African ancestry are found in the two highest GS quintiles, as compared to children of European (9.2 %) and East Asian (6.4%) ancestry. In both pediatric cohorts the mean BMD-GS values were highest in the Sub-Saharan African group (Generation R: 0.537, BMDCS: 0.542), while remaining similar in the European (Generation R: 0.495, BMDCS: 0.493) and East Asian (Generation R: 0.487, BMDCS: 0.491) groups.

Computing BMD-increasing allele scores using additional sets of markers associated with BMD at lower levels of significance showed that the greatest difference in mean GS values between subjects of African and European background was observed with the original BMD-GS including 61 markers (Online Resource Figure S1). The addition of more (less significantly-associated) SNPs to the score decreased the mean allele score difference observed between the two ethnic groups and did not improve significantly the BMD explained variance.

Worldwide geographical distribution of the BMD-increasing alleles

We then sought to understand the observed ethnic mean differences in BMD-GS by decomposing it and considering each of the BMD-increasing alleles independently. In the Generation R Study, where the sample size allows more accurate statistical inferences, we found that in children of Sub-Saharan African ancestry 23 of these alleles had ≥10% higher frequency as compared to only 12 in Europeans (P=0.04). When examining HapMap data [14], similar allele frequency differences were observed between the CEU and YRI populations, independent of the CEU SNP minor allele frequency (MAF) (Online Resource Table S1).

In order to broaden the scope of our findings, we analyzed the spatial distribution of the BMD-GS SNPs worldwide, using the populations from the Centre D'Etude du Polymorphism Humaine-Human Genome Diversity Project (CEPH-HGDP) panel [15] (Online Resource Table S3). As shown on Figure 3, examining the BMD-GS across all individuals

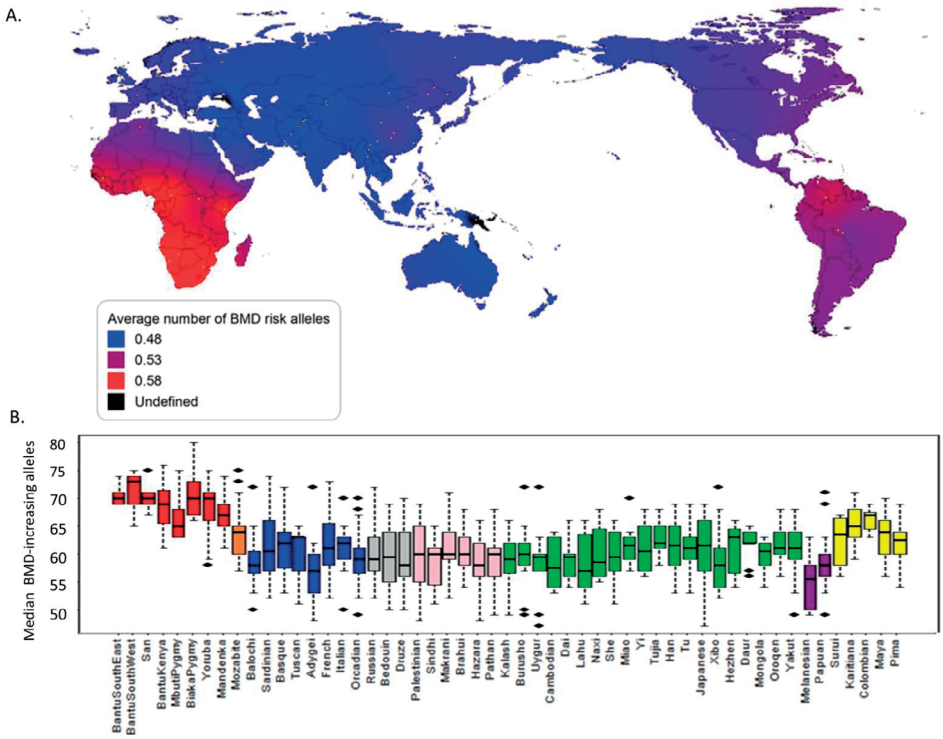


Figure 3. Worldwide geospatial distribution of BMD-increasing alleles. A. The density map of the combined frequency of the BMD GS was generated by inverse to distance interpolation using the observed values across the 53 HGDP populations. **B.** Total number of BMD-increasing alleles per individual in the 53 populations of the HGDP. Populations are ordered by geographic groups: Sub-Saharan Africa (red), North-Africa (orange), Europe (blue), Middle East (grey), Central Asia (pink), East Asia (green), Oceania (purple), and Native Americans (yellow).

from the different populations, we confirmed that Sub-Saharan African populations had the highest number of BMD-increasing alleles, together with other two Latin-American populations located in the tropics. Likewise, the populations out of Africa in general had lower BMD-GS scores.

Signatures of natural selection in the BMD-GS SNPs and associated regions

We followed sixth different lines of evidence to discard that genetic drift was explaining the observed spatial distribution of the mean BMD-GS (Figure 3). First, we observed that the BMD-increasing alleles were enriched for ancestral alleles (73% actually constitute ancestral alleles; $P=1.8x10^{-4}$). Second, taking advantage of the geographical coverage of the HGDP, we quantified the differentiation between continents, where for each HGDP individual we generated 100,000 permuted scores by randomizing the coded effect allele in each of the BMD-GS variants. The BMD-GS differentiation between Sub-Saharan African and non-Sub-Saharan African populations was higher for the original BMD-GS (P=0.025) than for the different sets of sampled scores. Third, in line with the previous test, we evaluated the population specific departure of the BMD-GS score from its expected genomic value and determined that the proportion of variance attributable to the categorization into Sub-Saharan African and non-Sub-Saharan African populations, was significantly higher (P=0.011) for the original BMD-GS than that observed in 100,000 GS sets of 61 SNPs randomly sampled from the genome (after matching for factors related to GWAS discovery bias). These results provide additional statistical support to the contention that the observed differentiation of Sub-Saharan African populations, as shown by the multidimensional scaling (MDS) and ADMIXTURE analyses using only the markers of the BMD-GS (Online Resource Figure S2), is not due to genetic drift. Fourth, we estimated the Q_X statistic[6], which analyzes extreme patterns of genetic differentiation among populations, which was not significant (Qx = 46.67, P= 0.68). Fifth, to assess if selective pressures acting on the score explained these results, we scrutinized the YRI, CHB and CEU populations and found no signals of strong selective sweeps acting on the loci harboring the BMD-GS SNPs (Online Resource Table S4). Nonetheless, we did find evidence for departure by chance from neutrality in CEU and CHB (Hierarchical Bayesian binomial modeling, and Online Resource Table S5); resulting in a possible excess of intermediate-frequency alleles in the BMD-GS. Sixth, we applied tests for differences in mean MAF and determined that as compared to SNP sets randomly ascertained from the GWAS catalogue, the SNPs of the BMD-GS have significantly higher ($P=8.3x10^{-3}$) mean MAF in European (MAF=0.30 in CEU), but not in East Asian (MAF=0.26; P=0.13 in JPT+CHB) or African (MAF=0.21; P=0.83 in YRI) populations. Altogether, five out of sixth different lines of evidence argue against genetic drift explaining the observed ethnic differences in mean BMD-GS values.

DISCUSSION

In this study we established that ethnic differences in BMD are already present early in life and are partially explained by variation in BMD-associated loci. Furthermore, we demonstrate in two independent pediatric multiethnic cohorts that those alleles positively associated with BMD are systematically elevated in populations of Sub-Saharan African ancestry. The same pattern was also observed in a broader framework of catalogues of human genetic variation. Five out of six tests indicate deviation from neutrality and are suggestive of polygenic adaptation acting on the loci harboring the variants of the score. Our results confirm that the Sub-Saharan African versus non-Sub-Saharan African distribution of the BMD increasing alleles worldwide cannot be simply explained by demographic factors or genetic drift.

We assessed the existence of ethnic differences in BMD levels first within the Generation R Study. In contrast to ethnic comparisons conducted across different geographical areas [16-18], these research subjects were relatively standardized as they were all residents of the city of Rotterdam, The Netherlands; measured using the same DXA device; and within a similar age range. This study design reduces the impact of inter-machine variance, as well as local environmental and lifestyle exposures influencing BMD variation such as major differences in access to medical care, sunlight exposure, diet or physical activity. Children of Sub-Saharan African ancestry showed higher BMD in this set even after correction for lifestyle factors affecting both mothers and children. Nevertheless, since each child might have a varying degree of each of the three ancestral populations in their genome (Figure 1), actual ethnic BMD differences might be even higher than reported here when more stringent ethnic definitions are used. These differences found in the Generation R study and replicated in the BMDCS, together with the positive association of Sub-Saharan African genomic ancestry proportion and BMD, suggests that genetic factors play a role in the observed phenotypic differences in BMD.

The fact that the BMD-GS was significantly associated with BMD in children from both the Generation R and BMDCS (Figure 2), implicate developmental effects of the studied variants acting on the bone accrual process. The systematic higher frequency of the BMD increasing-alleles in children of Sub-Saharan ancestry mirrored their higher BMD, advocating that ethnic differences in BMD are to a given extent genetic in origin and not likely due to population stratification (enduring correction for genomic principal components).

Our results indicate that the BMD-GS is more predictive in subjects of European descent (where variants were discovered) than in other ethnic groups. Nevertheless, differences in the effect of the score across ethnic populations in our pediatric studies could simply

reflect greater uncertainty in estimating effect sizes of the BMD-SNPs for these populations given their small sample size. Despite not finding an association in children of East Asian background, previous studies in East Asian populations have replicated associations with BMD for several variants of the genetic score [19-21]. In fact, high trans ethnic replicability has been described between these populations for other complex traits [22]. Additionally, ethnic differences in the degree of association of the BMD-GS can also reflect population differences in the relationship between tagging and true underlying genetic variants [22]. This is especially evident in African populations where blocks of LD are the shortest and associations with phenotypes can result in dilution of effects [23]. Therefore, the true directional population differentiation at causal SNPs could be even larger than the directional population differentiation observed in tagging SNPs, likely comprising a large fraction of the variants in the GS we employed. Further, other variants more frequent in African or Asian populations may contain additional set of alleles contributing to BMD variation in those groups; this contention is also supported by the significant association of the fraction of Sub-Saharan African ancestry in Europeans, which remained significant independently of the BMD-GS.

In the GEnetic Factors of Osteoporosis (GEFOS) BMD meta-analysis a weighted version of the BMD-GS explained 5.8% of the femoral neck BMD variance in a cohort of unrelated postmenopausal Danish women [7]. The BMD-GS explained relatively lower variance in our two pediatric multiethnic cohorts (Generation R: 2.6%, BMDCS: 2.2%), likely as consequence of the different weighting scheme, age range [24], skeletal site specificity [25] or ethnic context. The use of a weighted scores could increase predictive ability but also introduce possible bias due to the above mentioned differences in study setting as well as residual population stratification intrinsic in GWAS studies [4].

By interrogating publicly available databases we could establish that the results observed in the two pediatric cohorts were paralleled in a worldwide scenario (Figure 3). Frequency differences in the Generation R Study reflected the differences observed in the HapMap YRI and CEU populations [14]. The magnitude of these differences was such, that by employing only the BMD-GS SNPs we could unequivocally separate Sub-Saharan African individuals from all others in the populations from the HGDP-CEPH panel (Online Resource Figure S2). This geographic pattern is in line with the geographical disparities in fracture risk worldwide and parallel a potential association with latitude [26-29]. However, we did not find any clear BMD-GS south to north gradient among the HGDP-CEPH populations. Yet, an association with latitude may be attenuated as the populations part of the HGDP-CEPH panel are unevenly distributed [30].

Likewise, careful consideration is needed when drawing conclusions about findings in African populations, as Africa is a continent characterized by profound genetic diversity and home to a vast array of heterogeneous lifestyles [31]. Given the limited representation of African populations in the HGDP-CEPH and HapMap datasets, the association between the BMD-GS and Sub-Saharan African ancestry is not to be directly extrapolated across the whole African continent. Nevertheless, our findings suggest that a fraction of the population differences in BMD levels are genetic in origin, without disregarding the large environmental influences contributing to trait variation and ethnic differences.

The overrepresentation of ancestral alleles in the BMD-GS, indicates that phenotypic states related to increased BMD, such as bone robustness [11], represent the ancestral state in humans and show an amount of phenotypic differentiation that cannot be uniquely explained by demographic factors. A similar result was observed by Wu and Zhang [32] when analyzing genes that had been associated with the skeletal system in humans. The authors concluded that such excess of derived variants in populations from European and East Asian ancestry was suggestive of selective pressures out of the African continent, although this could also be consequence of a relaxed constraint. We observed that the BMD-GS differentiation between Sub-Saharan African and non-Sub-Saharan African populations was significantly larger than expected under neutrality for both the randomization test of allele effects and when compared with the rest of the genome. In fact, departure of the BMD-GS from neutrality was supported by five of six different types of tests, including: 1) Detection of enrichment of ancestral alleles in the set of BMD increasing alleles in the BMD-GS; 2) Randomization of the BMD increasing allele and quantification of the differentiation between continents; 3) Population departure of the expected genomic value of BMD-GS; 4) Detection of departure of neutrality in sequence based tests; and 5) Presence of an excess of intermediate allele frequencies. We acknowledge that the statistical evidence derived from these five tests was marginal. However, such magnitude of statistical significance is in agreement with those reported for other complex phenotypes[5,33]. Only on the 6) Qx index test [6] we did not find any evidence for departure from neutrality.

This discrepancy could be due to the wide range of hypotheses about differentiation considered by the Qx, which models all different population levels. In contrast, the other tests (i.e., ANOVA) were motivated by the empirical observation that the BMD-GS is increased in Sub-Saharan populations and applied specifically to detect differences between two groups (Sub-Saharan vs. non-Sub-Saharan African populations). Our results, (Online Resource Table S4) as expected under polygenic adaptation, showed no enrichment for strong selective sweeps in the BMD-associated loci but rather displaying small allele frequency shifts between populations[34]. One plausible explanation is that methods for detecting strong selective sweeps are under powered in regions containing a relative excess

of intermediate-frequency alleles compared to the rest of the genome. We observed that in the European populations the BMD-GS SNPs have a systematic excess of intermediate-frequency alleles as compared to that observed in other SNPs reported in the GWAS catalogue[35]. This footprint of intermediate-frequency alleles in the BMD-GS SNPs can be produced by the presence of selective pressures increasing the genetic variability, such as balancing or directional selection on standing variation[36] or by SNP ascertainment bias on the GWAS discoveries[37]. Nevertheless, since our sampling SNPs sets were drawn from the GWAS catalogue, and controlled for multiple putative confounder factors, ascertainment bias is an unlikely explanation.

The skeletal system as a whole and its derived bone strength are likely to have been subject to natural selection during human evolution[11]. Multiple factors may have been responsible for selective pressures such as: long distance running and trekking (indispensable for hunting and scavenging), the establishment of a more sedentary lifestyle (with the advent of agriculture), which would have favored skeletal gracility [11,12]; adaptation to less sunlight at more distant latitudes from the equator (i.e., skin color and vitamin D metabolism) or differences in milk/calcium intake (i.e., lactase persistence), among several others. Therefore, it is conceivable that evolution would favor an optimal balance between bone thickness and bone strength within the context of the structural and metabolic functions that the skeleton must serve. However, the ultimate reason for the presence of this genetic signature *out of Africa*, remains to be elucidated. Understanding of the evolutionary and genetic mechanisms shaping BMD variation and peak bone mass acquisition across populations will facilitate pinpointing critical factors underlying ethnic differences in the risk of osteoporosis later in life.

In summary, we provide evidence that disparities in global patterns of BMD-increasing allele frequencies contribute to disparities in BMD levels across different ethnicities, effects which are already present at early ages, postulating a critical role of these variants in the developmental process of peak bone mass accrual. We also show that these allelic differences are compatible with the hypothesis of selective pressures acting on the genetic determinants of BMD. These findings constitute an explorative contribution to the role of selection on ethnic BMD differences and likely a new example of polygenic adaptation acting on a human trait.

MATERIALS AND METHODS

Study Populations

The Generation R Study

The Generation R Study is a population-based prospective cohort study from foetal life onwards in Rotterdam, the Netherlands [13]. Data from a total of 3,994 children with blood collected at birth, subsequently genotyped, and valid DXA scans at 6 years of age, was included in these analyses. Participants underwent DXA measurements with a GE-Lunar iDXA device (GE Healthcare Lunar, Madison, WI) following standard manufacturer protocols. DNA samples were genotyped either on the Illumina HumanHap 610 or Illumina HumanHap 660 chip. Duplicated samples or with excess of heterozygocity and gender mismatches were excluded from the dataset. SNPs with a MAF < 1%, call rate<98% or out of Hardy-Weinberg equilibrium ($P<10^{-6}$) were removed from further analyses. Imputations to the combined HapMap Phase II Build 36 Release 22 panel[14] , composed of dense genotypes from four populations: Utah residents of European Ancestry (CEU), Yoruba people from Ibadan, Nigeria(YRI), Han Chinese individuals from Beijing China and a Japanese population from the Tokyo area, Japan (CHB+JPT), were performed following a two-step procedure as implemented in the MACH/minimac suit.

Bone Mineral Density in Childhood Study (BMDCS)

The Bone Mineral Density in Childhood Study (BMDCS) is a longitudinal multicenter study in the U.S.A. designed with the goal of obtaining standard pediatric reference data for bone mineral density assessed using DXA[17,38]. In total, 2,521 boys and girls aged 5 to 20 years old were recruited between 2002 and 2007 and DXA measurements were obtained annually at five clinical centers in the United States for up to 7 measurements. Another 507 Caucasian children aged 5 to 18 year old were subsequently enrolled for a one-time visit in two of the five centers. Enrollment criteria were established to identify research subjects with normal development, including healthy bones. DXA scans were obtained using Hologic, Inc. (Bedford, MA) bone densitometers. The baseline measurements for 1,670 individuals were used for this analysis. All samples were genotyped on the HumanOmniExpressExome-8v1 chip. Samples with gender discrepancy, low genotype quality, sample replicates and siblings were excluded from the analysis. SNPs with MAF < 0.5% and call rate of < 95% were removed. Imputation was performed following a two-step procedure; haplotype phasing was carried out using ShapeIT while imputation to the combined HapMap Phase II Build 36 Release 22 panel was performed with Imputev2.

Centre D'Etude du Polymorphism Humaine-Human Genome Diversity Project panel (HGDP-CEPH)

The HGDP-CEPH is a collection of DNA samples from lymphoblastoid cell lines representing 1,064 individuals sampled from 53 populations throughout the world [15]. We downloaded Illumina *650Y* data for 1,043 samples from the HGDP-CEPH (http://www.hagsc.org/hgdp/files.html, last accessed February 2, 2014). After excluding duplicates, ethnical outliers, first or second degree relatives [39] and individuals with more than 2% missing SNPs, we selected a panel comprising 940 individuals (Online Resource Table S3). SNPs were removed if the MAF was less than 1%, call rate<98% or if they were out of Hardy-Weinberg equilibrium ($P<10^{-6}$). All samples were imputed to the combined HapMap Phase II Build 36 Release 22 panel. Imputations were performed following a two-step procedure as proposed by the MACH/minimac suit.

Individual BMD estimation

Total body bone mineral density was measured in the two population-based studies (Generation R Study and the BMDCS) using DXA scans. In both studies measurements were conducted by well-trained research assistants and daily quality control assurance was performed. Prior to the scan procedure, participants were asked to take off their shoes, heavy clothes and metal accessories. As recommended by the International Society for Clinical Densitometry [40], total body less head BMD was the measurement used in the analysis, as both studies are embedded in pediatric populations.

BMD associated Single Nucleotide Polymorphisms

For this study, we used 61 out of the 63 autosomal SNPs reported as genome wide associated with either Femoral Neck or Lumbar Spine BMD ($P\leq5x10^{-8}$) in the large scale GWAS meta-analyses of BMD from the GEFOS[7] successfully imputed in the different study populations. Description of these SNPs can be found in the Online Resource Table S1. Comparison of BMD-increasing allele frequency was completed based on the Generation R children of Sub Saharan African and European ancestry and on the YRI and CEU HapMap project[41] populations, in dbSNP [42].

Estimation of genomic principal components and genetic ancestry determination

Ancestry determination was performed in each of the studies, Generation R and BMDCS, separately but following the same protocol. Autosomal genotyped SNPs of all samples were pruned, so that no pair of SNPs within a window of 200 markers were in LD ($r^2=0.05$). Based on these ~35,000 SNPs we described the genetic ancestry of the whole dataset by means of classical MDS in PLINK, generating 10 genomic principal components. We then performed estimation of the genetic ancestry components of each individual on basis of the maximum likelihood using ADMIXTURE software [43]. This program models the probability of observed

genotypes using ancestry proportions and ancestral population allele frequencies. The clustering method was set to group individuals in 3 ancestral populations (K = 3), corresponding to the expected main Sub-Saharan African, European and East Asian ancestry components. Children were assigned to one of the three ancestry groups, labeled after the HapMap Phase II populations, based on their highest fraction of estimated ancestry (i.e., >0.50) proportions. In case none of the 3 ancestral populations reached this proportion, the participant was excluded from further analyses. Ancestry determination was also calculated for the HGDP individuals based exclusively in the 61 BMD-associated variants and assuming K = 3 ancestral proportions. Furthermore, we computed genomic principal components using the MDS routine and clustered the individuals with the R package *Mclust* (http://cran.r-project.org/web/packages/mclust/index.html, last accessed April 10, 2014) This algorithm assigns individuals to clusters by fitting multivariate normal distributions using the coordinates of the proposed dimensions and put forward the best clustering based on the Bayesian Information Criterion (BIC).

BMD Genetic Score calculation

Scores were obtained for each individual as an un-weighted sum across the 61 BMD-associated SNPs of the number of putative increaser alleles (0, 1 or 2) using the profile scoring routine in PLINK. BMD-GS calculations were performed for each population independently based on best guess genotypes obtained using GCTA software[44] after imputation and quality control of the imputed dataset. Extensions of the score to include SNPs associated at the meta-analysis level with BMD at different significance thresholds were also generated. We used the PLINK clumping function, following a strategy described elsewhere[4]. Briefly, we combined publicly available results of the stage 1 lumbar spine and femoral neck BMD meta-analyses [7], data publicly available (http://www.gefos.org/?q=content/data-release, last accessed March 20, 2015). We assigned to each SNP the lowest p-value reached genome-wide in any of the two association analyses by skeletal site. Then we clumped the dataset with an $r^2 \geq 0.1$ using as reference the most BMD-associated SNPs and pruned remaining SNPs within 0.5 Mb of each other. Using different thresholds of significance ($5x10^{-8}$, $5x10^{-7}$, $5x10^{-6}$, $5x10^{-5}$, $5x10^{-4}$, $5x10^{-3}$, 0.05 and 0.1) we generated eight different scores. Each cumulative threshold set contains the SNPs included in sets defined by previous thresholds (e.g., the set with threshold $5x10^{-6}$ includes all the SNPs with a p-value smaller than $5x10^{-6}$, comprising those at p-value thresholds of $5x10^{-8}$, $5x10^{-7}$, etc). For each score we calculated both the BMD explained variance and the mean difference of BMD increasing alleles between children of Sub-Saharan and those of European ancestry. Some differences between our score and the one including only GWS SNPs apply. The GEFOS meta-analysis involved a two-stage meta-analysis. In the first stage of the meta-analysis (discovery), only GWAS studies were included. For the second stage (replication), follow-up was pursued for only 96 SNPs. Since the genome-wide summary statistics are confined to the discovery setting,

only 40 SNPs are classified as associated at GWS level, with the rest achieving genome-wide significance in the meta-analysis of both stages.

Estimation of the genetic ancestry association with pediatric BMD

The effect of genetic ancestry on BMD was first evaluated by least-squares means using the R package *lsmeans* (http://cran.r-project.org/web/packages/lsmeans/index.html, last accessed April 10, 2014). BMD was adjusted for age, gender, height, fat and lean mass index (which mimic the equation of BMI by dividing the referred mass by the squared height) for all participants independently in both pediatric studies before comparison. Moreover, in a subsample of the Generation R Study (n= 1,750), a model that additionally included birth weight, hours of physical activity of the children at age 5, age and height of the mother right before pregnancy, smoking during pregnancy and breastfeeding was assessed. Furthermore, we evaluated the effect of Sub Saharan African and East Asian genetic ancestry proportions on adjusted BMD. For each individual genetically identified as European in the Generation R Study, the percentage of genetic ancestry estimated by ADMIXTURE [43] for either Sub-Saharan African or East Asian genetic ancestry was included as variable in the basic model.

Estimation of the BMD- GS association with pediatric BMD

BMD adjusted variability explained by the GS was evaluated based on the linear regression coefficient of determination for both multi-ethnic populations, namely the Generation R and the BMDCS, for the whole populations and by ethnic group. Analyses were further adjusted for 10 genomic principal components, when specified. Specific sets of principal components were generated in each of the two pediatric studies for the analyses within each of the ethnic clusters. Then results were pooled together. To quantify the magnitude of BMD increase across the genetic score in each study, we divided the obtained BMD-GS in quintiles. In each score-bin we calculated the mean adjusted BMD. Comparisons among quintiles were made with reference to the middle quintile using t-tests, and p-values were corrected for multiple testing. Deviation from the observed ethnic distribution across quintiles was evaluated using the Chi squared distribution, comparing the observed ethnic distribution with the one we would expect if the BMD-GS would be randomly distributed.

Worldwide geographic distribution of the BMD-increasing alleles

The BMD-GS was calculated for each individual across the 53 HGDP populations and box plots per population were generated. To visualize the spatial distribution of the GS alleles, we created a density map with the median BMD-GS value observed at each population using MapViewer 7.1.1767. Point interpolation of the locations where no data was available was performed by means of inverse to distance.

Signatures of natural selection in the 61 BMD SNPs and associated regions

In order to test whether the observed BMD-GS score differentiation pattern among populations was explained by genetic drift, we performed six different analyses:

1) Detection of enrichment of ancestral alleles in the set of BMD increasing alleles

Estimates of ancestral state were determined from the dbSNP database using BioMart [45]. Assessment of ancestral alleles enrichment in the BMD-GS was performed comparing the actual distribution of ancestral status to a Bernoulli distribution expected under random events. We based this calculation on 55 SNPs only, whose alleles were not palindromic (A/T or G/C) in order to avoid strand bias.

2) Randomization of the BMD increasing allele and quantification of the differentiation between continents by means of a two-way nested ANOVA design.

Each individual of the HGDP panel was hierarchically classified according to its sampling population, and the populations classified either to the Sub-Saharan African or non-Sub-Saharan African category. The amount of BMD-GS variation between the two ethnic categories was computed by means of a two-way nested ANOVA comprising three levels including 1) categories: Sub-Saharan and non-Sub-Saharan categories; 2) populations within categories; and 3) individuals within populations. In order to test whether the observed percentage of explained variation between these two categories was due to genetic drift, 100,000 permuted scores were generated, in which the sign of the effect for each of the 61 SNPs was randomly assigned to one of the two alleles. For each permuted dataset, we have constructed a new individual GS and determined the GS difference between African and non-African groups. A one-tail p-value was estimated as the number of times that the explained variance (SS_{among}/SS_{total}) of the sampled set was greater than the one obtained for the original BMD-GS (variance: 0.225).

3) Population departure of the expected genomic value of BMD-GS scores estimated by means of a two-way nested ANOVA design

We applied an empirical genomic-based approach in order to obtain the expected null genomic distribution of the BMD-GS differentiation among the Sub-Saharan and non-Sub-Saharan populations under the assumption of genetic drift. Under this contention we randomly sampled 100,000 datasets of 61 SNPs across the genome showing similar genomic characteristics as the BMD-GS SNPs. Following the SNP matching framework described by Berg and Coop[6], each BMD-GS SNP was matched by the imputed status at HGDP-CEPH dataset, the B-value[46] and the allelic frequency of each of the BMD-increasing alleles in a proxy population (Orcadians) to the one where the SNPs were initially associated to the BMD phenotype (Europeans). A genomic SNP was considered as putative proxy for a given BMD-SNP if it was at least >100Kb far away from the physical position of the BMD-SNP, the

B statistic difference was smaller than 0.05 (calculated by http://www.phrap.org/other-software.html, last accessed February 27, 2015) and had a similar allelic frequency (<0.05) of the ancestral allele in the Orcadian population. At each iteration, 61 SNPs were sampled at random from the pool of proxy SNPs and a new genetic scores was computed for each individual. A two-way nested ANOVA framework (described above) was used to compare the differentiation between Sub-Saharan and non-Sub-Saharan categories. One tail p-value was computed comparing the number of times that the explained variance (SS_{among}/SS_{total}) of the sampled set was bigger than the explained variance of the original BMD-GS.

4) Q_X statistic calculation

The Q_X statistic proposed by Berg and Coop[6], represents a measure of the among population variance in estimated genetic values that is not explained by drift and shared history. We assessed the covariance matrix matching by SNP ascertainment characteristics including imputed status, B-value and Orcadian allele frequency, as described above. In order to construct the null covariance F matrix, we sampled a maximum of 200 proxy SNPs for each BMD-GS SNP at a physical distance of at least 100Kb to ensure pairwise independence (total number of SNPs = 11,655). We computed the covariance F matrix and Q_X statistic as described [6], using formulas 17 and 10, respectively).

5) Detection of departure of neutrality in sequence based neutrality tests

We used the 1000 Genomes project selection browser[47] to detect signatures of positive selection in the 61 genomic regions harboring BMD-GS SNPs in the CEU, CHB and YRI populations. For each population, nine different summary statistics incorporated in the browser were used to evaluate the 61 regions simultaneously partitioning across windows of an average size of 30kb: CLR, XPCLR of PopA vs PopB and XPCLR of PopA vs PopC, Fay and Wu's H, Fu and Li's D, Fu and Li's F, R2, Tajima's D and Wall's B. Hierarchical Bayesian binomial model using a logit transformation, as implemented in Winbugs[48], was used to estimate the distribution of the frequency of statistically significant neutrality tests at 5% and asses if the 95% Credible Interval (CI) included the expected 5% by chance.

6) Mean MAF test

We compared the mean MAF in each population to that under the expectation of random genetic drift. In order to avoid the effect of GWAS ascertainment bias in our results, we first sampled 100,000 sets of 61 SNPs across the genome, assuring that the selected markers: i) have been associated in Europeans by GWAS to at least one of the phenotypes/traits described in the GWAS catalog[35], https://www.genome.gov/26525384, last accessed March 20, 2015) and ii) are also present in the HapMap population of interest. In the case of CEU individuals, the number of BMD-GS SNPs present in the HapMap Phase II dataset was 61 and the number of GWAS SNPs from where to sample was 8,349. In the case of YRI

individuals, the number of GS-BMD SNPs was 58 and the number of GWAS SNPs from where to sample was 7,500; for JPT+CHB individuals, there were 58 GS-BMD SNPs and 7,552 GWAS SNPs from where to sample. A one-tailed p-value was computed counting the number of times that the sampled mean was higher than the observed one.

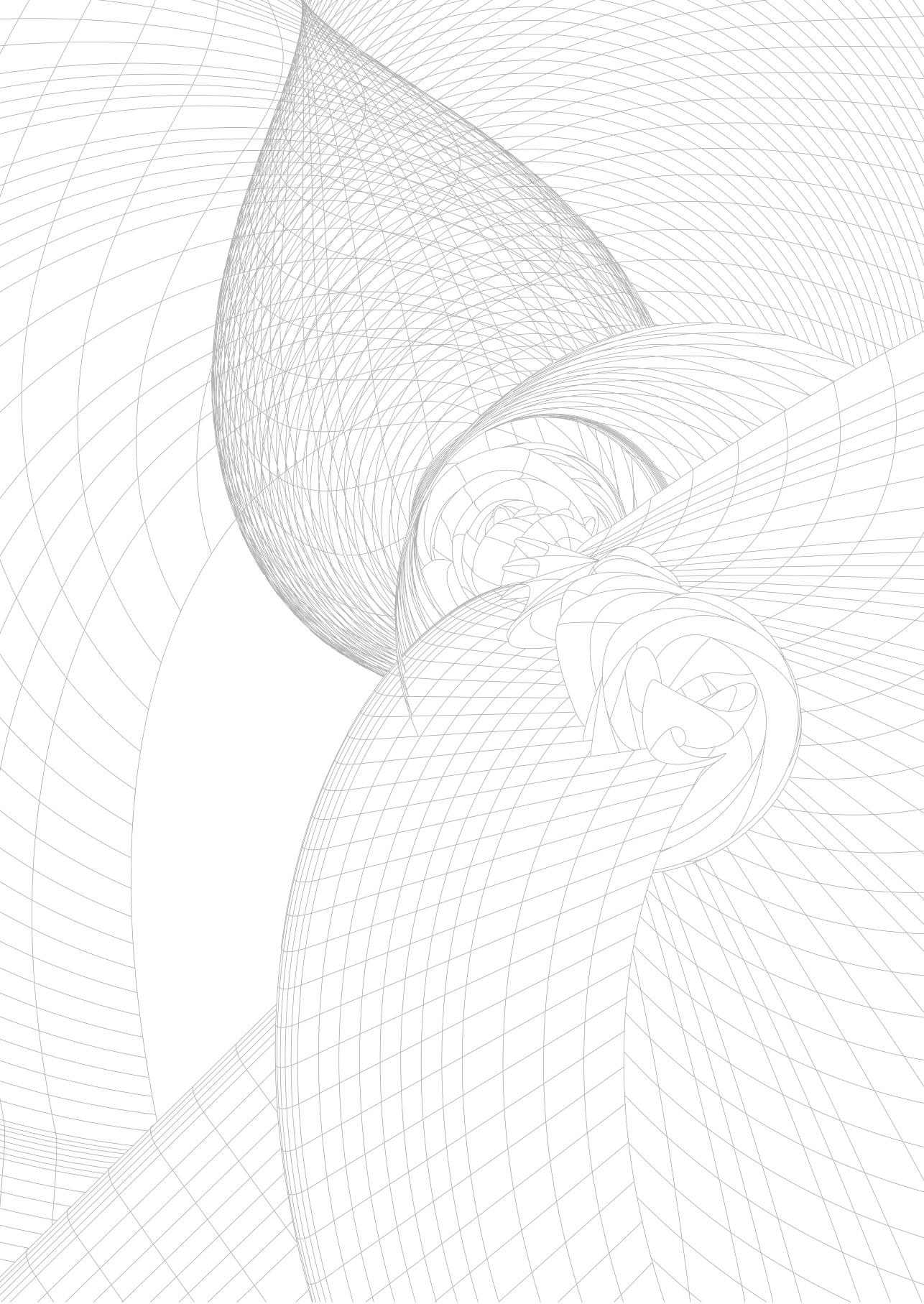
Detailed acknowledgements and online resources can be found in the published article online: http://mbe.oxfordjournals.org/content/early/2015/07/29/molbev.msv170.long

REFERENCES

1. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. **Am J Hum Genet**. 2012; 90(1):7-24.
2. McEvoy B, Beleza S, Shriver MD. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. **Human Molecular Genetics**. 2006;15:R176-R81.
3. Chen R, Corona E, Sikora M, et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. **PLoS Genet**. 2012;8(4):e1002621.
4. Turchin MC, Chiang CW, Palmer CD, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. **Nat Genet**. 2012;44(9):1015-9.
5. Corona E, Chen R, Sikora M, et al. Analysis of the genetic basis of disease in the context of worldwide human relationships and migration. **PLoS Genet**. 2013;9(5):e1003447.
6. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. **PLoS Genet**. 2014;10(8): e1004412.
7. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
8. Cummings SR, Black DM, Nevitt MC, et al. Bone-Density at Various Sites for Prediction of Hip-Fractures. **Lancet**. 1993;341(8837):72-5.
9. Finkelstein JS, Lee MLT, Sowers M, et al. Ethnic variation in bone density in premenopausal and early perimenopausal women: Effects of anthropometric and lifestyle factors. **J Clin Endocr Metab**. 2002; 87(7):3057-67.
10. Marshall LM, Zmuda JM, Chan BKS, et al. Race and ethnic variation in proximal femur structure and BMD among older men. **J Bone Miner Res**. 2008;23(1):121-30.
11. Nowlan NC, Jepsen KJ, Morgan EF. Smaller, weaker, and less stiff bones evolve from changes in subsistence strategy. **Osteoporos Int**. 2011;22(6):1967-80.
12. Ryan TM, Shaw CN. Gracility of the modern Homo sapiens skeleton is the result of decreased biomechanical loading. **Proc Natl Acad Sci U S A**. 2015;112(2):372-7.
13. Jaddoe VW, van Duijn CM, van der Heijden AJ, et al. The Generation R Study: design and cohort update 2010. **Eur J Epidemiol**. 2010;25(11):823-41.
14. International HapMap C. The International HapMap Project. **Nature**. 2003;426(6968):789-96.
15. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. **Science**. 2008;319(5866):1100-4.
16. Nam HS, Kweon SS, Choi JS, et al. Racial/ethnic differences in bone mineral density among older women. **J Bone Miner Metab**. 2013;31(2):190-8.
17. Zemel BS, Kalkwarf HJ, Gilsanz V, et al. Revised reference curves for bone mineral content and areal bone mineral density according to age and sex for black and non-black children: results of the bone mineral density in childhood study. **J Clin Endocrinol Metab**. 2011;96(10):3160-9.
18. Finkelstein JS, Lee ML, Sowers M, et al. Ethnic variation in bone density in premenopausal and early perimenopausal women: effects of anthropometric and lifestyle factors. **J Clin Endocrinol Metab**. 2002;87(7):3057-67.
19. Styrkarsdottir U, Halldorsson BV, Gudbjartsson DF, et al. European bone mineral density loci are also associated with BMD in East-Asian populations. **PLoS One**. 2010;5(10):e13217.
20. Ham S, Roh TY. A Follow-up Association Study of Genetic Variants for Bone Mineral Density in a Korean Population. **Genomics Inform**. 2014;12(3):114-20.
21. Deng YH, Zhao L, Zhang MJ, et al. The influence of the genetic and non-genetic factors on bone mineral density and osteoporotic fractures in Chinese women. **Endocrine**. 2013;43(1):127-35.
22. Marigorta UM, Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. **PLoS Genet**. 2013;9(6):e1003566.

23. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. **PLoS Biol**. 2013;11(9): e1001661.
24. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **Plos Genet**. 2012;8(7):e1002718.
25. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **Plos Genet**. 2014;10(6):e1004423.
26. Nilson F, Moniruzzaman S, Andersson R. A comparison of hip fracture incidence rates among elderly in Sweden by latitude and sunlight exposure. **Scand J Public Health**. 2014;42(2):201-6.
27. Kanis JA, Oden A, McCloskey EV, et al. A systematic review of hip fracture incidence and probability of fracture worldwide. **Osteoporos Int**. 2012;23(9):2239-56.
28. Cauley JA. Defining ethnic and racial differences in osteoporosis and fragility fractures. **Clin Orthop Relat Res**. 2011;469(7):1891-9.
29. Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. **Lancet**. 2002; 359(9319):1761-7.
30. Xing J, Watkins WS, Shlien A, et al. Toward a more uniform sampling of human genetic diversity: a survey of worldwide populations by high-density genotyping. **Genomics**. 2010;96(4):199-210.
31. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. **Science**. 2009;324(5930):1035-44.
32. Wu DD, Zhang YP. Positive selection drives population differentiation in the skeletal genes in modern humans. **Hum Mol Genet**. 2010;19(12):2341-6.
33. Perry GH, Foll M, Grenier JC, et al. Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. **Proc Natl Acad Sci U S A**. 2014;111(35):E3596-603.
34. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. **Curr Biol**. 2010;20(4):R208-15.
35. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. **Nucleic Acids Res**. 2014;42(Database issue):D1001-6.
36. Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. **Evolution**. 2005;59(11):2312-23.
37. Casto AM, Feldman MW. Genome-wide association study SNPs in the human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations? **PLoS Genet**. 2011;7(1):e1001266.
38. Kalkwarf HJ, Zemel BS, Gilsanz V, et al. The bone mineral density in childhood study: bone mineral content and density according to age, sex, and race. **J Clin Endocrinol Metab**. 2007;92(6):2087-99.
39. Rosenberg NA. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. **Ann Hum Genet**. 2006; 70(Pt 6):841-7.
40. Lewiecki EM, Gordon CM, Baim S, et al. Special report on the 2007 adult and pediatric Position Development Conferences of the International Society for Clinical Densitometry. **Osteoporos Int**. 2008;19(10):1369-78.
41. International HapMap C. A haplotype map of the human genome. **Nature**. 2005;437(7063):1299-320.
42. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. **Nucleic Acids Res**. 2001;29(1):308-11.
43. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Res**. 2009;19(9):1655-64.
44. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. **Am J Hum Genet**. 2011;88(1):76-82.

45. Kasprzyk A. BioMart: driving a paradigm change in biological data management. **Database (Oxford)**. 2011;2011:bar049.
46. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. **PLoS Genet**. 2009;5(5):e1000471.
47. Pybus M, Dall'Olio GM, Luisi P, et al. 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. **Nucleic Acids Res**. 2014;42(Database issue):D903-9.
48. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. **Stat Comput**. 2000;10(4):325-37.



Chapter 6

General Discussion

Bone homeostasis during childhood has a complex and dynamic etiology influenced by environmental and genetic factors. In this thesis, these factors were characterized by studying bone traits using dual-energy X-ray absorptiometry (DXA) in participants of the Generation R study, a multi-ethnic birth cohort based in Rotterdam the Netherlands. The bone health outcomes of participants included in the different investigations presented in this thesis were assessed using the same DXA device at a similar average age of 6 years, which constitutes a unique and favorable setting for both epidemiological and genetic investigations. In addition to studying the sexual and ethnic dimorphism of bone, the effects of maternal nutrition, early growth, and muscle mass in pediatric bone mass were also researched. This thesis placed particular focus on the influence of genetic factors underlying the heterogeneity of bone mineralization. As Generation R is one of the largest pediatric studies worldwide with both bone phenotype and genome-wide genotyping information, our contribution to the field of genetics of bone attainment has been substantial. One of the elements that have facilitated the advancement in the investigation of the genetic architecture of complex traits, such as bone accrual, is the successful application of the hypothesis-free genome-wide association studies (GWAS) approach. Notably, the Generation R Study has made part of the vast majority of GWAS leading to the discovery of variants associated with bone traits in children.

Throughout this thesis, I have presented compelling evidence that genetic factors account for a substantial proportion of the variance in pediatric bone mineral density (BMD). Lower bound heritability estimates, determined by single nucleotide polymorphisms (SNPs) present in commercial genotyping arrays, range between 35-50% (Chapter 4.3). Therefore, it is expected that the bone acquisition variance, explained by the additive effect of the complete set of polymorphisms in the genome, is even higher. The description of the genetic discoveries in the pediatric setting described in this section is partly based on a recently published review [1].

`1. SCIENTIFIC IMPACT OF THIS THESIS

The creation of the GWAS dataset of the Generation R Study has been of great scientific impact, as reflected by approximately 50 studies published to date in different research areas [2-50]. Moreover, this dissertation provides tools for the analysis of this dataset and brings new ideas to benefit from the multiethnic background of participants.

In the particular case of musculoskeletal traits, we have so far identified 14 different loci to be associated with pediatric BMD and 8 with the combined phenotype of BMD and lean mass in children, were only one was driven by the lean mass association (**Figure 1**). The studies presented here are providing valuable insights as to how molecular pathways influence bone growth and accrual. Genetic variants discovered so far unveil well-known

bone metabolism pathways, such as the Wnt signaling pathway, but also pinpoint novel genes and pathways not previously implicated in bone traits.

2. GWAS SET-UP OF THE MULTI-ETHNIC GENERATION R STUDY

In the last decade, GWAS have provided insights into the biological complexity of different diseases [51]. Chapter 2.1 describes the different challenges I coped with in setting up the GWAS dataset of the multiethnic Generation R Study. Back in 2011, (when the first efforts described in this thesis were started) the information regarding the steps for quality control,

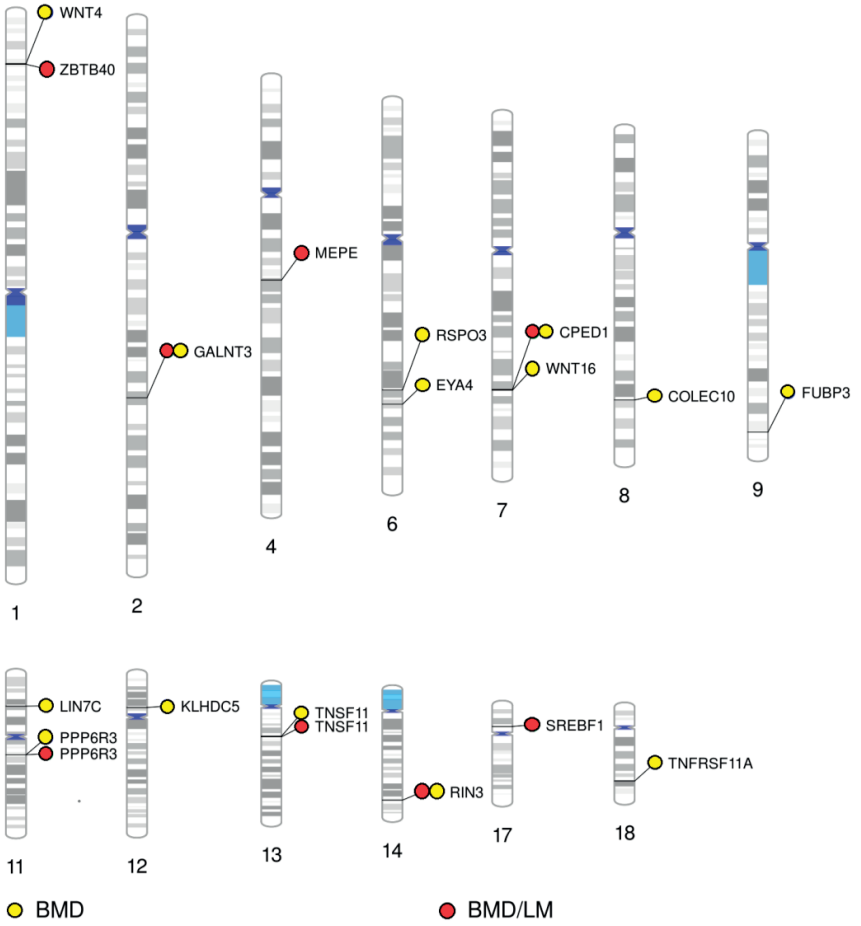


Figure 1. Musculoskeletal loci identified by GWAS in children and/or young populations described in this thesis. Loci are annotated with the most representative gene(s), as prioritized by the different GWAS publications. BMD= Bone Mineral Density, LM= Lean Mass.

study design and analytical procedures in admixed populations was scarce, considering that most GWAS have been performed in individuals with Northern-European ancestry. Exclusion of ethnic outliers from analyses in the Generation R Study, as was then the customary practice, was not an efficient strategy. This would have implied a high loss of statistical power for the study as well as a waste of resources given the large number of individuals of non-European origin (~50%) in the population. Likewise, the inclusion of admixed or non-European populations in the GWAS arena is steadily growing, bringing the opportunity to boost power by means of increasing sample size and, additionally, allowing the assessment of trans-ethnic replicability of GWAS findings. In this thesis, I have demonstrated that statistical methodologies (i.e., principal component analysis and linear mixed models) can successfully correct for high population substructure potential bias. This was evidenced using highly stratified traits like red hair pigmentation (Chapter 2.1, [52]) and Bechet's disease risk (Chapter 2.3, [9]) as proof of principle. This was then further corroborated by the successful application of the methodology across several musculoskeletal traits (Chapter 4, [3,7,8]). Nevertheless, differences in minor allele frequency (MAF) and linkage disequilibrium (LD) across populations from different ancestries can still result, for specific variants, in a dilution of the effect during the association analysis [53]. One such example is given in Chapter 2.1, where an association between rs13223036 and skull (SK-BMD) detected at genome-wide significance level (GWS, $P<5x10^{-8}$) in individuals of non-European ancestry, did not reach the GWS threshold in a similar number of individuals from the European ancestry subset. Yet, this variant was also associated at GWS level with SK-BMD in the combined dataset. Therefore, the GWAS design for admixed or multi-ethnic populations involves a tradeoff between the size of the study panel and the possible dilution of the effect of the variant (assuming potential stratification control is in place). There are alternatives at the meta-analytical level to combine data from studies of different ethnic background [54], but these are still not a solution for populations as the Generation R study withholding a large fraction of highly admixed individuals.

To note, the majority of the studies presented in this thesis are based on genotyped, or HapMap imputed data and consequently use common variants (MAF<0.05) in the analyses. It is plausible that the effect of MAF and/or LD differences across populations on the association results (e.g., false-negative associations) might be accentuated at a spectrum of lower allele frequencies. This would occur mainly because as shown by the 1000G project [55], rare variants are more specific to a given ethnic group, owing to demographic or evolutionary forces. Consequently, less frequent and rare variants are in general less correlated to each other than common variants[56].

3. LOW-FREQUENCY AND RARE VARIANTS IMPUTATION IMPROVEMENT

Aside from the analytical challenges posed by low frequency and rare variants (MAF<0.05), it is of high relevance to association studies to maximize the imputation quality of these kind of variants, in order to provide a more comprehensive picture of genetic factors influencing a specific trait. Whereas the development of next-generation arrays and new and larger sequence-based reference panels (e.g., the Genome of the Netherlands GoNL[57], the 1000 Genomes panel[58], UK10K [59] or more recently the Haplotype Reference Panel (HRC)[60]) facilitate imputation of low frequent single nucleotide polymorphisms (SNPs) by ensuring sufficient coverage, the efficiency of the imputation methodology remains determinant to increase the number of variants for genotype-phenotype analysis, due to the stringent quality checks used in GWAS. In Chapter 2.2, I describe a two-step strategy to improve the imputation performance by adding an intermediate step, the imputation to a local reference panel, previous to the imputation to a denser reference set from broader geographic origin. This strategy not only increased the number of high-quality markers (r^2>0.3) by 18% for variants in the lower frequency spectrum (MAF 0-5%) but also increased the general imputation quality (28% average increase) and genotype accuracy (overall accuracy up to 97.9%) [61]. It has to be noted, however, that this strategy was tested only in individuals of European ancestry and using a particular reference panel (e.g., the European subset from the 1,000 Genomes reference panel Version 3, November 2010). The increasing quality and size of the reference panels available (i.e., HRC: 64,976 haplotypes [60] as opposed to the 362 haplotypes from European ancestry from the 1000 Genomes [55] used in the study) may optimize significantly imputation performance and, therefore, decrease the boost provided by the proposed strategy. The applicability of this strategy to admixed populations, as the Generation R Study, is limited as no "local panel" would be properly defined. Recently, we have employed a specialized software for imputation of admixed populations (e.g., MaCH-Admix [62]), which uses *effective* reference panels to impute genomic regions based on their local ancestry. The use of this strategy resulted in a mean imputation quality of 95% based on 990,000 SNPs, for a combined set of Generation R and the Netherlands Twin Register (NTR) subsets (N~6,000) [45]. This result suggests that the use of local ancestry aware imputation has the potential to improve imputation performance, and thus increase GWAS coverage and power.

4. SOURCES OF HETEROGENEITY IN GWAS OF PEDIATRIC MUSCULOSKELETAL TRAITS

Up to 60 different loci have currently been described as associated, at GWS level, with adult BMD[63-68], most of them discovered by the large international collaboration em-

braced by the GEnetic Factors for Osteoporosis (GEFOS) consortium in elderly adults. Despite this success, only ~5.8% of the estimated heritability of adult BMD is accounted for by 56 of these loci in a GWAS that encompassed up to 84,000 adults [64], suggesting that many more genetic variants remain to be discovered. It is expected that a larger sample size GWAS of adult and elderly individuals would unveil more variants. Nonetheless, an alternative strategy is to perform BMD GWAS in cohorts of children and/or young adults. BMD measured early in the life-course, may be less influenced by the cumulative effect of non-genetic (i.e., environmental or lifestyle) factors [1]. Therefore, this alternative can prove to be more powerful as the effect of specific loci regulating bone acquisition and peak bone mass attainment, can be much less prominent in elderly populations (Chapters 4.1 & 4.3). Although, presented in the context of BMD, this strategy of identifying genetic variants in younger populations rather than older groups, might well be applied to other traits with large environmental effects as lean mass (Chapter 4.4).

Another possible source of heterogeneity inherent to the analysis of BMD (providing both challenges and opportunities) is that the genetic determinants of BMD can exert site-specific effects throughout the human body, as a reflection of the intrinsic complexity of the skeletal system. Previous GWAS have already shown the existence of skeletal site heterogeneity, in which specific loci are more strongly associated with BMD at the femoral neck (FN) than BMD at the lumbar spine (LS) or vice versa [64,66,67]. The exact reason underlying this difference is still to be fully elucidated. Studies of pediatric BMD represent an ideal setting in which to investigate skeletal-site differences, as these routinely include total body (TB-) DXA scans, whereas most adult studies use DXA to measure BMD only at sites where fracture usually occurs (e.g., the hip, spine, and forearm). Nevertheless, since total body DXA is employed for the assessment of body composition, TB-BMD can also be used in an integrative approach in children and adults (Chapter 4.2). As performed in Chapter 4.4, the TB-measurement can be further decomposed in order to study BMD at sites not normally investigated, such as the skull (SK), lower limbs (LL) and upper limbs (UL).

The availability of healthy children with DXA measurements is less than in adult populations, constituting a limitation for the large sample sizes needed by GWAS. In Chapter 4.5, a multivariate approach was implemented incorporating BMD-correlated measurements also obtained by DXA. While this is a way to maximize power by benefiting from using the correlation between measurements obtained in the same scans, it also examines pleiotropic relations, which feature another important source of heterogeneity reflecting the complex genetic architecture of developmental musculoskeletal traits.

4.1 Age-dependent effects as a possible source of BMD genetic heterogeneity

The first GWAS of TB-BMD involving the Generation R Study reported an association of this trait with variants in the 7q31.31 locus [3]. As described in Chapter 4.1, this analysis comprised ~2,600 six-year-old children from this study, as discovery population, and an additional five cohorts that represented distinct age groups, ranging from 10 – 75 years (n = 11,052). Results from the discovery set showed a significant association of variants in this locus with TB-BMD and SK-BMD, that was successfully replicated across the other five studies. Variants in this locus, harboring the *WNT16*, *CPED1,* and *FAM3C* genes, had been associated with adult FN- and LS-BMD, in a previous GEFOS GWAS meta-analysis [64]. Since then, this association has been replicated by us (Chapters 4.1 and 4.2) and others in several studies across diverse bone traits [69-73]. The sample sizes needed to detect this association signal at GWS-level in our analysis in children (discovery set N=2,660), and another study of pre-menopausal women (discovery set N= 4,061) [69] are considerably smaller than the one in the initial meta-analysis (N=32,000), mostly composed of elderly subjects [64]. In Chapter 4.1, in the SK-BMD GWAS, but not in the TB-BMD analysis, I describe the age heterogeneity of rs7801723, an SNP mapping to *CPED1*, with a larger effect in young populations, that seems to mimic the natural trajectory of BMD, decreasing with age after peak bone mass attainment. Nonetheless, there was not statistical evidence for the effect of variants in this locus to be different between children and adolescents (4-15 years; N= 11,800), and the elderly population (>60 years; N=20,300) in the TB-BMD GWAS meta-analysis following a life-course approach (Chapter 4.3). However, ten variants from this locus were nominally significantly ($P<0.05$) different between the two groups. Interestingly, those variants mapping in the vicinity of the 3'-end of *CPED1* had effects 1.8 times larger in the elderly as compared to the children and adolescents, while those elsewhere in *CPED1* had effect sizes almost two times larger in children as compared to the elderly individuals.

In Chapter 4.1, via conditional analysis, the existence of allelic heterogeneity (at least two different signals) in this locus was demonstrated, sheding some light on the understanding of the seemingly contradictory results described above. In addition to variants in this complex locus, in Chapter 4.3, age-dependent effects for variants mapping within or near *ESR1, RIN3, RANKL* and other loci were found, most of them showing the strongest associations in children. Variants in the *ESR1* locus had pretty weak associations with BMD in children. This is likely explained by the majority (>80%) of children in this group being pre-pubertal with low sex hormone level and not expected to play a role on BMD levels. Whereas in adults estrogen and sex hormones are key regulators of bone. The origin of the age effect heterogeneity underlying the other two loci remains to be elucidated.

As shown by the findings described in this thesis, age is an important source of heterogeneity in the effect estimate of genetic variants assessed by meta-analysis. This can also

be reflected on less power for the discovery of genetic associations and restrict the probability of success for replication of true findings (Chapter 4.3). Age-dependent effects have been assessed in GWAS by the introduction of "age X SNP" interaction in the association models [74]. The main drawback of this strategy is the need of cohorts with a sufficiently broad age range to contribute valuable information to a large meta-analysis. Accordingly, relationships as the ones displayed in Chapter 4.1 or Chapter 4.3 will most likely not be captured by that approach. In Chapter 4.3, the effect of the majority of the loci associated with BMD displaying a high degree of age heterogeneity was smaller in older adults than in the pediatric population. This suggests that the study of BMD and other complex traits with a large influence of lifestyle factors in young populations will increase power for the discovery of associated genetic factors as also shown for blood pressure [49] and BMI [75]. A better understanding of age-dependent effects would also provide insight into the role of gene-environment interactions in the determination of complex traits as accumulative effects of most of the environmental factors evaluated so far (i.e., smoking [76,77] or physical activity [78,79]) relate to aging.

4.2 Site-specific effects as a possible source of BMD genetic heterogeneity

As mentioned above, the first time the 7q31.31 locus was reported as associated with BMD was in a meta-analysis including more than 32,000 individuals, comprising adult and elderly populations part of the second GEFOS meta-analysis [64]. However, no variants in these loci were identified in the first GEFOS effort encompassing 19,195 subjects in 2009 [63]. So, the larger sample size required to detect the signal in the elderly is in contrast to the smaller sample size required to identify the same GWAS signal in younger populations. Although these results point to age-specific effects, I show in Chapter 4.3 a GWS association mapping to this region with TB-BMD in about 20,000 individuals >60 years old. Moreover, Zheng and colleagues also depicted the association at forearm (FA-) BMD in 5,672 individuals around their fifth decade of life [73]. Therefore, skeletal site-specificity might be underlying the larger power of our study. Indeed, the second GEFOS and our meta-analyses were performed assessing distinct skeletal sites (e.g., LS-, FN-BMD vs. TB-BMD).

In Chapter 4.4, I explored the heterogeneity underlying skeletal site-specificity in the BMD-association analyses. It was found that, in children, BMD heritability estimates differ across skeletal sites. Actually, common variants explained a larger proportion of the overall BMD-variance at the skull as compared to BMD measured at the appendicular sites (UL- upper limb and LL- lower limb). Moreover, there was a higher genetic correlation between BMD measured at the appendicular sites (UL-BMD and LL-BMD) than between any of these measurements and SK-BMD. This observation could reflect the differential exposure of each skeletal site to varying environmental stimuli that influence BMD. Specifically, the skull may be less influenced by factors acting through mechanical loading, as compared to

appendicular sites. Similarly, these differences could be a consequence of the biological processes that govern growth and maintenance of the bone at these sites (e.g., intramembranous ossification at the skull and endochondral ossification at the long bones) or reflecting the composition of bone at each skeletal site (e.g., trabecular vs. cortical fraction). For instance, variants upstream of the *EYA4* gene, which are not associated with BMD at other skeletal sites, are strongly associated with SK-BMD [7].

Even if the biological mechanism driving this association is still unknown, variants in the same locus have been associated with cortical volumetric BMD, assessed by peripheral quantitative computed tomography (pQCT) at the tibia in children and young adults [80], supporting the role of this locus on bone metabolism. In addition to *EYA4*, variants showing site-dependent effects located within or near *TNFRSF11A, TNFRSF11B, RSPO3,* and *LGR4* were found. Most interestingly, the locus *CPED1/WNT16* displayed once more complex patterns of association in this study. Variants located closer to *CPED1* were associated with SK- and UL-BMD, but not with LL-BMD; whereas variants at close proximity or within *WNT16* were most strongly related to UL-BMD when compared to the LL- and SK-BMD. This is in line with the allelic heterogeneity described above.

Importantly, I demonstrated in Chapter 4.2 that TB-BMD is a suitable phenotype for the identification of genetic factors influencing BMD. Through the study of this phenotype, 53 different loci were detected, many of them previously reported in the literature and showing site (e.g., LS-, FN-, FA-BMD) or compartmental specificity (e.g., trabecular bone, cortical bone). I also described novel loci, whose role on bone metabolism remains to be elucidated, although with a large fraction enriched by genes involved in the WNT-pathway. Taken together, all the independent variants identified in this chapter explained approximately 8% of the TB-BMD variance, suggesting that larger GWAS would still provide the opportunity for finding many more variants associated with this TB-BMD trait.

4.3 Genetic pleiotropic effects of BMD and lean mass

There is an evident mechanical interaction between bone and muscle mass, and also at a molecular level, biochemical cross-talk. Consequently, it is likely that genetic factors have pleiotropic effects, influencing both muscle and bone phenotypes. Lean mass (LM) or fat-free mass is highly correlated with muscle mass and derived from the same DXA scan as TB-BMD, allowing the study of variants exerting simultaneous effects on both bone and muscle development. Modeling simultaneously two correlated phenotypes increases the effective sample size for analysis, providing a crucial improvement in statistical power of GWAS. This strategy was followed in Chapter 4.5, where I describe a bivariate analysis of TB-BMD and TB-LM yielding eight GWS signals of association mapping within or near *WNT4, GALNT3, MEPE, CPED1/WNT16, PPP6R3/LRP5,* and *TOM1L2/SREBF1*. The association with TB-BMD drove for a large fraction the GWS signals of these implicated variants except for those

mapping to the last two loci. Whereas, there is evidence of association in the bivariate analysis for *PPP6R3/LRP5* for TB-LM ($P<10^{-6}$), the association of this locus is stronger with TB-BMD ($P<10^{-9}$). There is robust evidence supporting the role of *LRP5* in mechanotransduction (reviewed elsewhere [81]), what might explain the bivariate association observed in our analysis. Conversely, the association arising within the *TOM1L2/SREBF1* locus was driven by both TB-BMD (beta=0.043SD; $P=2x10^{-3}$) and TB-LM (beta=-0.056SD; $P=5.5x10^{-5}$), although stronger in the latter. Noteworthy, by comparing results from this study I confirmed that the bivariate analysis outperforms the univariate analysis, particularly when the direction of the effect is opposite in sign [82].

Examining further the *TOM1L2/SREBF1* locus, substantial evidence pointing to *SREBF1* being the gene underlying this association was collected. Active SREBP-1, the product of *SREBF1*, is known as a negative regulator of myogenesis [83] and protein synthesis in differentiated myotubes [84]. Similarly, a high concentration of the active form of this protein induces mineralization of osteoblastic cultures *in-vitro* and bone formation *in-vivo* [84]. Therefore, the cross-phenotype association suggests biological pleiotropy, meaning the same gene causing different effects across different tissues. Nevertheless, *in-vitro* or *in-vivo* models evaluating both muscle and bone simultaneously are needed to determine if the effect of the associated allele is direct or indirect in both myogenesis and osteoblast mineralization. Overall, SREBP1 affects both bone and muscle tissues yet in opposite directions; this contention is in agreement with the opposite effect of the leading SNP observed in both univariate and bivariate association analyses.

It is important to highlight that the genetic architecture underlying muscle mass remains predominantly unknown and consequently, there is considerable room for genetic discoveries. Unfortunately, considering that large GWAS meta-analyses have yielded a limited number of associated loci with muscle traits (i.e. lean mass, muscle strength), innovative analytical methodologies (like bivariate/multivariate analysis) would be proven fruitful. Genes associated with LM as *FTO* [85], and *SREBF1* (Chapter 4.5) are also known to be key factors in adipogenesis. This implies that a multivariate analysis including the third major component of body composition, body fat, would increase our understanding of both muscle and bone metabolism and in general, of human physiology. Alternatively, additional analytical models managing to reduce the phenotypic variance of muscle mass by including important cofounders as physical activity, for example, should be explored. Early this year Korostishevsky and colleagues implemented a combined metabolomics and genomics approach to study lean mass, setting an example of integrative alternatives for the discovery of genetic polymorphisms associated with LM [86].

4.4 Trans-ethnic genetic studies of BMD

I explained above the challenges in the GWAS set-up of multiethnic studies. In Chapter 5, in turn, an original approach to study the well-documented racial differences in BMD [87-92] by combining genomic and evolutionary methods was proposed, profiting this way from the ethnic heterogeneity intrinsic to the Generation R Study. It was found that indeed, children of Sub-Saharan ancestry had significantly higher levels of TB-BMD when compared with children of European or East-Asian ancestry. This result was mirrored at the genetic level after it was confirmed that the compound effect of variants associated with BMD in adults[64] is associated with infant BMD. The frequency of those alleles associated with increased BMD was systematically elevated in individuals of Sub-Saharan African ancestry, observation that was paralleled across a worldwide population distribution (i.e. Human Genome Diversity Panel [93]). Certainly, it is also possible that variants, not captured by the current GWAS in European populations, may contain an additional set of alleles contributing to BMD variation in African or Asian populations. The overrepresentation of ancestral alleles in the set of alleles increasing BMD indicates that phenotypic states related to increased BMD represent the ancestral state in humans. Studies of hominin fossil records have reached similar conclusions for phenotypes as bone stiffness [94] or high trabecular density [95]. Most importantly, they have provided strong evidence that the gracility of modern humans is the result of decreased biomechanical loading due to a tendency to ever increasing levels of sedentariness [96]. As the amount of BMD differentiation between sub-Saharan and not sub-Saharan populations, cannot be uniquely explained by demographic factors, the presented results are compatible with the hypothesis of selective pressures acting on the genetic determinants of BMD, thus, providing a new example of polygenic adaptation to a human trait [8].

Trans-ethnic efforts have traditionally used admixture mapping to foster the discovery power in admixed populations [97]. Admixture mapping might be applied to populations that have recently undergone gene flow from two ancestrally isolated populations, when these ancestral populations differ both in allele frequencies and disease prevalence, as usually applied to case-control designs[98]. In Chapter 2.1, I acknowledge that the complex structure of the Generation R Study, where admixture of individuals cannot be easily discerned just by assessing the combination of two ancestral populations, constrains the use of this technique. Nevertheless, in Chapter 5, following a similar approach based only on participants with European Ancestry, it was shown that the proportion of genetic Sub-Saharan African ancestry in their genomes was associated with higher BMD levels.

As the catalogue of genetic variation in the non-European subpopulations increases, and the sample size of BMD-GWAS for these populations enlarges, we would get a more complete picture of the genetic factors underlying ethnic BMD variation and partially explaining disparities in fracture risk worldwide.

5. EPIDEMIOLOGICAL ANALYSES OF THE HETEROGENEITY OF BONE ATTAINMENT

5.1 Sexual dimorphism of bone density and strength

Along with the skeletal changes associated with age, there is a pronounced sexual dimorphism in the skeletal system. In Chapter 3.3, I investigated this kind of heterogeneity between sexes from an epidemiological perspective. However, SNP-sex interaction at a large scale was not evaluated in any of the studies presented in this dissertation. Sexual dimorphism in genetic loci has been evaluated for BMD previously in powerful settings [64,99] and only variants mapping to the X chromosome, not scrutinized in our association studies, showed statistical evidence of sex-dependent effects [64].

Bone development processes alter the amount of bone tissue within bones as well as its spatial distribution. Both of these parameters contribute independently to explain bone strength. Sexual dimorphism in bone strength is evident in the elderly population, conferring higher osteoporotic fracture risk to women. The risk of fragility fractures is 2.5 times higher in women compared to men [100]. In Chapter 3.3, I investigated pediatric hip geometry, by implementing the Hip Structural Analysis (HSA) software for analysis of the hip DXA scans of the participants of the Generation R Study. It was found that there are significant differences in the bone structure, even at pre-pubertal age, between boys and girls. Girls have thinner cortices and wider femurs, a configuration that confers their bones greater stress [101]. This study suggests that both men and women construct their bone differently since an early age, and girls have an innate structural disadvantage with respect to boys. Yet, such sex differences are attenuated after adjustment by lean mass. This indicates that the sexual dimorphism of bone is partly explained by differences in muscle mass. In line with these findings, a recent study on children of the same age showed that longitudinal changes in lean mass predict tibial bone geometry and mineralization, as assessed by pQCT[102]. In sum, these results highlight the importance of the establishment of exercise programs during the pre- and peri-pubertal years. These programs would be particularly beneficial for girls, as they have on average lower physical activity than boys of their same age [103] and "weaker" bone configurations. Indeed, different investigations have demonstrated that intervention strategies, targeting physical activity at early ages result in gains in peak bone mass that persist throughout life[104,105]. Noteworthy, palaeontological studies have allowed researchers to implicate patterns of physical activity, rather than nutrition, as the most important factor influencing the decrease in bone strength in modern human populations[96], as also discussed in this section.

5.2 Parental, fetal and infant determinants of bone accrual

In Chapter 3.2, the relationship between early growth patterns and childhood bone mass was evaluated. It was found that both fetal growth and infant growth independently contribute to explain bone mass variability at school-age. Furthermore, growth during the first year appeared to have the largest impact on bone mineral accrual at age six years. This relation between longitudinal growth and bone accretion might be, at least partially, mediated by an overlapping genetic architecture of these traits, or the hormonal levels that govern skeletal growth, which also (to a large extent) are genetically determined. To demonstrate this, assessing the genetic correlation between growth and bone parameters (like height and BMD) is needed. Nevertheless, the calculation of the genetic correlation between height and BMD might prove difficult as DXA measurements of pediatric BMD must be adjusted by size to avoid biases due to area changes during skeletal development and growth.

Chapter 3.1 focused on evaluating the effect of maternal first-trimester diet, a modifiable factor, on bone parameters at school age, we found a positive effect of diverse dietary nutrients like calcium and protein intake, homocysteine and vitamin B-12 blood concentrations and maternal phosphorus on the bone of the offspring at 6 years[106]. While maternal levels of calcium and phosphorous, the main bone forming minerals, and protein intake, a component of the organic bone matrix, are thought to exert a direct effect on the early fetal skeletal growth, other molecular mechanisms may play a significant role. For example, maternal dietary intake is known to modulate the fetal epigenome[107]. Nutrients can reverse or change epigenetic marks such as DNA methylation and histone modifications, thereby modifying gene expression[108]. For instance, vitamin B-12 can act as a methyl donor and accordingly affect the DNA-methylation patterns[107]. Likewise, protein intake is associated with reduced methylation of the glucocorticoid receptor in mice experiments[109]. Based on all these findings, it is clear that epigenetic mechanisms (further discussed below) will constitute a next step to understand the regulatory mechanisms underlying peak bone mass acquisition.

Altogether, these results (e.g., a positive relationship between early growth and bone mass at school age) show how early life processes can "track" bone development at later ages. In addition, they confirm the beneficial effect of a balanced maternal diet during pregnancy, including minerals, vitamins and proteins, in the offspring bone health with effects lasting at least to their school-age. The post-natal nutrition during the first years might be just as important as maternal diet. Notwithstanding, it was found the growth during the first year to be an important determinant of bone health at six years. There is also evidence that the trajectory of skeletal growth may be modified by factors as nutrient availability. While in the analysis presented in Chapter 3.1, we adjusted for breastfeeding, no evaluation of the maternal nutritional status or an assessment of other nutrients consumed by the infant during

this post-natal period that might affect bone homeostasis were included. In an additional study assessing nutrition at 13 months in participants of the Generation R Study (N=2,850), we found that adherence to a diet pattern rich in whole grains and dairy was positively associated with BMD at six years[110]. Hence, fetal and early post-natal nutritional exposures may permanently influence bone development.

6. FUNCTIONAL FOLLOW-UP OF LOCI WITH PUTATIVE CLINICAL RELEVANCE

Studies in Chapter 4 highlighted the influence of WNT16 in bone mass acquisition. Most clinically relevant, *WNT16* has been associated with risk of any type [64] and forearm [73] fracture. This scientific evidence has led to a number of functional studies characterizing the important role of WNT16 in skeletal regulation. Wnt16 knockout (KO) mice had reduced total body BMD, and gene expression profiles in human bone biopsies supported a role of both *CPED1* and *WNT16* on the BMD phenotypes (Chapter 4.1). The initial hypothesis of *WNT16* influencing cortical bone, as the associated traits were mainly cortical (e.g., TB-BMD, skull BMD, femoral neck BMD, forearm BMD), was confirmed by KO-mice developing spontaneous fractures as a result of low cortical thickness and high cortical porosity[111]. No trabecular phenotype was observed in these mice. Likewise, mechanistic studies revealed that *WNT16* is predominantly expressed in osteoblasts, in the cortical bone microenvironment, and inhibits osteoclastogenesis [111]. However, as a matter of fact, the most recent study of this molecule found that overexpression of *Wnt16* results in higher TB-BMD in mice, particularly at expenses of higher trabecular bone mass[112]. Therefore, this work of Mòverare and colleagues demonstrated that the loss of endogenous Wnt16 results specifically in cortical bone loss, whereas its pharmacological overexpression increases mainly trabecular bone mass [112]. As *WNT16* signaling in bone does not require normal estrogen action, it may be useful for the treatment of postmenopausal trabecular bone loss, which is accelerated in menopausal women [112].

The clinical translation of this GWAS finding will be particularly meaningful as most of the pharmacological treatments for osteoporosis function as anti-resorptives that halt further bone loss. Only one osteoanabolic drug (i.e., Teriparatide) is presently FDA approved; however, it is far from ideal as it is very expensive and requires daily administration via injection to ensure adequate bone formation [113].

While mapping the same locus as *WNT16*, *CPED1* has received less attention from the bone research community. Nonetheless and in line with the allelic heterogeneity in this locus, where at least two independent signals are influencing bone accrual, *Cped1* is highly

expressed in osteoblasts isolated from neonatal C57BL/6J mice and robustly correlated with the expression of genes that compose the extracellular matrix in the work led by Dr. Ackert-Bicknell. Also, she and her group are performing functional follow-up in knockout mice models at the University of Rochester, NY [114].

7. FUTURE PERSPECTIVES

7.1 Refinement of the phenotype

The main disadvantage of DXA is its two-dimensional methodology of a 3D structure, which results in areal density estimates of a true volumetric BMD. By using this 2D technique, it is impossible to differentiate trabecular from cortical bone, and even more challenging to assess geometric and micro-architectural properties that modulate bone strength (i.e., periosteal expansion, cortical density and thickness, and trabecular number and thickness). In contrast, pQCT offers density values independent of bone size as well as measurements at cross-sections of predominantly cortical or trabecular bone, enabling different constituents of bone mass to be analyzed separately (e.g., cortical and trabecular volumetric density (vBMD)). Nowadays, pQCT is increasingly being used to study determinants of bone strength, even at the genetic level [80,115]. Most importantly, pQCT enables the study of the bone-muscle unit, as from the muscle cross-section, area and density are also measured. Nevertheless, given the high radiation dosages intrinsic to CT, assessments are limited to peripheral body regions (e.g., tibia or forearm).

The new round of assessments of the Generation R Study starting in 2015, in which participants are in average nine years old, has implemented pQCT measurements. Likewise, the inclusion of jumping mechanography can go beyond the evaluation of the musculoskeletal structure to evaluate its actual function (i.e. muscle strength and power).

The combination of pQCT -and mechanography- derived information, will allow the investigation of the functional muscle–bone relationships, as performed before [116], in a powerful pediatric setting. Overall, measurements derived from these assessments would help to elucidate further the role of muscle in the sexual dimorphism of bone. In the same way, the study of pleiotropy (Chapter 4.5) could benefit from the refined phenotypes obtained by these techniques.

7.2 Longitudinal assessment of bone accrual

The results presented in this thesis are in agreement with the notion that both, genetic and environmental factors are interconnected and determine bone metabolism. The contribution of these factors as causal explanations in the variability of bone accrual is best studied from a longitudinal perspective, with the additional benefit of testing their impact on bone accrual rate. The Generation R Study has DXA-measurements of bone mass at six

months in a subset of 252 individuals [117], what has somehow limited our power to conduct a follow-up study at six years of age. Besides the measurements at six years (school age period), follow-up measurements are currently taking place (focus at 9) and programmed again when participants will have a mean age of 13 years. The acquisition of this prospective data, can provide a better framework to test the effect of lean mass in bone mass acquisition, or the evaluation not only of mother nutrition but infant nutrition in the rate of bone accrual. Moreover, studies analyzing the effect of BMD-associated SNPs in the rate of bone accrual from 9-13 years, identified loci not necessarily detected in the cross-sectional pediatric studies presented here [118]. Similar approaches have been applied to understand the interaction of these BMD-SNPs with sex and or maturation status [119,120]. While the signals detected by these studies are not novel and are captured by our life-course approach, the longitudinal study of their effect in the Generation R Study may contribute a missing piece to the bone accrual puzzle.

7.3 Enhancement of bone-related GWAS

The GWAS of bone traits of children and young adults have already proven its value. Nevertheless, our current knowledge concerning the impact of genetic factors in the determination of pediatric bone health may still be limited due to the lack of availability of other large pediatric cohorts with GWAS data. Also, the dynamic nature of BMD (i.e., changes with age), particularly during the growth years, and the influence of several other modifiable factors like physical activity and nutrition might decrease the statistical power of GWAS for this trait. Nonetheless, it remains critical that we stay on the path of uncovering the complete genetic architecture of pediatric BMD. One obvious route forward is to extend the number of studies involved in the research of genetic factors affecting bone accrual. As the price for GWAS genotyping is steadily declining, this array technology can soon be within reach for cohorts that at this moment have only collected bone measurements and DNA.

In addition, the availability of larger reference panels including samples from diverse populations and deeper sampling within populations like the Haplotype Reference Consortium (HRC)[60] would permit more accurate imputation of GWAS cohorts. The benefits will be greatest at low-frequency variants (MAF<0.05) since there is a higher chance for these variants to be poorly represented in smaller reference panels, and for studies in more ethnically diverse settings. The Generation R GWAS data has already been imputed to this panel. According to the developers, using this resource will lead to accurate genotype imputation at MAF as low as 0.1%, and a large increase in the number of SNPs tested within association studies [60]. Performing GWAS across very large samples of children and young adults imputed to reference panels of extended variant coverage would make possible to identify the great majority of genetic variants that are responsible for the variation in bone acquisition, as recently suggested for other complex traits such as body height or body mass index [75].

Alternatively, studies are applying other technologies such as whole-exome sequencing (WES) or whole-genome sequencing (WGS) to explore further the genetic architecture of complex traits. WES reads represent only ~2% of the genome, reducing storage and analysis cost. Nonetheless, many variants of interest map to non-coding regions and are not in LD with coding exons, hence not being captured by WES. The importance of these non-coding variants has been recently recognized, thanks to the ENCODE project [121], as they play an essential role in the regulation of gene expression and thus, in the observable phenotypic differences. Therefore, in contrast to the success observed for the mapping of monogenetic conditions, WES is not expected to yield much for the identification of genetic variants underlying complex traits, like BMD and other musculoskeletal phenotypes. WGS, on the other hand, allows the examination of non-coding and coding regions even though the generation of larger data significantly increases the computational burden for analysis. In osteoporosis research, a handful of BMD-associated variants with MAF<0.5% has been identified through the analysis of a combination of WGS data and imputed-GWAS data [65,66,68]. At this moment, the real benefit of sequencing approaches will only be seen through its application in large consortia efforts.

7.4 Novel omics technologies

While this dissertation has focused on genomics as charactherized by GWAS data, epigenetics and transcriptomics provide extra layers of information of the multiple levels of cell regulation that ultimately generate the observable phenotypic variability (**Figure 2**). These data sources are dynamic in nature and thus, feature high variability, not only at the individual level but also within an individual, for instance, at the tissue level. This dynamic attribute poses new challenges, from collecting sample sizes large enough to overcome the inherent error introduced per-se by the variability of the data, to the tissue-specificity of its

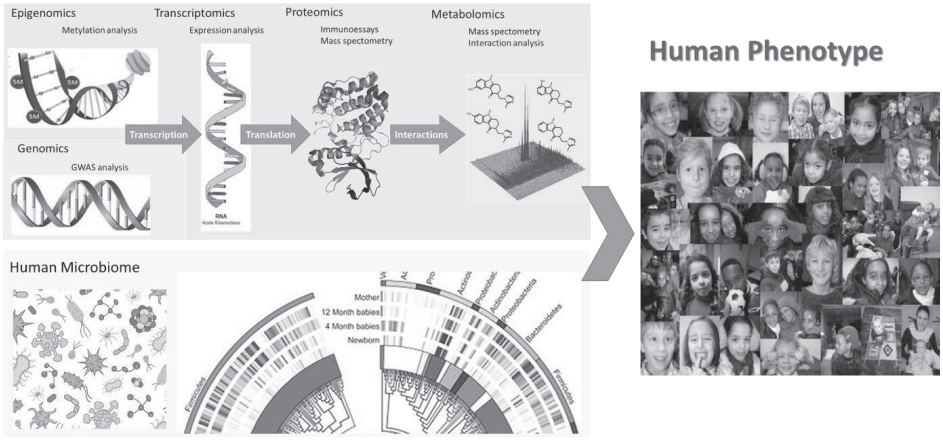


Figure 2. Scheme of the omics space. Integration of the different omics to define human phenotype.

signatures. On the top of these issues, the identification of the tissue most relevant to the disease/trait under investigation is not a straightforward task. Hence, there is undoubtedly a need for enlarging initiatives as GTEx[122], capturing gene expression in different human tissues or even further generate of a cell-specific *omics* atlas.

DNA methylation is key during human fetal development. Therefore, this data is essential in the study of early environmental and genetic causes of normal and abnormal growth, development and health from fetal life until young adulthood [123], the aim of the Generation R Study. In the immediate future, it will be possible to take advantage of the deep phenotyping of mothers and children acquired in this study for the investigation of differential methylation patterns as a link between environmental factors and trait variability. The Generation R Study methylation data is available for a subsample of around 900 children, collected from umbilical cord white blood cells. This small sample size provides limited statistical power for an hypothesis-free approach; nonetheless, it has already been used in association with specific genes[124]. Opportunely, the Pregnancy And Child Epigenetic (PACE) consortium is pulling a big effort to bring together all birth cohorts with methylation data, providing sufficient statistical power to perform genome-wide methylation surveys.

Similarly, there are ongoing efforts to characterize the association of the microbiome with phenotypic variability in which, with no doubt growth and bone health are to be investigated [125]. For instance, recent studies have demonstrated that gut microbiota regulates bone health via effects on the immune system in mice, which in turn regulates osteoclastogenesis [126]. Microbiome data from more than thousand participants of Generation R is currently being generated. As exploited by the GWAS approach described thoroughly in this thesis, the multiethnic nature of the Generation R Study will result in diverse challenges as well as opportunities for the analysis of microbiome data. With regard to the latter, some of the most interesting research questions will relate to obtaining a better understanding of health disparities across ethnic groups. Like for every new technology, in the microbiome field, new analytical frameworks and replication strategies are required. Therefore, some time will pass before it is feasible to apply such approximation on population-based studies in thousands of individuals.

Alternatively, the ENCODE project [121] has generated results of several experiments across the genome, epigenome and transcriptome for the evaluation of DNA methylation (e.g., Methyl RRB-seq), Open chromatin (e.g., DNase-seq), RNA binding proteins (e.g., RIP-seq), RNA profiling (e.g., RNA-seq), Transcription factor binding sites and histones (i.e., ChIP-seq), that are publicly available. The integration of this data into meta-dimensional analyses, even if still to be fully developed will unravel complex biological processes [127].

8. CONCLUDING REMARK

In this thesis, I have approached the etiology of bone health outcomes from different perspectives. The presented results show that modifiable factors, prone to interventions, as maternal diet and post-natal nutritional exposures influence bone development. The presented findings stress the importance of exercise programs from pre-pubertal ages to build up strong bones, particularly in girls. We have been pioneers in the use of GWAS to identify genetic factors influencing bone mineralization in children. Many of the genetic variants associated with BMD in adults, have effects on bone since an early age and the strategy of investigating young populations in several cases could provide gains in statistical power for gene discovery. Additionally, I have shown the complex manners in which genetic factors influence bone, varying by skeletal-site or age of evaluation or even exerting pleiotropic effect. This thesis has brought new ideas to the horizon of bone research as a result of the study of pediatric populations, the implementation of new statistical methods and the embracement of an evolutionary perspective to the traditional genetic studies. This research provides a first glance to the genetics of bone accrual and highlights the importance of a healthy lifestyle for early life for bone health, which could effectively reduce the risk of osteoporosis later in life.

REFERENCES

1. Kemp JP, Medina-Gomez C, Tobias JH, Rivadeneira F, Evans DM. The case for genome-wide association studies of bone acquisition in paediatric and adolescent populations. **BoneKEy Reports**. 2016; [In press].
2. Szekely E, Pappa I, Wilson JD, et al. Childhood peer network characteristics: genetic influences and links with early mental health trajectories. **J Child Psychol Psychiatry**. 2015.
3. Medina-Gomez C, Kemp JP, Estrada K, et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. **PLoS Genet**. 2012;8(7):e1002718.
4. Pappa I, St Pourcain B, Benke K, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. **Am J Med Genet B Neuropsychiatr Genet**. 2015.
5. van der Valk RJ, Kreiner-Moller E, Kooijman MN, et al. A novel common variant in DCST2 is associated with length in early life and height in adulthood. **Hum Mol Genet**. 2015;24(4):1155-68.
6. Stergiakouli E, Gaillard R, Tavare JM, et al. Genome-Wide Association Study of Height-Adjusted BMI in Childhood Identifies Functional Variant in ADCY3. **Obesity**. 2014;22(10):2252-9.
7. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. **PLoS Genet**. 2014;10(6):e1004423.
8. Medina-Gomez C, Chesi A, Heppe DH, et al. BMD Loci Contribute to Ethnic and Developmental Differences in Skeletal Fragility across Populations: Assessment of Evolutionary Selection Pressures. **Mol Biol Evol**. 2015;32(11):2961-72.
9. Kappen JH, Medina-Gomez C, van Hagen PM, et al. Genome-wide association study in an admixed case series reveals IL12A as a new candidate in Behcet disease. **PLoS One**. 2015;10(3):e0119085.
10. Mook-Kanamori DO, Marsh JA, Warrington NM, et al. Variants near CCNL1/LEKR1 and in ADCY5 and fetal growth characteristics in different trimesters. **J Clin Endocrinol Metab**. 2011;96(5):E810-5.
11. Mook-Kanamori DO, Ay L, Hofman A, et al. No association of obesity gene FTO with body composition at the age of 6 months. The Generation R Study. **J Endocrinol Invest**. 2011;34(1):16-20.
12. Sovio U, Mook-Kanamori DO, Warrington NM, et al. Association between Common Variation at the FTO Locus and Changes in Body Mass Index from Infancy to Late Childhood: The Complex Nature of Genetic Association through Growth and Development. **Plos Genetics**. 2011;7(2).
13. Szekely E, Herba CM, Arp PP, et al. Recognition of scared faces and the serotonin transporter gene in young children: the Generation R Study. **J Child Psychol Psyc**. 2011;52(12):1279-86.
14. Pluess M, Velders FP, Belsky J, et al. Serotonin Transporter Polymorphism Moderates Effects of Prenatal Maternal Anxiety on Infant Negative Emotionality. **Biol Psychiat**. 2011;69(6):520-5.
15. Bradfield JP, Taal HR, Timpson NJ, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. **Nat Genet**. 2012;44(5):526-+.
16. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. **Nat Genet**. 2012;44(2):187-92.
17. Ikram MA, Fornage M, Smith AV, et al. Common variants at 6q22 and 17q21 are associated with intracranial volume. **Nat Genet**. 2012;44(5):539-+.
18. Leermakers ETM, Taal HR, Bakker R, Steegers EAP, Hofman A, Jaddoe VWV. A Common Genetic Variant at 15q25 Modifies the Associations of Maternal Smoking during Pregnancy with Fetal Growth: The Generation R Study. **PLoS One**. 2012;7(4).
19. Cents RAM, Tiemeier H, Velders FP, et al. Maternal smoking during pregnancy and child emotional problems: The relevance of maternal and child 5-HTTLPR genotype. **Am J Med Genet B**. 2012;159B(3): 289-97.
20. Steenweg-de Graaff J, Roza SJ, Steegers EAP, et al. Maternal folate status in early pregnancy and child emotional and behavioral problems: the Generation R Study. **Am J Clin Nutr**. 2012;95(6):1413-21.

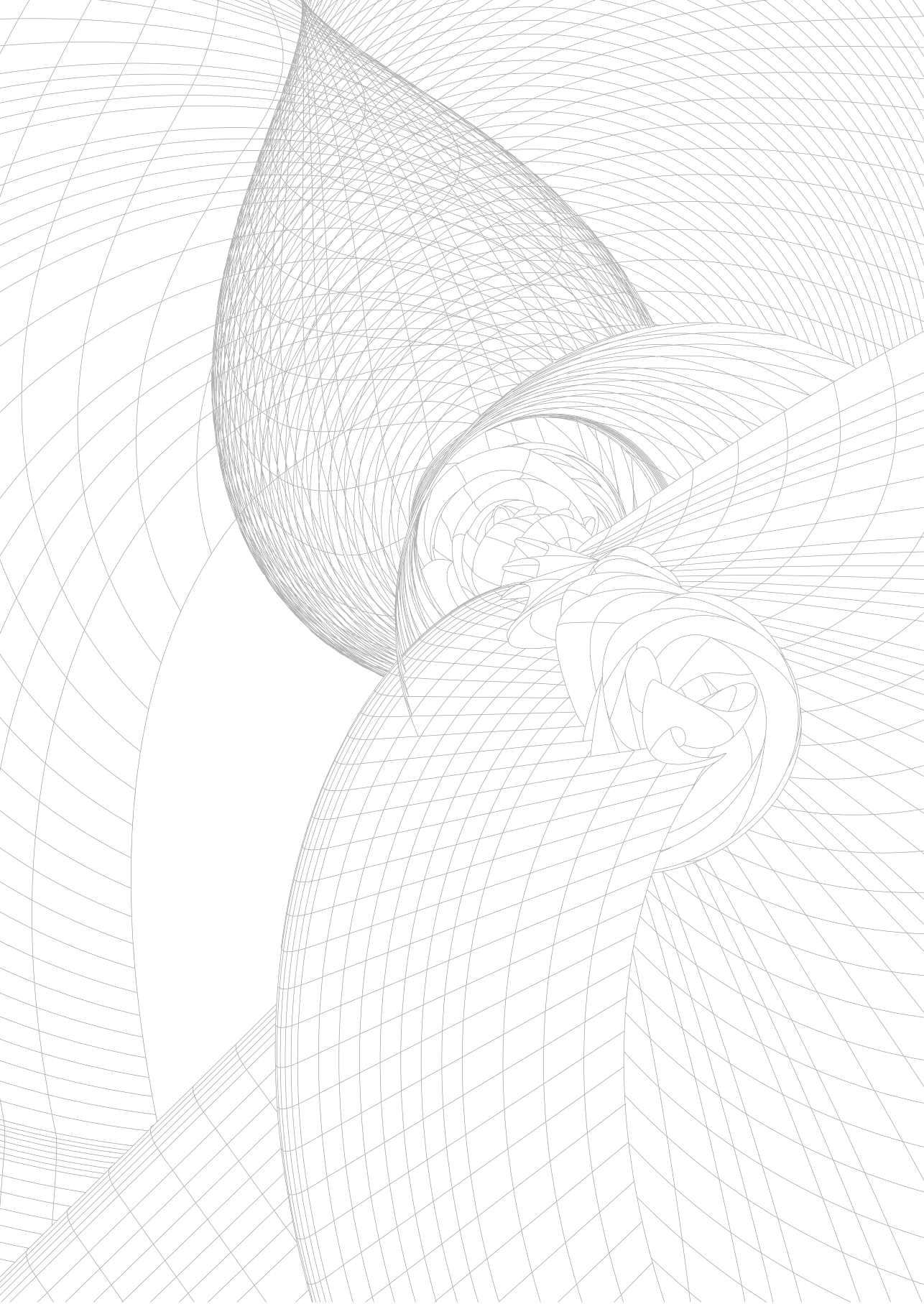
21. Taal HR, van den Hil LC, Hofman A, van der Heijden AJ, Jaddoe VW. Genetic variants associated with adult blood pressure and kidney function do not affect fetal kidney volume. The Generation R Study. **Early Hum Dev**. 2012;88(9):711-6.
22. Taal HR, St Pourcain B, Thiering E, et al. Common variants at 12q15 and 12q24 are associated with infant head circumference. **Nat Genet**. 2012;44(5):532-+.
23. Tyrrell J, Huikari V, Christie JT, et al. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5CHRNA3CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. **Human Molecular Genetics**. 2012;21(24):5344-58.
24. van der Valk RJP, Duijts L, Kerkhof M, et al. Interaction of a 17q12 variant with both fetal and infant smoke exposure in the development of childhood asthma-like symptoms. **Allergy**. 2012;67(6): 767-74.
25. Velders FP, Dieleman G, Cents RAM, et al. Variation in the Glucocorticoid Receptor Gene at rs41423247 Moderates the Effect of Prenatal Maternal Psychological Symptoms on Child Cortisol Reactivity and Behavior. **Neuropsychopharmacol**. 2012;37(11):2541-9.
26. Velders FP, De Wit JE, Jansen PW, et al. FTO at rs9939609, Food Responsiveness, Emotional Control and Symptoms of ADHD in Preschool Children. **PLoS One**. 2012;7(11).
27. Cousminer DL, Berry DJ, Timpson NJ, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. **Human Molecular Genetics**. 2013;22(13):2735-47.
28. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. **Nat Genet**. 2013;45(1): 76-U115.
29. Kok R, Bakermans-Kranenburg MJ, van IJzendoorn MH, et al. The role of maternal stress during pregnancy, maternal discipline, and child COMT Val158Met genotype in the development of compliance. **Dev Psychobiol**. 2013;55(5):451-64.
30. Parmar PG, Marsh JA, Taal HR, et al. Polymorphisms in genes within the IGF-axis influence antenatal and postnatal growth. **J Dev Orig Hlth Dis**. 2013;4(2):157-69.
31. Bouthoorn SH, van Lenthe FJ, Kiefte-de Jong JC, et al. Genetic taste blindness to bitter and body composition in childhood: a Mendelian randomization design. **Int J Obesity**. 2014;38(7):1005-10.
32. Gaillard R, Durmus B, Hofman A, Mackenbach JP, Steegers EA, Jaddoe VW. Risk factors and outcomes of maternal obesity and excessive weight gain during pregnancy. **Obesity (Silver Spring)**. 2013;21(5): 1046-55.
33. Bouzigon E, Nadif R, Thompson EE, et al. A common variant in RAB27A gene is associated with fractional exhaled nitric oxide levels in adults. **Clin Exp Allergy**. 2015;45(4):797-806.
34. Benke KS, Nivard MG, Velders FP, et al. A Genome-wide Association Meta-analysis of Preschool Internalizing Problems. **J Am Acad Child Psy**. 2014;53(6):667-76.
35. Benyamin B, St Pourcain B, Davis OS, et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. **Mol Psychiatr**. 2014;19(2):253-8.
36. Bonnelykke K, Sleiman P, Nielsen K, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. **Nat Genet**. 2014;46(1):51-+.
37. Cents RAM, Kok R, Tiemeier H, et al. Variations in maternal 5-HTTLPR affect observed sensitive parenting. **J Child Psychol Psyc**. 2014;55(9):1025-32.
38. van der Knaap NJF, El Marroun H, Klumpers F, et al. Beyond Classical Inheritance: The Influence of Maternal Genotype upon Child's Brain Morphology and Behavior. **J Neurosci**. 2014;34(29):9516-21.
39. Rietveld CA, Esko T, Davies G, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. **P Natl Acad Sci USA**. 2014;111(38):13790-4.
40. St Pourcain B, Cents RA, Whitehouse AJ, et al. Common variation near ROBO2 is associated with expressive vocabulary in infancy. **Nature Communications**. 2014;5:4831.
41. van der Valk RJ, Duijts L, Timpson NJ, et al. Fraction of exhaled nitric oxide values in childhood are associated with 17q11.2-q12 and 17q12-q21 variants. **J Allergy Clin Immunol**. 2014;134(1):46-55.

42. Felix JF, Bradfield JP, Monnereau C, et al. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. **Hum Mol Genet**. 2016;25(2):389-403.
43. Mous SE, Hammerschlag AR, Polderman TJ, et al. A Population-Based Imaging Genetics Study of Inattention/Hyperactivity: Basal Ganglia and Genetic Pathways. **J Am Acad Child Adolesc Psychiatry**. 2015;54(9):745-52.
44. Pappa I, Fedko IO, Mileva-Seitz VR, et al. Single Nucleotide Polymorphism Heritability of Behavior Problems in Childhood: Genome-Wide Complex Trait Analysis. **J Am Acad Child Adolesc Psychiatry**. 2015;54(9):737-44.
45. Fedko IO, Hottenga JJ, Medina-Gomez C, et al. Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. **Behav Genet**. 2015;45(5):514-28.
46. Punwasi RV, Monnereau C, Hofman A, Jaddoe VW, Felix JF. The Influence of Known Genetic Variants on Subclinical Cardiovascular Outcomes in Childhood: Generation R Study. **Circ Cardiovasc Genet**. 2015;8(4):596-602.
47. Qi Q, Downer MK, Kilpelainen TO, et al. Dietary Intake, FTO Genetic Variants, and Adiposity: A Combined Analysis of Over 16,000 Children and Adolescents. **Diabetes**. 2015;64(7):2467-76.
48. van der Valk RJP, Kreiner-Moller E, Kooijman MN, et al. A novel common variant in DCST2 is associated with length in early life and height in adulthood. **Human Molecular Genetics**. 2015;24(4):1155-68.
49. Parmar PG, Taal HR, Timpson NJ, et al. International GWAS Consortium Identifies Novel Loci Associated with Blood Pressure in Children and Adolescents. **Circ Cardiovasc Genet**. 2016.
50. Miliku K, Vogelezang S, Franco OH, Hofman A, Jaddoe VW, Felix JF. Influence of common genetic variants on childhood kidney outcomes. **Pediatr Res**. 2016.
51. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. **Nucleic Acids Res**. 2014;42(Database issue):D1001-6.
52. Medina-Gomez C, Felix JF, Estrada K, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. **Eur J Epidemiol**. 2015;30(4):317-30.
53. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. **PLoS Biol**. 2013;11(9):e1001661.
54. Morris AP. Transethnic meta-analysis of genomewide association studies. **Genet Epidemiol**. 2011;35(8):809-22.
55. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**. 2012;491(7422):56-65.
56. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. **Genome Med**. 2015;7(1):16.
57. Genome of the Netherlands C. Whole-genome sequence variation, population structure and demographic history of the Dutch population. **Nat Genet**. 2014;46(8):818-25.
58. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. **Nature**. 2015;526(7571):75-81.
59. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. **Nature Communications**. 2015;6.
60. McCarthy S, Das S, Kretzschmar W, Durbin R, Abecasis G, Marchini J. A reference panel of 64,976 haplotypes for genotype imputation. **bioRxiv**. 2015.
61. Kreiner-Moller E, Medina-Gomez C, Uitterlinden AG, Rivadeneira F, Estrada K. Improving accuracy of rare variant imputation with a two-step imputation approach. **Eur J Hum Genet**. 2015;23(3):395-400.
62. Liu EY, Li M, Wang W, Li Y. MaCH-admix: genotype imputation for admixed populations. **Genet Epidemiol**. 2013;37(1):25-37.
63. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. **Nat Genet**. 2009;41(11):1199-U58.

64. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. **Nat Genet**. 2012;44(5):491-501.
65. Styrkarsdottir U, Thorleifsson G, Eiriksdottir B, et al. Two Rare Mutations in the COL1A2 Gene Associate With Low Bone Mineral Density and Fractures in Iceland. **J Bone Miner Res**. 2015.
66. Zheng HF, Forgetta V, Hsu YH, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. **Nature**. 2015;526(7571):112-7.
67. Styrkarsdottir U, Thorleifsson G, Gudjonsson SA, et al. Sequence variants in the PTCH1 gene associate with spine bone mineral density and osteoporotic fractures. **Nature Communications**. 2016;7:10129.
68. Styrkarsdottir U, Thorleifsson G, Sulem P, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. **Nature**. 2013;497(7450):517-20.
69. Koller DL, Zheng HF, Karasik D, et al. Meta-analysis of genome-wide studies identifies WNT16 and ESR1 SNPs associated with bone mineral density in premenopausal women. **J Bone Miner Res**. 2013; 28(3):547-58.
70. Chesi A, Mitchell JA, Kalkwarf HJ, et al. A trans-ethnic genome-wide association study identifies gender-specific loci influencing pediatric aBMD and BMC at the distal radius. **Hum Mol Genet**. 2015; 24(17):5053-9.
71. Moayyeri A, Hsu YH, Karasik D, et al. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. **Human Molecular Genetics**. 2014;23(11):3054-68.
72. Correa-Rodriguez M, Schmidt Rio-Valle J, Rueda-Medina B. Polymorphisms of the WNT16 gene are associated with the heel ultrasound parameter in young adults. **Osteoporos Int**. 2015.
73. Zheng HF, Tobias JH, Duncan E, et al. *WNT16* influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. **PLoS Genet**. 2012;8(7):e1002745.
74. Shirts BH, Hasstedt SJ, Hopkins PN, Hunt SC. Evaluation of the gene-age interactions in HDL cholesterol, LDL cholesterol, and triglyceride levels: The impact of the SORT1 polymorphism on LDL cholesterol levels is age dependent. **Atherosclerosis**. 2011;217(1):139-41.
75. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. **PLoS Genet**. 2015;11(10): e1005378.
76. Hancock DB, Artigas MS, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. **PLoS Genet**. 2012;8(12):e1003098.
77. Taylor AE, Morris RW, Fluharty ME, et al. Stratification by smoking status reveals an association of CHRNA5-A3-B4 genotype with body mass index in never smokers. **PLoS Genet**. 2014;10(12):e1004799.
78. Andreasen CH, Stender-Petersen KL, Mogensen MS, et al. Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. **Diabetes**. 2008;57(1):95-101.
79. Demerath EW, Lutsey PL, Monda KL, et al. Interaction of FTO and physical activity level on adiposity in African-American and European-American adults: the ARIC study. **Obesity (Silver Spring)**. 2011;19(9): 1866-72.
80. Paternoster L, Lorentzon M, Lehtimaki T, et al. Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure. **PLoS Genet**. 2013;9(2):e1003247.
81. Kang KS, Robling AG. New Insights into Wnt-Lrp5/6-beta-Catenin Signaling in Mechanotransduction. **Front Endocrinol (Lausanne)**. 2014;5:246.
82. Evans DM. The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between variables. **Am J Hum Genet**. 2002;70(6):1599-602.
83. Lecomte V, Meugnier E, Euthine V, et al. A new role for sterol regulatory element binding protein 1 transcription factors in the regulation of muscle mass and muscle cell differentiation. **Mol Cell Biol**. 2010;30(5):1182-98.
84. Dessalle K, Euthine V, Chanon S, et al. SREBP-1 transcription factors regulate skeletal muscle cell size by controlling protein synthesis through myogenic regulatory factors. **PLoS One**. 2012;7(11):e50878.

85. Livshits G, Malkin I, Moayyeri A, Spector TD, Hammond CJ. Association of FTO gene variants with body composition in UK twins. **Ann Hum Genet**. 2012;76:333-41.
86. Korostishevsky M, Steves CJ, Malkin I, Spector T, Williams FM, Livshits G. Genomics and metabolomics of muscular mass in a community-based sample of UK females. **Eur J Hum Genet**. 2016;24(2): 277-83.
87. Nam HS, Shin MH, Zmuda JM, et al. Race/ethnic differences in bone mineral densities in older men. **Osteoporos Int**. 2010;21(12):2115-23.
88. Horlick M, Thornton J, Wang J, Levine LS, Fedun B, Pierson RN, Jr. Bone mineral in prepubertal children: gender and ethnicity. **J Bone Miner Res**. 2000;15(7):1393-7.
89. Marshall LM, Zmuda JM, Chan BKS, et al. Race and ethnic variation in proximal femur structure and BMD among older men. **Journal of Bone and Mineral Research**. 2008;23(1):121-30.
90. Henry YM, Eastell R. Ethnic and gender differences in bone mineral density and bone turnover in young adults: effect of bone size. **Osteoporos Int**. 2000;11(6):512-7.
91. Micklesfield LK, Norris SA, Pettifor JM. Determinants of bone size and strength in 13-year-old South African children: the influence of ethnicity, sex and pubertal maturation. **Bone**. 2011;48(4):777-85.
92. Nam HS, Kweon SS, Choi JS, et al. Racial/ethnic differences in bone mineral density among older women. **J Bone Miner Metab**. 2013;31(2):190-8.
93. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. **Nat Rev Genet**. 2005;6(4):333-40.
94. Nowlan NC, Jepsen KJ, Morgan EF. Smaller, weaker, and less stiff bones evolve from changes in subsistence strategy. **Osteoporos Int**. 2011;22(6):1967-80.
95. Chirchir H, Kivell TL, Ruff CB, et al. Recent origin of low trabecular bone density in modern humans. **Proc Natl Acad Sci U S A**. 2015;112(2):366-71.
96. Ryan TM, Shaw CN. Gracility of the modern Homo sapiens skeleton is the result of decreased biomechanical loading. **P Natl Acad Sci USA**. 2015;112(2):372-7.
97. Gomez F, Wang L, Abel H, Zhang Q, Province MA, Borecki IB. Admixture mapping of coronary artery calcification in African Americans from the NHLBI family heart study. **BMC Genet**. 2015;16:42.
98. Li YR, Keating BJ. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. **Genome Med**. 2014;6(10):91.
99. Liu CT, Estrada K, Yerges-Armstrong LM, et al. Assessment of gene-by-sex interaction effect on bone mineral density. **J Bone Miner Res**. 2012;27(10):2051-64.
100. Schlecht SH, Bigelow EM, Jepsen KJ. How Does Bone Strength Compare Across Sex, Site, and Ethnicity? **Clin Orthop Relat Res**. 2015;473(8):2540-7.
101. Medina-Gomez C, Heppe DH, Yin JL, et al. Bone Mass and Strength in School Age Children Exhibit Sexual Dimorphism Related to Differences in Lean Mass: The Generation R Study. **J Bone Miner Res**. 2015.
102. Moon RJ, Cole ZA, Crozier SR, et al. Longitudinal changes in lean mass predict pQCT measures of tibial geometry and mineralisation at 6-7 years. **Bone**. 2015;75:105-10.
103. Detter F, Rosengren BE, Dencker M, Lorentzon M, Nilsson JA, Karlsson MK. A 6-year exercise program improves skeletal traits without affecting fracture risk: a prospective controlled study in 2621 children. **J Bone Miner Res**. 2014;29(6):1325-36.
104. Bonjour JP, Chevalley T, Ferrari S, Rizzoli R. The importance and relevance of peak bone mass in the prevalence of osteoporosis. **Salud Publica Mex**. 2009;51 Suppl 1:S5-17.
105. Warden SJ, Mantila Roosa SM, Kersh ME, et al. Physical activity when young provides lifelong benefits to cortical bone size and strength in men. **Proc Natl Acad Sci U S A**. 2014;111(14):5337-42.
106. Heppe DH, Medina-Gomez C, Hofman A, Franco OH, Rivadeneira F, Jaddoe VW. Maternal first-trimester diet and childhood bone mass: the Generation R Study. **Am J Clin Nutr**. 2013;98(1):224-32.
107. Chango A, Pogribny IP. Considering Maternal Dietary Modulators for Epigenetic Regulation and Programming of the Fetal Epigenome. **Nutrients**. 2015;7(4):2748-70.
108. Choi SW, Friso S. Epigenetics: A New Bridge between Nutrition and Health. **Adv Nutr**. 2010;1(1):8-16.

109. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. **Br J Nutr**. 2007;97(6): 1064-73.
110. van den Hooven EH, Heppe DH, Kiefte-de Jong JC, et al. Infant dietary patterns and bone mass in childhood: the Generation R Study. **Osteoporos Int**. 2015;26(5):1595-604.
111. Moverare-Skrtic S, Henning P, Liu X, et al. Osteoblast-derived WNT16 represses osteoclastogenesis and prevents cortical bone fragility fractures. **Nat Med**. 2014;20(11):1279-88.
112. Moverare-Skrtic S, Wu J, Henning P, et al. The bone-sparing effects of estrogen and WNT16 are independent of each other. **Proc Natl Acad Sci U S A**. 2015;112(48):14972-7.
113. Baron R, Hesse E. Update on bone anabolics in osteoporosis treatment: rationale, current status, and perspectives. **J Clin Endocrinol Metab**. 2012;97(2):311-25.
114. Pathology Department UR. Pathways of Human Disease. Rochester, NY: University of Rochester Medical Center 2015.
115. Paternoster L, Lorentzon M, Vandenput L, et al. Genome-wide association meta-analysis of cortical bone mineral density unravels allelic heterogeneity at the RANKL locus and potential pleiotropic effects on bone. **PLoS Genet**. 2010;6(11):e1001217.
116. Verroken C, Zmierczak HG, Goemaere S, Kaufman JM, Lapauw B. Association of Jumping Mechanography-Derived Indices of Muscle Function with Tibial Cortical Bone Geometry. **Calcif Tissue Int**. 2015.
117. Ay L, Jaddoe VW, Hofman A, et al. Foetal and postnatal growth and bone mass at 6 months: the Generation R Study. **Clin Endocrinol (Oxf)**. 2011;74(2):181-90.
118. Warrington NM, Kemp JP, Tilling K, Tobias JH, Evans DM. Genetic variants in adult bone mineral density and fracture risk genes are associated with the rate of bone mineral density acquisition in adolescence. **Hum Mol Genet**. 2015;24(14):4158-66.
119. Mitchell JA, Chesi A, Elci O, et al. Genetic Risk Scores Implicated in Adult Bone Fragility Associate With Pediatric Bone Density. **J Bone Miner Res**. 2015.
120. Mitchell JA, Chesi A, Elci O, et al. Genetics of Bone Mass in Childhood and Adolescence: Effects of Sex and Maturation Interactions. **J Bone Miner Res**. 2015;30(9):1676-83.
121. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. **Nature**. 2012; 489(7414):57-74.
122. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. **Science**. 2015;348(6235):648-60.
123. Jaddoe VW, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. **Eur J Epidemiol**. 2012;27(9):739-56.
124. Bouwland-Both MI, van Mil NH, Tolhoek CP, et al. Prenatal parental tobacco smoking, gene specific DNA methylation, and newborns size: the Generation R study. **Clin Epigenetics**. 2015;7(1):83.
125. Steves CJ, Bird S, Williams FM, Spector TD. The Microbiome and Musculoskeletal Conditions of Aging: A Review of Evidence for Impact and Potential Therapeutics. **J Bone Miner Res**. 2016;31(2):261-9.
126. Ohlsson C, Sjogren K. Effects of the gut microbiota on bone mass. **Trends Endocrin Met**. 2015;26(2): 69-74.
127. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. **Nat Rev Genet**. 2015;16(2):85-97.



Chapter 7

Summary/Samenvatting

SUMMARY

Bone accrual during childhood and adolescence largely determines the peak (maximal) bone mass achieved approximately in the third decade of human life. This process is essential to assess the risk of fracture throughout life. Therefore, understanding the factors influencing bone accrual is crucial to design strategies aiming to prevent the early onset of osteoporosis and low bone mass later in life. To date, a considerable amount of literature has been published on lifestyle choices that influence peak bone mass. Physical activity and infant calcium intake are the strongest factors for which there is evidence of a positive effect on bone growth and mineralization. In addition, genetic studies in families have demonstrated that more than 50% of the variability of bone mineral density (BMD) is attributable to genetic variants and that this heritability tends to decrease with age, probably as a result of the accumulation of environmental perturbations over many years.

As described in **Chapter 1**, throughout this thesis we aimed to disentangle the multifactorial process of bone accretion in children by assessing the influence of factors like genetic background, fetal exposures and infant body size and composition on bone status in children at school age. The studies of this thesis are embedded in the Generation R Study, a multiethnic population-based prospective cohort study from fetal life onwards, based in Rotterdam, The Netherlands.

In **Chapter 2**, the methodological framework to perform genome-wide association studies (GWAS) in admixed populations is detailed. In **Chapter 2.1**, an exhaustive description of the Generation R genetic data generation and possible analysis procedures to cope with its multiethnic nature. The generation of this data for 5,732 participants of Generation R has contributed already to at least 50 publications in fields as diverse as kidney research, childhood growth or infant behavioral problems. In **Chapter 2.2**, we propose a imputation method which adds an intermediate step, the imputation to a local reference panel, prior to the imputation to a denser reference set from broader geographic origin (e.g., two-step imputation). This method improved in 28% the quality of the imputation, representing an increase of 18% in the number of genetic markers to be surveyed by GWAS studies, compared to the traditional one-step imputation method. In **Chapter 2.3**, we use the Generation R genetic data from an admixed background as a suitable set of controls for the study of genetic factors influencing the risk of the low prevalent Behçet's disease in which cases were from different genetic ancestries. By using linear mixed models and principal component adjustment strategies, we were able to replicate the association of variants in the HLA locus, already reported in homogeneous populations of Asian origin, as well as identify a novel association mapping to an intronic region in *IL2A*.

Chapter 3 is dedicated to epidemiological studies of pediatric BMD. In **Chapter 3.1**, we show that factors influencing bone status at school age date back to fetal life. Maternal phosphorous intake and homocysteine concentrations strongly influence the amount of bone mass at six years, whilst vitamin B-12 concentration and protein intake predicted BMD levels. In **Chapter 3.2**, we prove that both fetal growth and infant growth contribute independently to bone mass at six years of age, with the first year of age being the most critical period in the contribution. In **Chapter 3.3**, we make emphasis on bone geometry, another parameter that together with bone mass determines bone strength. Our results indicate that on average the femur of boys has better resistance to axial and bending forces in comparison to girls. However, these sex differences were partially explained by differences in the amount of muscle mass between boys and girls. Considering the evidence of this and other studies, we promote the establishment of exercise programs during the pre- and peri-pubertal years, particularly for girls whose bones have a relatively "weaker" configuration.

In **Chapter 4**, we focus on genetic studies of pediatric BMD. In **Chapter 4.1**, we identified a strong association signal for pediatric BMD in the 7q31.31 locus. *WNT16* and *CPED1* are the genes underlying this association and different variants mapping in these two genes influence both total body less head (TBLH-) BMD and skull BMD. By means of conditional analysis, we found that at least two different signals arise from this genomic region and that their effect on skull BMD could vary with aging. Most importantly, Wnt16 knockout (KO) mice had reduced total body BMD and gene expression profiles in human bone biopsies support a role of CPED1 and WNT16 on the BMD phenotypes observed at the human population level. Altogether we provided different lines of evidence for the role of these molecules in the process of bone mineralization. In **Chapter 4.2**, we demonstrate that TB-BMD is a relevant trait for genetic studies of bone quality not only in children but also in the general population and thus suitable for the study of osteoporosis. We identified variants known to influence different bone compartments at different skeletal sites. Moreover, we identified 18 novel genetic loci influencing BMD variation: 1p13.3 (*CSF1*), 2p11.2 (*TCF7L1*), 2p22.1 (*SLC8A1*), 2q33.1 (*FZD7*), 5q22.2 (*APC*), 5q23.2 (*CNKIG3*), 7p14.3 (*AQP1*), 10q26.13 (*BTBD16*), 11p13 (*LDLRAD3*), 11p15.5 (*RIC8A*), 12q21.33 (*ATP2B1*), 12q21.33 (*BTG1*), 15q14 (*TMCO5A*), 15q22.33 (*SMAD3*), 16q22.3 (*ZFHX3*), 17q24.1 (*CEP112*), 20q12 (*MAFB*) and 21q22.12 (*RUNX1*). Several of these regions code for factors embedded in Wnt signaling, yet the possible role of others remain to be elucidated. In **Chapter 4.3**, we analyse the same TB-BMD data but stratifying by age-bins of 15 years. This manner of analysing the data reached three additional association signals. One mapping to the 14q32.12 (*RIN3*) evident only in children, another one in 11p15.1 (*PTPN5*) evident in the 30-45 years group, and the third one in 7q22.1 (*ZNF789/ZSCAN5*) evident only in the old adults (e.g., >60 years). Moreover, we found variants in different loci whose effect magnitude depends on age, most of them with

larger effects in children. The strongest evidence of this varying effect with age came from variants located in 14q32.12 (*RIN3*) and 6q25.1 (*ESR1/CCDC70*). In **Chapter 4.4**, based on pediatric populations we pinpoint genetic factors whose relative contribution to explain BMD variability was uneven across skeletal sites. The genetic correlation between appendicular sites was higher than the one reported for each of these sites and the skull. We described thirteen different loci associated with BMD in one or more skeletal sites (i.e. SNPs at *WNT4, GALNT3, CPED1/WNT16/FAM3C, RSPO3, FUBP3, PTHLH, TNFSF11, TNFRSF11B, TNFRSF11A, LRP5, LGR4, RIN3* and *EYA4*). In **Chapter 4.5**, we found evidence for genetic variants within or nearby *SREBF1* exert effects on bone and muscle tissues simultaneously and in opposite directions. Using publicly available databases we compiled evidence that these SNPs could regulate the expression of *SREBF1*, whose product SREBP-1 indeed downregulates muscle cell differentiation, while also increasing the number of mineralized foci in osteoblasts.

In **Chapter 5**, we confirmed that at school age children from African ancestry have higher BMD levels than their counterparts from European or Asian ancestry. We found that the combined effect of variants reported in literature as associated with BMD in adults was significantly associated with pediatric BMD, suggesting that many of these variants exert effects in bone since early age. Children from African background were found to carry a higher number of BMD-increasing variants than children of non-African background, suggesting that the known BMD ethnic differences are at least partially from genetic origin. This result was parallel by comparing genetic data from different populations worldwide. Moreover, after applying multiple cutting-edge evolutionary analyses, we found evidence pointing to natural selection as the engine driving these genetic patterns.

Finally, in **Chapter 6**, the results are interconnected and discussed more generally. Also, methodological issues regarding the studies contained in this thesis are elaborated. Implications of our findings and future directions for research are also presented in this chapter.

SAMENVATTING

Botvorming en groei bij kinderen bepaald grotendeels de maximale hoeveelheid botmassa die wordt verkregen, meestal in het derde decennium, tijdens het menselijke leven. Dit proces is essentieel voor het risico op botfracturen tijdens het leven. Het is daarom cruciaal om te begrijpen welke factoren een rol spelen bij deze botvorming, en hieruit strategieën te bedenken om de eerste symptomen van osteoporose en lage bot dichtheid als volwassene te voorkomen. Een aanzienlijke hoeveelheid literatuur is gepubliceerd over de keuzes in leefstijl die de maximale botmassa beïnvloeden. Gezond bewegen en voldoende calcium inname als kind zijn de sterkste bekende factoren die botgroei en mineralisatie positief beïnvloeden. Daarnaast hebben genetische studies in families laten zien dat meer dan 50% van de variatie in bot-mineraal dichtheid (BMD) toegewezen kan worden aan genetische variaties, en dat deze erfelijkheid toeneemt met leeftijd, waarschijnlijk vanwege een opsomming van verstorende omgevingsfactoren tijdens het leven. Zoals beschreven in **Hoofdstuk 1**, streven wij in dit proefschrift naar het ontleden van het multifactoriële proces van botvorming bij kinderen door de invloed van factoren als genetische achtergrond, blootstellingen van buitenaf aan de foetus, en de grootte en samenstelling van het lichaam als kind op de status van het bot bij schoolgaande kinderen. De studies uit dit proefschrift maken onderdeel uit van de Generation R Studie, een multi-etnische prospectieve populatie cohort studie vanaf foetaal leven, opgezet in Rotterdam, Nederland.

In **Hoofdstuk 2** wordt de methodologische benadering van genoom-wijde associatie studies (GWAS) in gemengde populaties in detail beschreven. In **Hoofdstuk 2.1** volgt een uitgebreide beschrijving van het generen van de genetische data voor Generation R en worden verschillende analysemethoden voorgesteld om met de multi-ethnische samenstelling van deze dataset om te gaan. Deze genetische data van 5,732 deelnemers aan Generation R heeft inmiddels al bijgedragen aan ten minste 50 wetenschappelijke publicaties in onderzoeksgebieden als nieronderzoek, groei bij kinderen of gedragsproblemen bij jonge kinderen. In **Hoofdstuk 2.2** stellen wij een imputatie methode voor die gebruik maakt van een tussenstap, imputatie naar een lokaal referentiepanel, voor de imputatie naar een geografisch uitgebreidere referentie dataset (zogenaamde "two-step imputation"). Deze methode verbeterde de imputatie in 28% van de gevallen, met daardoor 18% meer genetische markers tijdens de GWAS studie, in vergelijking met de traditionele "one-step imputation" methode. In **Hoofdstuk 2.3** gebruiken wij de genetische data uit de gemengde populatie van Generation R als controle dataset voor het bestuderen van de genetische factoren die het risico op de ziekte van Behçet beinvloeden, waarbij de aangedane mensen van verschillende genetische achtergronden kwamen. Door middel van "linear mixed models" en correctie van principale componenten waren wij in staat om de associatie van varianten in de HLA locus, welke al eerder gerapporteerd waren in homogene Aziatische

populaties, te bevestigen, maar ook een nieuwe associatie in het intronische gebied van *IL2A* te identificeren.

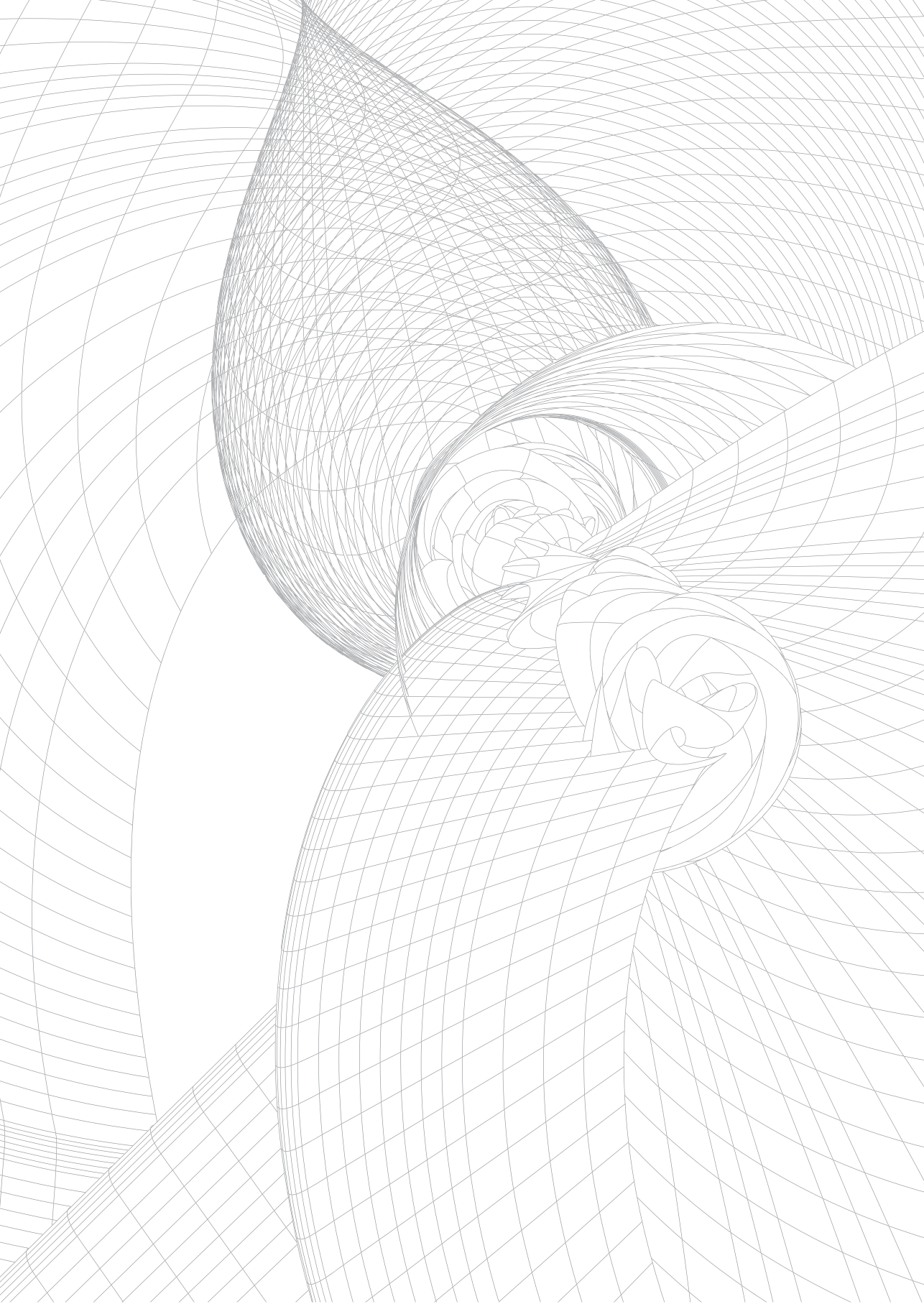
Hoofdstuk 3 is toegewijd aan epidemiologische studies bij pediatrische BMD. In **Hoofdstuk 3.1** laten we zien dat de factoren die botstatus bij schoolgaande kinderen beïnvloeden terug te herleiden zijn tot de periode als foetus. De inname van fosfor van de moeder en homocysteïne concentraties beïnvloeden sterk de hoeveelheid botmassa op 6-jarige leeftijd, terwijl vitamine B-12 concentraties en eiwit inname BMD niveaus voorspellen. In **Hoofdstuk 3.2** laten we zien dat zowel foetale groei en de groei als klein kind onafhankelijk bijdragen aan de botmassa op 6-jarige leeftijd, waarbij het eerste levensjaar de meest kritieke bijdrage levert. In **Hoofdstuk 3.3** leggen we de nadruk op de bot-geometrie, een andere parameter die samen met botmassa de sterkte van het bot bepaald. Onze resultaten laten zien dat het dijbeen (de "femur") van jongens gemiddeld een hogere weerstand heeft tegen axiale en buigkrachten dan dat van meisjes. Een deel van deze verschillen werd echter verklaard door het verschil in spiermassa tussen jongens en meisjes. Met dit resultaat, en andere studies in acht nemend, moedigen wij het oprichten van sportprogramma's voor en tijdens de puberteit aan, vooral voor meisjes waarbij het bot een relatief "zwakkere" configuratie heeft.

In **Hoofdstuk 4** gaan we in op genetische studies bij pediatrische BMD. In **Hoofdstuk 4.1** hebben we een sterke associatie gevonden voor pediatrische BMD in de 7q31.31 locus. *WNT16* en *CPED1* zijn de genen die deze associatie teweegbrengen, en verschillende varianten in deze twee genen beïnvloeden zowel BMD van het hele lichaam (zonder hoofd; TBLH-BMD) als van de schedel (Skull-BMD). Door middel van conditionele analyses hebben we ten minste twee verschillende signalen gevonden in dit genomische gebied, waarbij hun effect op BMD van de schedel met leeftijd kan variëren. Belangrijker, een Wnt16 knockout muis had een verlaagde ("total body"-BMD) TB-BMD, en gen expressie profielen in humane bot biopsies ondersteunen een rol voor CPED1 en WNT16 bij de BMD fenotypen op een populatie niveau. Alles bij elkaar genomen hebben we meerdere aanwijzingen gevonden voor de rol van deze moleculen in het botmineralisatie proces. In **Hoofdstuk 4.2** demonstreren we dat TB-BMD niet alleen relevant is voor genetische studies bij botkwaliteit bij kinderen, maar ook toepasbaar voor het bestuderen van osteoporose in de hele populatie. We hebben varianten gevonden waarvan bekend is dat zij invloed hebben op verschillende boteigenschappen in verschillende gebieden op het skelet. Daarnaast hebben wij 18 nieuwe genetische gebieden gevonden die invloed hebben op de variatie van BMD: 1p13.3 (*CSF1*), 2p11.2 (*TCF7L1*), 2p22.1 (*SLC8A1*), 2q33.1 (*FZD7*), 5q22.2 (*APC*), 5q23.2 (*CNK1G3*), 7p14.3 (*AQP1*), 10q26.13 (*BTBD16*), 11p13 (*LDLRAD3*), 11p15.5 (*RIC8A*), 12q21.33 (*ATP2B1*), 12q21.33 (*BTG1*), 15q14 (*TMCO5A*), 15q22.33 (*SMAD3*), 16q22.3 (*ZFHX3*), 17q24.1 (*CEP112*), 20q12 (*MAFB*) en 21q22.12 (*RUNX1*). Verschillende van deze

gebieden coderen voor onderdelen van de Wnt signaling pathway, terwijl de mogelijke rol van anderen nog nagegaan moet worden. In **Hoofdstuk 4.3** hebben we dezelfde TB-BMD data geanalyseerd in leeftijd categorieen van 15 jaar. Deze manier van analyseren leverde nog eens drie additionele signalen op. Een daarvan op 14q32.12 (*RIN3*) was alleen gevonden bij kinderen, een andere in 11p15.1 (*PTPN5*) alleen in de categorie van 30-45 jaar, en de laatste in 7q22.1 (*ZNF789/ZSCAN5*) alleen de in de oudste categorie (60 jaar of ouder). Daarnaast hebben we varianten in verschillende regios gevonden waarvan het effect afhankelijk is van de leeftijd, meestal met sterkere effecten in kinderen. Het sterkste voorbeeld van dit varierende effect kwam van de varianten in 14q32.12 (*RIN3*) en 6q25.1 (*ESR1/CCDC70*). In **Hoofdstuk 4.4** bekijken we genetische factoren waarvan het effect op BMD verschilde tussen skeletale gebieden in onze pediatrische populaties. De genetische correlatie was hoger tussen skeletale gebieden in verschillende ledematen dan tussen deze ledematen en de schedel. We beschrijven dertien verschillende regios die geassocieerd zijn met BMD in een of meerdere skeletale gebieden (bijvoorbeeld SNPs in *WNT4, GALNT3, CPED1/WNT16/FAM3C, RSPO3, FUBP3, PTHLH, TNFSF11, TNFRSF11B, TNFRSF11A, LRP5, LGR4, RIN3* en *EYA4*). In **Hoofdstuk 4.5** laten we zien dat genetische variaties in of bij *SREBF1* op hetzelfde moment een tegengesteld effect kunnen hebben op bot en spierweefsel. Met behulp van publieke databases hebben we bewijs verzameld dat deze SNPs de expressie van *SREBF1* kunnen reguleren, waarbij het product SREBP-1 inderdaad spiercel differentiatie onderdrukt, terwijl het het aantal gemineraliseerde foci in osteoblasten verhoogd.

In **Hoofdstuk 5** bevestigen we dat schoolgaande kinderen van Afrikaanse afkomst hogere BMD levels hebben dan kinderen van Europese of Aziatische afkomst. We vonden dat het gecombineerde effect van de varianten die in de literatuur geassocieerd zijn met BMD in volwassenen significant geassocieerd was met pediatric BMD, wat suggereerd dat veel van deze varianten hun effect al uitoefenen op jonge leeftijd. Kinderen van Afrikaanse afkomst hadden een groter aantal BMD-verhogende varianten dan kinderen van andere afkomsten, suggererend dat de bekende ethnische verschillen in BMD ten minste deels van genetische oorsprong zijn. Dit resultaat was hetzelfde na het bestuderen van populaties over de hele wereld. Daarnaast, na het uitvoeren van verschillende cutting-edge evolutionaire analyses, vonden we bewijs dat natuurlijke selectie aanwijst als de veroorzaker van deze genetische patronen.

Uiteindelijk worden de resultaten in **Hoofdstuk 6** en in meer algemene zin besproken. Ook worden de methodologische aspecten met betrekking tot de studies in dit proefschrift toegelicht, en worden de implicaties van onze bevindingen en suggesties voor verder onderzoek in dit hoofdstuk gepresenteerd.



Chapter 8

Appendices

LIST OF PUBLICATIONS

PUBLICATIONS BASED ON THE STUDIES DESCRIBED IN THIS THESIS

*Joint First Author

Chapter 2.1

- **Medina-Gomez C**, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, Duijts L, Hofman A, van Duijn CM, Uitterlinden AG, Jaddoe VW, Rivadeneira F. 2015. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. ***Eur J Epidemiol.*** Apr;30(4):317-30.

Chapter 2.2

- Kreiner-Møller E, **Medina-Gomez C**, Uitterlinden AG, Rivadeneira F, Estrada K. 2015. Improving accuracy of rare variant imputation with a two-step imputation approach. ***Eur J Hum Genet***. Mar;23(3):395-400.

Chapter 2.3

- Kappen JH, **Medina-Gomez C**, van Hagen PM, Stolk L, Estrada K, Rivadeneira F, Uitterlinden AG, Stanford MR, Ben-Chetrit E, Wallace GR, Soylu M, van Laar JA. 2015. Genome-wide association study in an admixed case series reveals IL12A as a new candidate in Behçet disease. ***PLoS One.*** Mar 23;10(3):e0119085.

Chapter 3.1

- Heppe DH**, Medina-Gomez C**, Hofman A, Franco OH, Rivadeneira F, Jaddoe VW. 2013. Maternal first-trimester diet and childhood bone mass: the Generation R Study. ***Am J Clin Nutr***. 98: 224-232.

Chapter 3.2

- Heppe DH, **Medina-Gomez C**, Hofman A, Rivadeneira F, Jaddoe VW. 2015. Does fetal smoke exposure affect childhood bone mass? The Generation R Study. ***Osteoporos Int***. Apr;26(4):1319-29.

Chapter 3.3

- **Medina-Gomez C**, Heppe DH, Yin JL, Trajanoska K, Uitterlinden AG, Beck TJ, Jaddoe VW, Rivadeneira F. 2015. Bone Mass and Strength in School Age Children Exhibit Sexual Dimorphism Related to Differences in Lean Mass: The Generation R Study. ***J Bone Miner Res***. 2015 Nov 24.

Chapter 4.1

- **Medina-Gomez C***, Kemp JP*, Estrada K, Eriksson J, Liu J, Reppe S, Evans DM, Heppe DH, Vandenput L, Herrera L, Ring SM, Kruithof CJ, Timpson NJ, Zillikens MC, Olstad OK, Zheng HF, Richards JB, St Pourcain B, Hofman A, Jaddoe VW, Smith GD, Lorentzon M, Gautvik KM, Uitterlinden AG, Brommage R, Ohlsson C, Tobias JH, Rivadeneira F.2012. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. ***PLoS Genet***. 8(7):e1002718.

Chapter 4.2

- **Medina-Gomez C***, Kemp JP*, Chesi A, Kreiner-Møller E, Ahluwalia T, Mook D, Liu Y, Hartwig FP, Evans D, Joro R, van Duijn CM, Nedeljkovic I, Mullin B, Eriksson J, Richards B, Jackson R, Karasik D, Van der Velde N, Ikram A, Zemel BS, Harris T, Zhou Y, Robins J, Li R, Psaty B, Nielson C, van der Eerden B, van de Peppel

J, Scott W, Horta BL, Lakka T, Grant SF, McGuigan F, Wilson J, Styrkársdóttir U, Koller D, Zhu K, Kiel D, Ohlsson C, Ackert-Bicknell CL, Ntzani E, Evangelis E, Uitterlinden AG, Jaddoe VW, Tobias JH, Dave M. Evans, Rivadeneira F. Genetic studies of TB-BMD yield eighteen novel loci involved in bone biology. ***(In preparation)***

Chapter 4.3

- **Medina-Gomez C***, Kemp JP*, Chesi A, Kreiner-Møller E, Ahluwalia T, Mook D, Liu Y, Hartwig FP, Evans D, Joro R, van Duijn CM, Nedeljkovic I, Mullin B, Eriksson J, Richards B, Jackson R, Karasik D, Van der Velde N, Ikram A, Zemel BS, Harris T, Zhou Y, Robins J, Li R, Psaty B, Nielson C, van der Eerden B, van de Peppel J, Scott W, Horta BL, Lakka T, Grant SF, McGuigan F, Wilson J, Styrkársdóttir U, Koller D, Zhu K, Kiel D, Ohlsson C, Ackert-Bicknell CL, Ntzani E, Evangelis E, Uitterlinden AG, Jaddoe VW, Tobias JH, Dave M. Evans, Rivadeneira F. A Large-Scale Genome-Wide Investigation of Age-dependent Effects in Genetic Associations with TB-BMD. ***(In preparation)***

Chapter 4.4

- Kemp JP*, **Medina-Gomez C***, Estrada K, St Pourcain B, Heppe DH, Warrington NM, Oei L, Ring SM, Kruithof CJ, Timpson NJ, Wolber LE, Reppe S, Gautvik K, Grundberg E, Ge B, van der Eerden B, van de Peppel J, Hibbs MA, Ackert-Bicknell CL, Choi K, Koller DL, Econs MJ, Williams FM, Foroud T, Zillikens MC, Ohlsson C, Hofman A, Uitterlinden AG, Davey Smith G, Jaddoe VW, Tobias JH, Rivadeneira F, Evans DM. 2014. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. ***PLoS Genet*** .10(6):e1004423.

Chapter 4.5

- **Medina-Gomez C**, Kemp JP, Dimou NL, Kreiner-Møller E, Chesi A, Heppe DHM, Zemel BS, Bønnelykke K, Boer C, Bisggard H, Evangelou E, Bonewald LF, Gorski JP, Maurano MT, van der Eerden B, Ackert-Bicknell CL, Karasik D, van de Peppel J, Jaddoe VWV, Uitterlinden AG, Tobias JH, Duque G, Grant SF, Bagos PG, Evans DM, Rivadeneira F. Bivariate genome-wide association analysis unveil the pleiotropic effects of *SREBF1* on BMD and lean mass in children. ***(In Preparation)***

Chapter 5

- **Medina-Gómez C**, Chesi A, Heppe DH, Zemel BS, Yin JL, Kalkwarf HJ, Hofman A, Lappe JM, Kelly A, Kayser M, Oberfield SE, Gilsanz V, Uitterlinden AG, Shepherd JA, Jaddoe VW, Grant SF, Lao O, Rivadeneira F. 2015. BMD Loci Contribute to Ethnic And Developmental Differences in Skeletal Fragility Across Populations: Assessment of Evolutionary Selection Pressures. ***Mol Biol Evol***. Nov;32(11):2961-72.

OTHER PUBLICATIONS

- Kemp JP*, **Medina-Gomez C***, , Rivadeneira F, Tobias JH, Evans DM. 2016. The case for genome-wide association studies of bone acquisition in paediatric and adolescent populations. **BoneKEy**. *In press*.
- Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, Cousminer DL, Dastani Z, Drong AW, Esko T, Evans DM, Falchi M, Feitosa MF, Ferreira T, Hedman ÅK, Haring R, Hysi PG, Iles MM, Justice AE, Kanoni S, Lagou V, Li R, Li X, Locke A, Lu C, Mägi R, Perry JR, Pers TH, Qi Q, Sanna M, Schmidt EM, Scott WR, Shungin D, Teumer A, Vinkhuyzen AA, Walker RW, Westra HJ, Zhang M, Zhang W, Zhao JH, Zhu Z, Afzal U, Ahluwalia TS, Bakker SJ, Bellis C, Bonnefond A, Borodulin K, Buchman AS, Cederholm T, Choh AC, Choi HJ, Curran JE, de Groot LC, De Jager PL, Dhonukshe-Rutten RA, Enneman AW, Eury E, Evans DS, Forsen T, Friedrich N, Fumeron F, Garcia ME, Gärtner S, Han BG, Havulinna AS, Hayward C, Hernandez D, Hillege

H, Ittermann T, Kent JW, Kolcic I, Laatikainen T, Lahti J, Mateo Leach I, Lee CG, Lee JY, Liu T, Liu Y, Lobbens S, Loh M, Lyytikäinen LP, **Medina-Gomez C**, Michaëlsson K, Nalls MA, Nielson CM, Oozageer L, Pascoe L, Paternoster L, Polašek O, Ripatti S, Sarzynski MA, Shin CS, Narančić NS, Spira D, Srikanth P, Steinhagen-Thiessen E, Sung YJ, Swart KM, Taittonen L, Tanaka T, Tikkanen E, van der Velde N, van Schoor NM, Verweij N, Wright AF, Yu L, Zmuda JM, Eklund N, Forrester T, Grarup N, Jackson AU, Kristiansson K, Kuulasmaa T, Kuusisto J, Lichtner P, Luan J, Mahajan A, Männistö S, Palmer CD, Ried JS, Scott RA, Stancáková A, Wagner PJ, Demirkan A, Döring A, Gudnason V, Kiel DP, Kühnel B, Mangino M, Mcknight B, Menni C, O'Connell JR, Oostra BA, Shuldiner AR, Song K, Vandenput L, van Duijn CM, Vollenweider P, White CC, Boehnke M, Boettcher Y, Cooper RS, Forouhi NG, Gieger C, Grallert H, Hingorani A, Jørgensen T, Jousilahti P, Kivimaki M, Kumari M, Laakso M, Langenberg C, Linneberg A, Luke A, Mckenzie CA, Palotie A, Pedersen O, Peters A, Strauch K, Tayo BO, Wareham NJ, Bennett DA, Bertram L, Blangero J, Blüher M, Bouchard C, Campbell H, Cho NH, Cummings SR, Czerwinski SA, Demuth I, Eckardt R, Eriksson JG, Ferrucci L, Franco OH, Froguel P, Gansevoort RT, Hansen T, Harris TB, Hastie N, Heliövaara M, Hofman A, Jordan JM, Jula A, Kähönen M, Kajantie E, Knekt PB, Koskinen S, Kovacs P, Lehtimäki T, Lind L, Liu Y, Orwoll ES, Osmond C, Perola M, Pérusse L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Rivadeneira F, Rudan I, Salomaa V, Sørensen TI, Stumvoll M, Tönjes A, Towne B, Tranah GJ, Tremblay A, Uitterlinden AG, van der Harst P, Vartiainen E, Viikari JS, Vitart V, Vohl MC, Völzke H, Walker M, Wallaschofski H, Wild S, Wilson JF, Yengo L, Bishop DT, Borecki IB, Chambers JC, Cupples LA, Dehghan A, Deloukas P, Fatemifar G, Fox C, Furey TS, Franke L, Han J, Hunter DJ, Karjalainen J, Karpe F, Kaplan RC, Kooner JS, McCarthy MI, Murabito JM, Morris AP, Bishop JA, North KE, Ohlsson C, Ong KK, Prokopenko I, Richards JB, Schadt EE, Spector TD, Widén E, Willer CJ, Yang J, Ingelsson E, Mohlke KL, Hirschhorn JN, Pospisilik JA, Zillikens MC, Lindgren C, Kilpeläinen TO, Loos RJ. 2016. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. ***Nat Commun***. Feb 1;7:10495.

- Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, Czajkowski J, Esko T, Fall T, Kilpeläinen TO, Lu Y, Mägi R, Mihailov E, Pers TH, Rüeger S, Teumer A, Ehret GB, Ferreira T, Heard-Costa NL, Karjalainen J, Lagou V, Mahajan A, Neinast MD, Prokopenko I, Simino J, Teslovich TM, Jansen R, Westra HJ, White CC, Absher D, Ahluwalia TS, Ahmad S, Albrecht E, Alves AC, Bragg-Gresham JL, de Craen AJ, Bis JC, Bonnefond A, Boucher G, Cadby G, Cheng YC, Chiang CW, Delgado G, Demirkan A, Dueker N, Eklund N, Eiriksdottir G, Eriksson J, Feenstra B, Fischer K, Frau F, Galesloot TE, Geller F, Goel A, Gorski M, Grammer TB, Gustafsson S, Haitjema S, Hottenga JJ, Huffman JE, Jackson AU, Jacobs KB, Johansson Å, Kaakinen M, Kleber ME, Lahti J, Mateo Leach I, Lehne B, Liu Y, Lo KS, Lorentzon M, Luan J, Madden PA, Mangino M, McKnight B, **Medina-Gomez C**, Monda KL, Montasser ME, Müller G, Müller-Nurasyid M, Nolte IM, Panoutsopoulou K, Pascoe L, Paternoster L, Rayner NW, Renström F, Rizzi F, Rose LM, Ryan KA, Salo P, Sanna S, Scharnagl H, Shi J, Smith AV, Southam L, Stančáková A, Steinthorsdottir V, Strawbridge RJ, Sung YJ, Tachmazidou I, Tanaka T, Thorleifsson G, Trompet S, Pervjakova N, Tyrer JP, Vandenput L, van der Laan SW, van der Velde N, van Setten J, van Vliet-Ostaptchouk JV, Verweij N, Vlachopoulou E, Waite LL, Wang SR, Wang Z, Wild SH, Willenborg C, Wilson JF, Wong A, Yang J, Yengo L, Yerges-Armstrong LM, Yu L, Zhang W, Zhao JH, Andersson EA, Bakker SJ, Baldassarre D, Banasik K, Barcella M, Barlassina C, Bellis C, Benaglio P, Blangero J, Blüher M, Bonnet F, Bonnycastle LL, Boyd HA, Bruinenberg M, Buchman AS, Campbell H, Chen YD, Chines PS, Claudi-Boehm S, Cole J, Collins FS, de Geus EJ, de Groot LC, Dimitriou M, Duan J, Enroth S, Eury E, Farmaki AE, Forouhi NG, Friedrich N, Gejman PV, Gigante B, Glorioso N, Go AS, Gottesman O, Gräßler J, Grallert H, Grarup N, Gu YM, Broer L, Ham AC, Hansen T, Harris TB, Hartman CA, Hassinen M, Hastie N, Hattersley AT, Heath AC, Henders AK, Hernandez D, Hillege H, Holmen O, Hovingh KG, Hui J, Husemoen LL, Hutri-Kähönen N, Hysi PG, Illig T, De Jager PL, Jalilzadeh S, Jørgensen T, Jukema JW, Juonala M, Kanoni S, Karaleftheri M, Khaw KT, Kinnunen L, Kittner SJ, Koenig W, Kolcic I, Kovacs P, Krarup NT, Kratzer W, Krüger J, Kuh D, Kumari M, Kyriakou T, Langenberg C, Lannfelt L, Lanzani C, Lotay V, Launer LJ, Leander K, Lindström J, Linneberg A, Liu YP, Lobbens S, Luben R, Lyssenko V, Männistö S, Magnusson PK, McArdle WL, Menni C, Merger S, Milani L, Montgomery GW, Morris AP, Narisu N, Nelis M, Ong KK, Palotie A, Pérusse L, Pichler I, Pilia MG, Pouta A, Rheinberger M, Ribel-Madsen R, Richards M, Rice KM, Rice TK, Rivolta C, Salomaa V, Sanders AR, Sarzynski MA, Scholtens S, Scott RA, Scott WR, Sebert S, Sengupta S, Sennblad B, Seufferlein

T, Silveira A, Slagboom PE, Smit JH, Sparsø TH, Stirrups K, Stolk RP, Stringham HM, Swertz MA, Swift AJ, Syvänen AC, Tan ST, Thorand B, Tönjes A, Tremblay A, Tsafantakis E, van der Most PJ, Völker U, Vohl MC, Vonk JM, Waldenberger M, Walker RW, Wennauer R, Widén E, Willemsen G, Wilsgaard T, Wright AF, Zillikens MC, van Dijk SC, van Schoor NM, Asselbergs FW, de Bakker PI, Beckmann JS, Beilby J, Bennett DA, Bergman RN, Bergmann S, Böger CA, Boehm BO, Boerwinkle E, Boomsma DI, Bornstein SR, Bottinger EP, Bouchard C, Chambers JC, Chanock SJ, Chasman DI, Cucca F, Cusi D, Dedoussis G, Erdmann J, Eriksson JG, Evans DA, de Faire U, Farrall M, Ferrucci L, Ford I, Franke L, Franks PW, Froguel P, Gansevoort RT, Gieger C, Grönberg H, Gudnason V, Gyllensten U, Hall P, Hamsten A, van der Harst P, Hayward C, Heliövaara M, Hengstenberg C, Hicks AA, Hingorani A, Hofman A, Hu F, Huikuri HV, Hveem K, James AL, Jordan JM, Jula A, Kähönen M, Kajantie E, Kathiresan S, Kiemeney LA, Kivimaki M, Knekt PB, Koistinen HA, Kooner JS, Koskinen S, Kuusisto J, Maerz W, Martin NG, Laakso M, Lakka TA, Lehtimäki T, Lettre G, Levinson DF, Lind L, Lokki ML, Mäntyselkä P, Melbye M, Metspalu A, Mitchell BD, Moll FL, Murray JC, Musk AW, Nieminen MS, Njølstad I, Ohlsson C, Oldehinkel AJ, Oostra BA, Palmer LJ, Pankow JS, Pasterkamp G, Pedersen NL, Pedersen O, Penninx BW, Perola M, Peters A, Polašek O, Pramstaller PP, Psaty BM, Qi L, Quertermous T, Raitakari OT, Rankinen T, Rauramaa R, Ridker PM, Rioux JD, Rivadeneira F, Rotter JI, Rudan I, den Ruijter HM, Saltevo J, Sattar N, Schunkert H, Schwarz PE, Shuldiner AR, Sinisalo J, Snieder H, Sørensen TI, Spector TD, Staessen JA, Stefania B, Thorsteinsdottir U, Stumvoll M, Tardif JC, Tremoli E, Tuomilehto J, Uitterlinden AG, Uusitupa M, Verbeek AL, Vermeulen SH, Viikari JS, Vitart V, Völzke H, Vollenweider P, Waeber G, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Zeggini E; CHARGE Consortium; DIAGRAM Consortium; GLGC Consortium; Global-BPGen Consortium; ICBP Consortium; MAGIC Consortium, Chakravarti A, Clegg DJ, Cupples LA, Gordon-Larsen P, Jaquish CE, Rao DC, Abecasis GR, Assimes TL, Barroso I, Berndt SI, Boehnke M, Deloukas P, Fox CS, Groop LC, Hunter DJ, Ingelsson E, Kaplan RC, McCarthy MI, Mohlke KL, O'Connell JR, Schlessinger D, Strachan DP, Stefansson K, van Duijn CM, Hirschhorn JN, Lindgren CM, Heid IM, North KE, Borecki IB, Kutalik Z, Loos RJ. 2015. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. ***PLoS Genet***. Oct 1;11(10):e1005378.

- Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, Dahia CL, Park-Min KH, Tobias JH, Kooperberg C, Kleinman A, Styrkarsdottir U, Liu CT, Uggla C, Evans DS, Nielson CM, Walter K, Pettersson-Kymmer U, McCarthy S, Eriksson J, Kwan T, Jhamai M, Trajanoska K, Memari Y, Min J, Huang J, Danecek P, Wilmot B, Li R, Chou WC, Mokry LE, Moayyeri A, Claussnitzer M, Cheng CH, Cheung W, **Medina-Gómez C**, Ge B, Chen SH, Choi K, Oei L, Fraser J, Kraaij R, Hibbs MA, Gregson CL, Paquette D, Hofman A, Wibom C, Tranah GJ, Marshall M, Gardiner BB, Cremin K, Auer P, Hsu L, Ring S, Tung JY, Thorleifsson G, Enneman AW, van Schoor NM, de Groot LC, van der Velde N, Melin B, Kemp JP, Christiansen C, Sayers A, Zhou Y, Calderari S, van Rooij J, Carlson C, Peters U, Berlivet S, Dostie J, Uitterlinden AG, Williams SR, Farber C, Grinberg D, LaCroix AZ, Haessler J, Chasman DI, Giulianini F, Rose LM, Ridker PM, Eisman JA, Nguyen TV, Center JR, Nogues X, Garcia-Giralt N, Launer LL, Gudnason V, Mellström D, Vandenput L, Amin N, van Duijn CM, Karlsson MK, Ljunggren Ö, Svensson O, Hallmans G, Rousseau F, Giroux S, Bussière J, Arp PP, Koromani F, Prince RL, Lewis JR, Langdahl BL, Hermann AP, Jensen JE, Kaptoge S, Khaw KT, Reeve J, Formosa MM, Xuereb-Anastasi A, Åkesson K, McGuigan FE, Garg G, Olmos JM, Zarrabeitia MT, Riancho JA, Ralston SH, Alonso N, Jiang X, Goltzman D, Pastinen T, Grundberg E, Gauguier D, Orwoll ES, Karasik D, Davey-Smith G; AOGC Consortium, Smith AV, Siggeirsdottir K, Harris TB, Zillikens MC, van Meurs JB, Thorsteinsdottir U, Maurano MT, Timpson NJ, Soranzo N, Durbin R, Wilson SG, Ntzani EE, Brown MA, Stefansson K, Hinds DA, Spector T, Cupples LA, Ohlsson C, Greenwood CM; UK10K Consortium, Jackson RD, Rowe DW, Loomis CA, Evans DM, Ackert-Bicknell CL, Joyner AL, Duncan EL, Kiel DP, Rivadeneira F, Richards JB. 2015. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. ***Nature.*** Oct 1;526(7571):112-7.
- Robinson MR, Hemani G, **Medina-Gomez C**, Mezzavilla M, Esko T, Shakhbazov K, Powell JE, Vinkhuyzen A, Berndt SI, Gustafsson S, Justice AE, Kahali B, Locke AE, Pers TH, Vedantam S, Wood AR, van Rheenen W, Andreassen OA, Gasparini P, Metspalu A, Berg LH, Veldink JH, Rivadeneira F, Werge TM, Abecasis GR, Boomsma DI, Chasman DI, de Geus EJ, Frayling TM, Hirschhorn JN, Hottenga JJ, Ingelsson E, Loos RJ, Magnusson PK, Martin NG, Montgomery GW, North KE, Pedersen NL, Spector TD, Speliotes EK, Goddard ME,

Yang J, Visscher PM. 2015. Population genetic differentiation of height and body mass index across Europe. ***Nat Genet***. Nov;47(11):1357-62.

- Fedko IO, Hottenga JJ, **Medina-Gomez C**, Pappa I, van Beijsterveldt CE, Ehli EA, Davies GE, Rivadeneira F, Tiemeier H, Swertz MA, Middeldorp CM, Bartels M, Boomsma DI. 2015. Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. ***Behav Genet***. Jun 3.
- van den Hooven EH, Heppe DH, Kiefte-de Jong JC, **Medina-Gomez C**, Moll HA, Hofman A, Jaddoe VW, Rivadeneira F, Franco OH. 2015. Infant dietary patterns and bone mass in childhood: the Generation R Study. ***Osteoporos Int.*** May;26(5):1595-604.
- van Leeuwen EM, Karssen LC, Deelen J, Isaacs A, **Medina-Gomez C**, Mbarek H, Kanterakis A, Trompet S, Postmus I, Verweij N, van Enckevort DJ, Huffman JE, White CC, Feitosa MF, Bartz TM, Manichaikul A, Joshi PK, Peloso GM, Deelen P, van Dijk F, Willemsen G, de Geus EJ, Milaneschi Y, Penninx BW, Francioli LC, Menelaou A, Pulit SL, Rivadeneira F, Hofman A, Oostra BA, Franco OH, Mateo Leach I, Beekman M, de Craen AJ, Uh HW, Trochet H, Hocking LJ, Porteous DJ, Sattar N, Packard CJ, Buckley BM, Brody JA, Bis JC, Rotter JI, Mychaleckyj JC, Campbell H, Duan Q, Lange LA, Wilson JF, Hayward C, Polasek O, Vitart V, Rudan I, Wright AF, Rich SS, Psaty BM, Borecki IB, Kearney PM, Stott DJ, Adrienne Cupples L; Genome of The Netherlands Consortium, Jukema JW, van der Harst P, Sijbrands EJ, Hottenga JJ, Uitterlinden AG, Swertz MA, van Ommen GJ, de Bakker PI, Eline Slagboom P, Boomsma DI, Wijmenga C, van Duijn CM. 2015. Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. ***Nat Commun***. Mar 9;6:6065.
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua Zhao J, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, **Medina-Gomez C**, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL, Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen YD, Clarke R, Daw EW, de Craen AJ, Delgado G, Dimitriou M, Doney AS, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H, Grammer TB, Gräßler J, Grönberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM, Johansson Å, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Sin Lo K, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PK, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Müller G, Müller-Nurasyid M, Musk AW, Nagaraja R, Nöthen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen AC, Tan ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q; LifeLines Cohort Study, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gådin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JR, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, Van't Hooft FM, Vinkhuyzen AA, Westra HJ, Zheng W, Zondervan KT; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIO-

GRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium, Heath AC, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC, Hyppönen E, Illig T, Jacobs KB, Jarvelin MR, Jöckel KH, Johansen B, Jousilahti P, Jukema JW, Jula AM, Kaprio J, Kastelein JJ, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PA, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PE, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tönjes A, Trégouët DA, Tremblay A, Tremoli E, Virtamo J, Vohl MC, Völker U, Waeber G, Willemsen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJ, Speliotes EK.2015. Genetic studies of body mass index yield new insights for obesity biology. ***Nature***. Feb 12;518(7538):197-206.

- Van der Valk RJ, Kreiner-Møller E, Kooijman MN, Guxens M, Stergiakouli E, Sääf A, Bradfield JP, Geller F, Hayes MG, Cousminer DL, Körner A, Thiering E, Curtin JA, Myhre R, Huikari V, Joro R, Kerkhof M, Warrington NM, Pitkänen N, Ntalla I, Horikoshi M, Veijola R, Freathy RM, Teo YY, Barton SJ, Evans DM, Kemp JP, St Pourcain B, Ring SM, Davey Smith G, Bergström A, Kull I, Hakonarson H, Mentch FD, Bisgaard H, Chawes B, Stokholm J, Waage J, Eriksen P, Sevelsted A, Melbye M; Early Genetics and Lifecourse Epidemiology (EAGLE) Consortium, van Duijn CM, **Medina-Gomez C**, Hofman A, de Jongste JC, Taal HR, Uitterlinden AG; Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Armstrong LL, Eriksson J, Palotie A, Bustamante M, Estivill X, Gonzalez JR, Llop S, Kiess W, Mahajan A, Flexeder C, Tiesler CM, Murray CS, Simpson A, Magnus P, Sengpiel V, Hartikainen AL, Keinanen-Kiukaanniemi S, Lewin A, Da Silva Couto Alves A, Blakemore AI, Buxton JL, Kaakinen M, Rodriguez A, Sebert S, Vaarasmaki M, Lakka T, Lindi V, Gehring U, Postma DS, Ang W, Newnham JP, Lyytikäinen LP, Pahkala K, Raitakari OT, Panoutsopoulou K, Zeggini E, Boomsma DI, Groen-Blokhuis M, Ilonen J, Franke L, Hirschhorn JN, Pers TH, Liang L, Huang J, Hocher B, Knip M, Saw SM, Holloway JW, Melén E, Grant SF, Feenstra B, Lowe WL, Widén E, Sergeyev E, Grallert H, Custovic A, Jacobsson B, Jarvelin MR, Atalay M, Koppelman GH, Pennell CE, Niinikoski H, Dedoussis GV, Mccarthy MI, Frayling TM, Sunyer J, Timpson NJ, Rivadeneira F, Bønnelykke K, Jaddoe VW; Early Growth Genetics (EGG) Consortium. A novel common variant in DCST2 is associated with length in early life and height in adulthood. ***Hum Mol Genet.*** Feb 15;24(4):1155-68.
- Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JM, Buchkovich ML, Heard-Costa NL, Roman TS, Drong AW, Song C, Gustafsson S, Day FR, Esko T, Fall T, Kutalik Z, Luan J, Randall JC, Scherag A, Vedantam S, Wood AR, Chen J, Fehrmann R, Karjalainen J, Kahali B, Liu CT, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M,

Bragg-Gresham JL, Buyske S, Demirkan A, Ehret GB, Feitosa MF, Goel A, Jackson AU, Johnson T, Kleber ME, Kristiansson K, Mangino M, Mateo Leach I, **Medina-Gomez C**, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Stanc̆áková A, Ju Sung Y, Tanaka T, Teumer A, Van Vliet-Ostaptchouk JV, Yengo L, Zhang W, Albrecht E, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Böhringer S, Bonnet F, Böttcher Y, Bruinenberg M, Carba DB, Caspersen IH, Clarke R, Daw EW, Deelen J, Deelman E, Delgado G, Doney AS, Eklund N, Erdos MR, Estrada K, Eury E, Friedrich N, Garcia ME, Giedraitis V, Gigante B, Go AS, Golay A, Grallert H, Grammer TB, Gräßler J, Grewal J, Groves CJ, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heikkilä K, Herzig KH, Helmer Q, Hillege HL, Holmen O, Hunt SC, Isaacs A, Ittermann T, James AL, Johansson I, Juliusdottir T, Kalafati IP, Kinnunen L, Koenig W, Kooner IK, Kratzer W, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Mach F, Magnusson PK, Mahajan A, McArdle WL, Menni C, Merger S, Mihailov E, Milani L, Mills R, Moayyeri A, Monda KL, Mooijaart SP, Mühleisen TW, Mulas A, Müller G, Müller-Nurasyid M, Nagaraja R, Nalls MA, Narisu N, Glorioso N, Nolte IM, Olden M, Rayner NW, Renstrom F, Ried JS, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Sennblad B, Seufferlein T, Sitlani CM, Vernon Smith A, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen AC, Tayo BO, Thorand B, Thorleifsson G, Tomaschitz A, Troffa C, van Oort FV, Verweij N, Vonk JM, Waite LL, Wennauer R, Wilsgaard T, Wojczynski MK, Wong A, Zhang Q, Hua Zhao J, Brennan EP, Choi M, Eriksson P, Folkersen L, Franco-Cereceda A, Gharavi AG, Hedman ÅK, Hivert MF, Huang J, Kanoni S, Karpe F, Keildson S, Kiryluk K, Liang L, Lifton RP, Ma B, McKnight AJ, McPherson R, Metspalu A, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Olsson C, Perry JR, Reinmaa E, Salem RM, Sandholm N, Schadt EE, Scott RA, Stolk L, Vallejo EE, Westra HJ, Zondervan KT; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium, Amouyel P, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Blangero J, Brown MJ, Burnier M, Campbell H, Chakravarti A, Chines PS, Claudi-Boehm S, Collins FS, Crawford DC, Danesh J, de Faire U, de Geus EJ, Dörr M, Erbel R, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gieger C, Gudnason V, Haiman CA, Harris TB, Hattersley AT, Heliövaara M, Hicks AA, Hingorani AD, Hoffmann W, Hofman A, Homuth G, Humphries SE, Hyppönen E, Illig T, Jarvelin MR, Johansen B, Jousilahti P, Jula AM, Kaprio J, Kee F, Keinanen-Kiukaanniemi SM, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuulasmaa K, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Musk AW, Möhlenkamp S, Morris AD, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Palmer LJ, Penninx BW, Peters A, Pramstaller PP, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PE, Shuldiner AR, Staessen JA, Steinthorsdottir V, Stolk RP, Strauch K, Tönjes A, Tremblay A, Tremoli E, Vohl MC, Völker U, Vollenweider P, Wilson JF, Witteman JC, Adair LS, Bochud M, Boehm BO, Bornstein SR, Bouchard C, Cauchi S, Caulfield MJ, Chambers JC, Chasman DI, Cooper RS, Dedoussis G, Ferrucci L, Froguel P, Grabe HJ, Hamsten A, Hui J, Hveem K, Jöckel KH, Kivimaki M, Kuh D, Laakso M, Liu Y, März W, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sinisalo J, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Veronesi G, Walker M, Wareham NJ, Watkins H, Wichmann HE, Abecasis GR, Assimes TL, Berndt SI, Boehnke M, Borecki IB, Deloukas P, Franke L, Frayling TM, Groop LC, Hunter DJ, Kaplan RC, O'Connell JR, Qi L, Schlessinger D, Strachan DP, Stefansson K, van Duijn CM, Willer CJ, Visscher PM, Yang J, Hirschhorn JN, Zillikens MC, McCarthy MI, Speliotes EK, North KE, Fox CS, Barroso I, Franks PW, Ingelsson E, Heid IM, Loos RJ, Cupples LA, Morris AP, Lindgren CM, Mohlke KL. 2015. New genetic loci link adipose and insulin biology to body fat distribution. ***Nature.*** Feb 12;518(7538):187-96.

- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan

A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, **Medina-Gomez C**, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney AS, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LC, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PK, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen AC, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FV, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin MR, Jousilahti P, Jula AM, Kaprio J, Kastelein JJ, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lupoli S, Madden PA, Männistö S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PE, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif JC, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl MC; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M, Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJ, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. ***Nat Genet***. 46(11):1173-86

- Reppe S, Noer A, Grimholt RM, Halldórsson BV, **Medina-Gomez C**, Gautvik VT, Olstad OK, Berg JP, Datta H, Estrada K, Hofman A, Uitterlinden AG, Rivadeneira F, Lyle R, Collas P, Gautvik KM. 2014. Methylation of bone SOST, its mRNA, and serum sclerostin levels correlate strongly with fracture risk in postmenopausal women. ***J Bone Miner Res***. Aug 22.
- Deelen P, Menelaou A, van Leeuwen EM, Kanterakis A, van Dijk F, **Medina-Gomez C**, Francioli LC, Hottenga JJ, Karssen LC, Estrada K, Kreiner-Møller E, Rivadeneira F, van Setten J, Gutierrez-Achury J, Westra HJ, Franke L, van Enckevort D, Dijkstra M, Byelas H, van Duijn CM; Genome of Netherlands Consortium, de Bakker PI, Wijmenga C, Swertz MA. 2014. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. ***Eur J Hum Genet***. 2014 Nov;22(11):1321-6.
- Oei L, Estrada K, Duncan EL, Christiansen C, Liu CT, Langdahl BL, Obermayer-Pietsch B, Riancho JA, Prince RL, van Schoor NM, McCloskey E, Hsu YH, Evangelou E, Ntzani E, Evans DM, Alonso N, Husted LB, Valero C, Hernandez JL, Lewis JR, Kaptoge SK, Zhu K, Cupples LA, **Medina-Gómez C**, Vandenput L, Kim GS, Hun Lee S, Castaño-Betancourt MC, Oei EH, Martinez J, Daroszewska A, van der Klift M, Mellström D, Herrera L, Karlsson MK, Hofman A, Ljunggren Ö, Pols HA, Stolk L, van Meurs JB, Ioannidis JP, Zillikens MC, Lips P, Karasik D, Uitterlinden AG, Styrkarsdottir U, Brown MA, Koh JM, Richards JB, Reeve J, Ohlsson C, Ralston SH, Kiel DP, Rivadeneira F. 2014. Genome-wide association study for radiographic vertebral fractures: a potential role for the 16q24 BMD locus. ***Bone.*** Feb;59:20-7**.**
- Moayyeri A, Hsu YH, Karasik D, Estrada K, Xiao SM, Nielson C, Srikanth P, Giroux S, Wilson SG, Zheng HF, Smith AV, Pye SR, Leo PJ, Teumer A, Hwang JY, Ohlsson C, McGuigan F, Minster RL, Hayward C, Olmos JM, Lyytikäinen LP, Lewis JR, Swart KM, Masi L, Oldmeadow C, Holliday EG, Cheng S, van Schoor NM, Harvey NC, Kruk M, del Greco M F, Igl W, Trummer O, Grigoriou E, Luben R, Liu CT, Zhou Y, Oei L, **Medina-Gomez C**, Zmuda J, Tranah G, Brown SJ, Williams FM, Soranzo N, Jakobsdottir J, Siggeirsdottir K, Holliday KL, Hannemann A, Go MJ, Garcia M, Polasek O, Laaksonen M, Zhu K, Enneman AW, McEvoy M, Peel R, Sham PC, Jaworski M, Johansson Å, Hicks AA, Pludowski P, Scott R, Dhonukshe-Rutten RA, van der Velde N, Kähönen M, Viikari JS, Sievänen H, Raitakari OT, González-Macías J, Hernández JL, Mellström D, Ljunggren O, Cho YS, Völker U, Nauck M, Homuth G, Völzke H, Haring R, Brown MA, McCloskey E, Nicholson GC, Eastell R, Eisman JA, Jones G, Reid IR, Dennison EM, Wark J, Boonen S, Vanderschueren D, Wu FC, Aspelund T, Richards JB, Bauer D, Hofman A, Khaw KT, Dedoussis G, Obermayer-Pietsch B, Gyllensten U, Pramstaller PP, Lorenc RS, Cooper C, Kung AW, Lips P, Alen M, Attia J, Brandi ML, de Groot LC, Lehtimäki T, Riancho JA, Campbell H, Liu Y, Harris TB, Akesson K, Karlsson M, Lee JY, Wallaschofski H, Duncan EL, O'Neill TW, Gudnason V, Spector TD, Rousseau F, Orwoll E, Cummings SR, Wareham NJ, Rivadeneira F, Uitterlinden AG, Prince RL, Kiel DP, Reeve J, Kaptoge SK. 2014. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. ***Hum Mol Genet.*** Jun 1;23(11):3054-68.
- Oei L, Hsu YH, Styrkarsdottir U, Eussen BH, de Klein A, Peters MJ, Halldorsson B, Liu CT, Alonso N, Kaptoge SK, Thorleifsson G, Hallmans G, Hocking LJ, Husted LB, Jameson KA, Kruk M, Lewis JR, Patel MS, Scollen S, Svensson O, Trompet S, van Schoor NM, Zhu K, Buckley BM, Cooper C, Ford I, Goltzman D, González-Macías J, Langdahl BL, Leslie WD, Lips P, Lorenc RS, Olmos JM, Pettersson-Kymmer U, Reid DM, Riancho JA, Slagboom PE, Garcia-Ibarbia C, Ingvarsson T, Johannsdottir H, Luben R, **Medina-Gómez C**, et al. 2014. A genome-wide copy number association study of osteoporotic fractures points to the 6p25.1 locus. ***J Med Genet***. Feb;51(2):122-31.
- Oei L, Estrada K, Duncan EL, Christiansen C, Liu CT, Langdahl BL, Obermayer-Pietsch B, Riancho JA, Prince RL, van Schoor NM, McCloskey E, Hsu YH, Evangelou E, Ntzani E, Evans DM, Alonso N, Husted LB, Valero C, Hernandez JL, Lewis JR, Kaptoge SK, Zhu K, Cupples LA, **Medina-Gómez C**, et al. 2013. Genome-wide Association Study for Radiographic Vertebral Fractures: A Potential Role for the 16q24 BMD Locus versus Lessons Learned from Challenging Phenotype Definition. ***Bone.*** Oct 25. pii: S8756-3282(13)00425-0.
- Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, Kilpeläinen TO, Esko T, Mägi R, Li S, Workalemahu T, Feitosa MF, Croteau-Chonka DC, Day FR, Fall T, Ferreira T, Gustafsson S, Locke AE, Mathieson I, Scherag A, Vedantam S, Wood AR, Liang L, Steinthorsdottir V, Thorleifsson G, Dermitzakis ET, Dimas AS, Karpe F, Min JL, Nicholson G, Clegg DJ, Person T, Krohn JP, Bauer S, Buechler C, Eisinger K; DIAGRAM

Consortium, Bonnefond A, Froguel P; MAGIC Investigators, Hottenga JJ, Prokopenko I, Waite LL, Harris TB, Smith AV, Shuldiner AR, McArdle WL, Caulfield MJ, Munroe PB, Grönberg H, Chen YD, Li G, Beckmann JS, Johnson T, Thorsteinsdottir U, Teder-Laving M, Khaw KT, Wareham NJ, Zhao JH, Amin N, Oostra BA, Kraja AT, Province MA, Cupples LA, Heard-Costa NL, Kaprio J, Ripatti S, Surakka I, Collins FS, Saramies J, Tuomilehto J, Jula A, Salomaa V, Erdmann J, Hengstenberg C, Loley C, Schunkert H, Lamina C, Wichmann HE, Albrecht E, Gieger C, Hicks AA, Johansson A, Pramstaller PP, Kathiresan S, Speliotes EK, Penninx B, Hartikainen AL, Jarvelin MR, Gyllensten U, Boomsma DI, Campbell H, Wilson JF, Chanock SJ, Farrall M, Goel A**, Medina-Gomez C**, Rivadeneira F, Estrada K, Uitterlinden AG, Hofman A, Zillikens MC, den Heijer M, Kiemeney LA, Maschio A, Hall P, Tyrer J, Teumer A, Völzke H, Kovacs P, Tönjes A, Mangino M, Spector TD, Hayward C, Rudan I, Hall AS, Samani NJ, Attwood AP, Sambrook JG, Hung J, Palmer LJ, Lokki ML, Sinisalo J, Boucher G, Huikuri H, Lorentzon M, Ohlsson C, Eklund N, Eriksson JG, Barlassina C, Rivolta C, Nolte IM, Snieder H, Van der Klauw MM, Van Vliet-Ostaptchouk JV, Gejman PV, Shi J, Jacobs KB, Wang Z, Bakker SJ, Mateo Leach I, Navis G, van der Harst P, Martin NG, Medland SE, Montgomery GW, Yang J, Chasman DI, Ridker PM, Rose LM, Lehtimäki T, Raitakari O, Absher D, Iribarren C, Basart H, Hovingh KG, Hyppönen E, Power C, Anderson D, Beilby JP, Hui J, Jolley J, Sager H, Bornstein SR, Schwarz PE, Kristiansson K, Perola M, Lindström J, Swift AJ, Uusitupa M, Atalay M, Lakka TA, Rauramaa R, Bolton JL, Fowkes G, Fraser RM, Price JF, Fischer K, Krjutå Kov K, Metspalu A, Mihailov E, Langenberg C, Luan J, Ong KK, Chines PS, Keinanen-Kiukaanniemi SM, Saaristo TE, Edkins S, Franks PW, Hallmans G, Shungin D, Morris AD, Palmer CN, Erbel R, Moebus S, Nöthen MM, Pechlivanis S, Hveem K, Narisu N, Hamsten A, Humphries SE, Strawbridge RJ, Tremoli E, Grallert H, Thorand B, Illig T, Koenig W, Müller-Nurasyid M, Peters A, Boehm BO, Kleber ME, März W, Winkelmann BR, Kuusisto J, Laakso M, Arveiler D, Cesana G, Kuulasmaa K, Virtamo J, Yarnell JW, Kuh D, Wong A, Lind L, de Faire U, Gigante B, Magnusson PK, Pedersen NL, Dedoussis G, Dimitriou M, Kolovou G, Kanoni S, Stirrups K, Bonnycastle LL, Njølstad I, Wilsgaard T, Ganna A, Rehnberg E, Hingorani A, Kivimaki M, Kumari M, Assimes TL, Barroso I, Boehnke M, Borecki IB, Deloukas P, Fox CS, Frayling T, Groop LC, Haritunians T, Hunter D, Ingelsson E, Kaplan R, Mohlke KL, O'Connell JR, Schlessinger D, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, McCarthy MI, Hirschhorn JN, Qi L, Loos RJ, Lindgren CM, North KE, Heid IM. 2013. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. ***PLoS Genet.*** Jun;9(6):e1003500.

- Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, Justice AE, Monda KL, Croteau-Chonka DC, Day FR, Esko T, Fall T, Ferreira T, Gentilini D, Jackson AU, Luan J, Randall JC, Vedantam S, Willer CJ, Winkler TW, Wood AR, Workalemahu T, Hu YJ, Lee SH, Liang L, Lin DY, Min JL, Neale BM, Thorleifsson G, Yang J, Albrecht E, Amin N, Bragg-Gresham JL, Cadby G, den Heijer M, Eklund N, Fischer K, Goel A, Hottenga JJ, Huffman JE, Jarick I, Johansson Å, Johnson T, Kanoni S, Kleber ME, König IR, Kristiansson K, Kutalik Z, Lamina C, Lecoeur C, Li G, Mangino M, McArdle WL, **Medina-Gomez C**, Müller-Nurasyid M, Ngwa JS, Nolte IM, Paternoster L, Pechlivanis S, Perola M, Peters MJ, Preuss M, Rose LM, Shi J, Shungin D, Smith AV, Strawbridge RJ, Surakka I, Teumer A, Trip MD, Tyrer J, Van Vliet-Ostaptchouk JV, Vandenput L, Waite LL, Zhao JH, Absher D, Asselbergs FW, Atalay M, Attwood AP, Balmforth AJ, Basart H, Beilby J, Bonnycastle LL, Brambilla P, Bruinenberg M, Campbell H, Chasman DI, Chines PS, Collins FS, Connell JM, Cookson WO, de Faire U, de Vegt F, Dei M, Dimitriou M, Edkins S, Estrada K, Evans DM, Farrall M, Ferrario MM, Ferrières J, Franke L, Frau F, Gejman PV, Grallert H, Grönberg H, Gudnason V, Hall AS, Hall P, Hartikainen AL, Hayward C, Heard-Costa NL, Heath AC, Hebebrand J, Homuth G, Hu FB, Hunt SE, Hyppönen E, Iribarren C, Jacobs KB, Jansson JO, Jula A, Kähönen M, Kathiresan S, Kee F, Khaw KT, Kivimäki M, Koenig W, Kraja AT, Kumari M, Kuulasmaa K, Kuusisto J, Laitinen JH, Lakka TA, Langenberg C, Launer LJ, Lind L, Lindström J, Liu J, Liuzzi A, Lokki ML, Lorentzon M, Madden PA, Magnusson PK, Manunta P, Marek D, März W, Mateo Leach I, McKnight B, Medland SE, Mihailov E, Milani L, Montgomery GW, Mooser V, Mühleisen TW, Munroe PB, Musk AW, Narisu N, Navis G, Nicholson G, Nohr EA, Ong KK, Oostra BA, Palmer CN, Palotie A, Peden JF, Pedersen N, Peters A, Polasek O, Pouta A, Pramstaller PP, Prokopenko I, Pütter C, Radhakrishnan A, Raitakari O, Rendon A, Rivadeneira F, Rudan I, Saaristo TE, Sambrook JG, Sanders AR, Sanna S, Saramies J, Schipf S, Schreiber S, Schunkert H, Shin SY, Signorini S, Sinisalo J, Skrobek B, Soranzo N, Stančáková A, Stark K, Stephens JC, Stirrups K, Stolk RP, Stumvoll M, Swift AJ, Theodoraki EV, Thorand B, Tregouet DA, Tremoli E,

Van der Klauw MM, van Meurs JB, Vermeulen SH, Viikari J, Virtamo J, Vitart V, Waeber G, Wang Z, Widén E, Wild SH, Willemsen G, Winkelmann BR, Witteman JC, Wolffenbuttel BH, Wong A, Wright AF, Zillikens MC, Amouyel P, Boehm BO, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Cupples LA, Cusi D, Dedoussis GV, Erdmann J, Eriksson JG, Franks PW, Froguel P, Gieger C, Gyllensten U, Hamsten A, Harris TB, Hengstenberg C, Hicks AA, Hingorani A, Hinney A, Hofman A, Hovingh KG, Hveem K, Illig T, Jarvelin MR, Jöckel KH, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kuh D, Laakso M, Lehtimäki T, Levinson DF, Martin NG, Metspalu A, Morris AD, Nieminen MS, Njølstad I, Ohlsson C, Oldehinkel AJ, Ouwehand WH, Palmer LJ, Penninx B, Power C, Province MA, Psaty BM, Qi L, Rauramaa R, Ridker PM, Ripatti S, Salomaa V, Samani NJ, Snieder H, Sørensen TI, Spector TD, Stefansson K, Tönjes A, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Vollenweider P, Wallaschofski H, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Abecasis GR, Assimes TL, Barroso I, Boehnke M, Borecki IB, Deloukas P, Fox CS, Frayling T, Groop LC, Haritunian T, Heid IM, Hunter D, Kaplan RC, Karpe F, Moffatt MF, Mohlke KL, O'Connell JR, Pawitan Y, Schadt EE, Schlessinger D, Steinthorsdottir V, Strachan DP, Thorsteinsdottir U, van Duijn CM, Visscher PM, Di Blasio AM, Hirschhorn JN, Lindgren CM, Morris AP, Meyre D, Scherag A, McCarthy MI, Speliotes EK, North KE, Loos RJ, Ingelsson E. 2013. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. ***Nat Genet***. May;45(5):501-12.

- Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, Waite L, Smith AV, Yerges-Armstrong LM, Monda KL, Hadley D, Mahajan A, Li G, Kapur K, Vitart V, Huffman JE, Wang SR, Palmer C, Esko T, Fischer K, Zhao JH, Demirkan A, Isaacs A, Feitosa MF, Luan J, Heard-Costa NL, White C, Jackson AU, Preuss M, Ziegler A, Eriksson J, Kutalik Z, Frau F, Nolte IM, Van Vliet-Ostaptchouk JV, Hottenga JJ, Jacobs KB, Verweij N, Goel A, **Medina-Gomez C**, Estrada K, Bragg-Gresham JL, Sanna S, Sidore C, Tyrer J, Teumer A, Prokopenko I, Mangino M, Lindgren CM, Assimes TL, Shuldiner AR, Hui J, Beilby JP, McArdle WL, Hall P, Haritunians T, Zgaga L, Kolcic I, Polasek O, Zemunik T, Oostra BA, Junttila MJ, Grönberg H, Schreiber S, Peters A, Hicks AA, Stephens J, Foad NS, Laitinen J, Pouta A, Kaakinen M, Willemsen G, Vink JM, Wild SH, Navis G, Asselbergs FW, Homuth G, John U, Iribarren C, Harris T, Launer L, Gudnason V, O'Connell JR, Boerwinkle E, Cadby G, Palmer LJ, James AL, Musk AW, Ingelsson E, Psaty BM, Beckmann JS, Waeber G, Vollenweider P, Hayward C, Wright AF, Rudan I, Groop LC, Metspalu A, Khaw KT, van Duijn CM, Borecki IB, Province MA, Wareham NJ, Tardif JC, Huikuri HV, Cupples LA, Atwood LD, Fox CS, Boehnke M, Collins FS, Mohlke KL, Erdmann J, Schunkert H, Hengstenberg C, Stark K, Lorentzon M, Ohlsson C, Cusi D, Staessen JA, Van der Klauw MM, Pramstaller PP, Kathiresan S, Jolley JD, Ripatti S, Jarvelin MR, de Geus EJ, Boomsma DI, Penninx B, Wilson JF, Campbell H, Chanock SJ, van der Harst P, Hamsten A, Watkins H, Hofman A, Witteman JC, Zillikens MC, Uitterlinden AG, Rivadeneira F, Zillikens MC, Kiemeney LA, Vermeulen SH, Abecasis GR, Schlessinger D, Schipf S, Stumvoll M, Tönjes A, Spector TD, North KE, Lettre G, McCarthy MI, Berndt SI, Heath AC, Madden PA, Nyholt DR, Montgomery GW, Martin NG, McKnight B, Strachan DP, Hill WG, Snieder H, Ridker PM, Thorsteinsdottir U, Stefansson K, Frayling TM, Hirschhorn JN, Goddard ME, Visscher PM. 2012. FTO genotype is associated with phenotypic variability of body mass index. ***Nature***. Oct 11;490(7419):267-72.
- Zheng HF, Tobias JH, Duncan E, Evans DM, Eriksson J, Paternoster L, Yerges-Armstrong LM, Lehtimäki T, Bergström U, Kähönen M, Leo PJ, Raitakari O, Laaksonen M, Nicholson GC, Viikari J, Ladouceur M, Lyytikäinen LP, **Medina-Gomez C**, Rivadeneira F, Prince RL, Sievanen H, Leslie WD, Mellström D, Eisman JA, Movérare-Skrtic S, Goltzman D, Hanley DA, Jones G, St Pourcain B, Xiao Y, Timpson NJ, Smith GD, Reid IR, Ring SM, Sambrook PN, Karlsson M, Dennison EM, Kemp JP, Danoy P, Sayers A, Wilson SG, Nethander M, McCloskey E, Vandenput L, Eastell R, Liu J, Spector T, Mitchell BD, Streeten EA, Brommage R, Pettersson-Kymmer U, Brown MA, Ohlsson C, Richards JB, Lorentzon M. 2012. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. ***PLoS Genet.*** Jul;8(7):e1002745.
- Liu CT, Estrada K, Yerges-Armstrong LM, Amin N, Evangelou E, Li G, Minster RL, Carless MA, Kammerer CM, Oei L, Zhou Y, Alonso N, Dailiana Z, Eriksson J, García-Giralt N, Giroux S, Husted LB, Khusainova RI, Koromila T, Kung AW, Lewis JR, Masi L, Mencej-Bedrac S, Nogues X, Patel MS, Prezelj J, Richards JB, Sham PC, Spector T, Vandenput L, Xiao SM, Zheng HF, Zhu K, Balcells S, Brandi ML, Frost M, Goltzman D, González-Macías J, Karlsson M, Khusnutdinova EK, Kollia P, Langdahl BL, Ljunggren O, Lorentzon M, Marc

J, Mellström D, Ohlsson C, Olmos JM, Ralston SH, Riancho JA, Rousseau F, Urreizti R, Van Hul W, Zarrabeitia MT, Castano-Betancourt M, Demissie S, Grundberg E, Herrera L, Kwan T, **Medina-Gómez C**, Pastinen T, Sigurdsson G, Thorleifsson G, Vanmeurs JB, Blangero J, Hofman A, Liu Y, Mitchell BD, O'Connell JR, Oostra BA, Rotter JI, Stefansson K, Streeten EA, Styrkarsdottir U, Thorsteinsdottir U, Tylavsky FA, Uitterlinden A, Cauley JA, Harris TB, Ioannidis JP, Psaty BM, Robbins JA, Zillikens MC, Vanduijn CM, Prince RL, Karasik D, Rivadeneira F, Kiel DP, Cupples LA, Hsu YH. 2012. Assessment of gene-by-sex interaction effect on bone mineral density. ***J Bone Miner Res***. Oct;27(10):2051-64.

- Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP, Koller DL, Li G, Liu CT, Minster RL, Moayyeri A, Vandenput L, Willner D, Xiao SM, Yerges-Armstrong LM, Zheng HF, Alonso N, Eriksson J, Kammerer CM, Kaptoge SK, Leo PJ, Thorleifsson G, Wilson SG, Wilson JF, Aalto V, Alen M, Aragaki AK, Aspelund T, Center JR, Dailiana Z, Duggan DJ, Garcia M, Garcia-Giralt N, Giroux S, Hallmans G, Hocking LJ, Husted LB, Jameson KA, Khusainova R, Kim GS, Kooperberg C, Koromila T, Kruk M, Laaksonen M, Lacroix AZ, Lee SH, Leung PC, Lewis JR, Masi L, Mencej-Bedrac S, Nguyen TV, Nogues X, Patel MS, Prezelj J, Rose LM, Scollen S, Siggeirsdottir K, Smith AV, Svensson O, Trompet S, Trummer O, van Schoor NM, Woo J, Zhu K, Balcells S, Brandi ML, Buckley BM, Cheng S, Christiansen C, Cooper C, Dedoussis G, Ford I, Frost M, Goltzman D, González-Macías J, Kähönen M, Karlsson M, Khusnutdinova E, Koh JM, Kollia P, Langdahl BL, Leslie WD, Lips P, Ljunggren Ö, Lorenc RS, Marc J, Mellström D, Obermayer-Pietsch B, Olmos JM, Pettersson-Kymmer U, Reid DM, Riancho JA, Ridker PM, Rousseau F, Slagboom PE, Tang NL, Urreizti R, Van Hul W, Viikari J, Zarrabeitia MT, Aulchenko YS, Castano-Betancourt M, Grundberg E, Herrera L, Ingvarsson T, Johannsdottir H, Kwan T, Li R, Luben R, **Medina-Gómez C**, Palsson ST, Reppe S, Rotter JI, Sigurdsson G, van Meurs JB, Verlaan D, Williams FM, Wood AR, Zhou Y, Gautvik KM, Pastinen T, Raychaudhuri S, Cauley JA, Chasman DI, Clark GR, Cummings SR, Danoy P, Dennison EM, Eastell R, Eisman JA, Gudnason V, Hofman A, Jackson RD, Jones G, Jukema JW, Khaw KT, Lehtimäki T, Liu Y, Lorentzon M, McCloskey E, Mitchell BD, Nandakumar K, Nicholson GC, Oostra BA, Peacock M, Pols HA, Prince RL, Raitakari O, Reid IR, Robbins J, Sambrook PN, Sham PC, Shuldiner AR, Tylavsky FA, van Duijn CM, Wareham NJ, Cupples LA, Econs MJ, Evans DM, Harris TB, Kung AW, Psaty BM, Reeve J, Spector TD, Streeten EA, Zillikens MC, Thorsteinsdottir U, Ohlsson C, Karasik D, Richards JB, Brown MA, Stefansson K, Uitterlinden AG, Ralston SH, Ioannidis JP, Kiel DP, Rivadeneira F.2012. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. ***Nat Genet.*** Apr 15;44(5):491-501.

PhD PORTFOLIO SUMMARY

Summary of PhD training and teaching activities

Name PhD student: Maria Carolina Medina Gomez
Erasmus MC Department: Department of Internal Medicine and Department of Epidemiology
Research School: NIHES: Erasmus University

PhD period: August 2010 – May 2016
Promotor(s): Prof. dr. A.G Uitterlinden
Supervisor: Dr. F. Rivadeneira

1. PhD training

	Year	Workload (Hours/ECTS)
GENERAL ACADEMIC SKILLS		
- Introduction to Medical Writing	2011	1.1
RESEARCH COURSES		
MSc. Genetic Epidemiology	2010	0.7
- Principles of Research in Medicine	2010	0.7
- Clinical Decision Analysis	2010	0.7
- Topics in Meta-analysis	2010	0.7
- Principles of Genetic Epidemiology	2010	4.3
- Study Design	2010	1.4
- Genomics in Molecular Medicine	2010	1.4
- Courses for the for the Quantitative Research	2011	1.4
- Genome Wide Association Analysis	2011	0.7
- Social Epidemiology	2011	5.7
- Genetic-epidemiologic Research Methods	2011	1.4
- SNP´s and Human Disease	2011	1.4
- Advances in Genome-Wide Association Studies	2011	1.4
- Family-based Genetic Analysis	2011	0.9
- Mendelian Randomization		
IN-DEPTH COURSES		
- From DNA to Phenotype	2011	1.4
- A first encounter with next-generation sequencing data	2012	1.4
The Children Hospital of Philadelphia	2014	
- CITI Course in The Protection of Human Research Subjects	2014	
- HIPPA for Research		
NATIONAL CONFERENCES AND PRESENTATIONS		
- **Department of Internal Medicine, Endocrinology research general meetings**	2010-2016	1.0
- **Department of Internal Medicine, Laboratory of Human genetics meetings**	2010-2016	1.0
- **The Generation R Group Research meetings**	2010-2015	1.0
- **Molecular Epidemiology Research meetings**	2014-2015	1.0
- **Wetenschappendagen Internal Medicine Department**	2012-2015	1.5

Poster presentations	2016	1.5
Oral Presentation	2011-2015	0.7
- **Dutch Society of Calcium and bone**	2012	
Oral Presentations		
- **Netherlands Consortium for Healthy Ageing**		
INTERNATIONAL CONFERENCES AND PRESENTATIONS		
- **European Calcified Tissue Society, Athens**	2011	1.2
- **American Society of Bone and Mineral Research, San Diego**	2011	1.5
Oral Presentation	2011	1.5
- **American Society of Human Genetics, Montreal**	2012	0.7
Poster Presentation	2012	1.2
- **EGG meeting, London**	2012	1.5
- **European Calcified Tissue Society of, Stockholm**	2012	1.5
Oral Presentation	2012	0.7
- **American Society of Bone and Mineral Research, Minneapolis**	2012	1.2
Plenary Poster Presentation	2013	1.2
- **American Society of Human Genetics, San Francisco**	2013	0.7
Poster Presentation	2013	1.0
- **CHARGE analysts meeting, Boston**	2013	1.5
- **International Society for Developmental Origins of Health and Disease, Rotterdam**	2014	1.2
Oral Presentation	2015	1.2
- **European Calcified Tissue Society of, Lisbon**	2015	1.0
Poster Presentation	2015	1.5
- **CHARGE analysts meeting, Rotterdam**	2015	1.5
Oral Presentation	2016	1.5
- **6th International Conference of Children's Bone Health, Rotterdam**		
Oral Presentation		
- **American Society of Bone and Mineral Research, Minneapolis**		
Poster Presentation		
- **European Calcified Tissue Society, Prague**		
Oral Presentation		
- **European Calcified Tissue Society, Rotterdam**		
Poster Presentation		
- **7th International Conference of Children's Bone Health, Salzburg**		
Oral Presentation		
- **American Society of Human Genetics, Baltimore**		
Oral Presentation		
- **American Society of Bone and Mineral Research, Seattle**		
Poster Presentation		
- **European Society of Calcified Tissues, Rome**		
Oral Presentation		

Seminars and workshops		
- Web-based Workshop: Repeated Measurements	2014	2.0
- Web-based Workshop: Graphics in R	2014	2.0
- Bioconductor	2015	2.0
- MasterClass: Eric Lander	2015	0.2
2. Teaching activities		
Supervising practicals and excursions		
- SNP course Practicals 2011-2015	2011-2015	2.0
- Genome Wide Association Analysis praticals	2012-2015	2.0
Supervising Master's theses		
- Tamara Pesic,	2015	2.0
- Biomedical Sciences VU Amsterdam University		
Other		
Peer review of articles for scientific journals: Journal of Bone and Mineral Research, Osteoporosis International, Scientific Reports and BoneKEy	2013-2016	2.0

ABOUT THE AUTHOR

Maria Carolina Medina Gomez was born in Bogota, Colombia, on June 26, 1980. In 1997, she completed her secondary education at the Colegio Siervas de San Jose. She obtained her BSc. in Chemical Engineering in 2006 and one year later a second BSc. in Microbiology at Universidad de Los Andes in Bogota, Colombia. During these years her interests moved towards the analysis of experimental biological data. Affinity she confirmed during her stage at the Research group of Peptide Chemistry of the Hungarian Academy of Sciences at ELTE university, combining the knowledge acquired in her two BSc. After working as a high school teacher for two years in Bogota, she enrolled in an MSc in Biostatistics at Hasselt University in Belgium, graduating in 2010. She moved to Leiden to write her MSc. Thesis in "Estimation of genetically homogenous clusters within Europe" at LUMC under the supervision of Professor Janine Houwing and Professor Boehringer. Genetic population substructure is still one of her favorite topics which she developed further during her PhD-project, whose results are presented in this dissertation. In the second semester of 2010, she received the ERACOL grant for her Ph.D. at the Departments of Internal Medicine and Epidemiology. In 2011, she received her NIHES MSc. in Genetic Epidemiology at ErasmusMC and expanded her research project in her current PhD-project entitled "Disentangling the Heterogeneity of Bone Accrual" under the supervision of Prof. dr. Andre G Uitterlinden and Dr. Fernando Rivadeneira. In 2014, a grant from the GeoCoDe mobility program enabled her to work as a research fellow at the Center of Applied Genomics at The children's Hospital of Philadelphia under the supervision Professor Struan Grant and in close collaboration with Professor Babette Zemel. During her Ph.D, she assisted several national and international conferences and her work received multiple investigation awards in both the bone research and genetics fields.

ACKNOWLEDGEMENTS

As extensive as it has been this work is also the number of people who has made it possible for me to accomplished this goal. Both, academically and at a personal level, I have received the company, advice and support to continue with this research work, and I have to admit all this encouragement kept me going on in the hard days - I will not use the metaphor *grey days* because then they would be too numerous-. In the end, I believe this is a collective effort from which many people have taken part and deserve recognition.

First of all, I would like to thank the ERACOL program which sponsored a great extent of my PhD period and to Prof. Andre Uitterlinden for giving me the opportunity to work in his group. Andre, these years on your team have provided me the opportunity of participating in many interesting projects, be at the edge of the state of art of genetic epidemiology, establish contact with leading persons in the field and to assist to different workshops, conferences, meetings, which have undoubtedly provided me the tools to get my PhD. Besides all that, I would also like to show you my gratitude for keeping our group as a big family and encourage us to have that loyalty feeling with each other, making it really special to be part of this group. Of course, I have no words to express my gratitude towards Fernando, I would, in any case, give it a try. Fer, you have been for me much more than a supervisor, and perhaps the least important things I have learned from all these years working under your shelter, are the ones concerning bone biology or clinical interpretation of my findings. You are a great human being and I admire your professionalism and continuous concern to really dig out the best of every person, your will to help and share your knowledge... that makes you really a leader rather than a boss, and is very touching. Thanks for so much, especially for making so palpable every day your faith on me, on my work!!. I know my political skills are not entirely developed yet, I have not been a fast learner as for basic research... but I hope at the end I won't disappoint you. As you have been part of my family, I would like to thank yours for great days and for letting me in.

I am really grateful to the members of the reading committee for their time and feedback. Professor Manfred Kayser, Professor Cornelia van Duijn, and Professor Babette Zemel. I have been able to get familiar with your work, through small collaborations and your publications. Also, I have had the opportunity to get close to people from your respective groups, so I really know how busy you are with your own projects, thanks. I would also like to take the opportunity to thank the members of the external committee Professor Jon Tobias, Professor Claes Ohlsson, and Professor Patrick Groenen, I hope we will have a great discussion.

At the beginning of my PhD I had the blessing to work and learn from one of the sharpest minds I have ever encountered, who shared with me his knowledge and has consistently been present all these years Dr. Karol Estrada. Capo, you have been a great older brother (jjajajajja), thanks for all the advices, the codes, for giving me so much inspiration at work, I should even thank you for marrying Liz from who I also learned a lot. You, together with BeBe are a great family and have warmed my days even from the distance.

Besides Karol, in our group of the laboratory of human genetics, I had the pleasure to work with great scientists and persons, even if with some of you I got closer, and then perhaps will extend a bit longer in my comments, I want to thank you for the excellent feedback not only for my projects but in general for my life, so to Lisette, Marjolein Peters, Marjolein de Kruif, Hanneke, Fatima, Robert, Carola, Joyce, Annemieke, Pooja, Jia Lian, Joost, Annelies, Anke, Cindy, Djawad, Ester, Tom, Linda, Jeroen, Olja, Enisa, Fjorda, BiBi, KT, Tamara, Marta, Nata, Ling, Eskil, and all other students who have joined our lab. Cindy and Linda, I know sometimes I am really not in the best of moods when working... but thanks for being patient, Cindy we will continue jumping I hope for a long time. Ling beauty, I always felt we were such a great team, you are very intelligent and so professional, now you are next!. Marta, usted mujer es una verraca en todo el sentido de la palabra, gracias por el apoyo, por dar siempre la cara, seguro la vida le esta recompensando ese carácter tan templado. Natis, amiguita, eres un ser humano excepcional, yo no sé como agradecerte haber compartido tanta luz conmigo, tenerme siempre en tus oraciones y estar tan pendiente, te quiero. Olja, Enisa, Fjorda, BiBi, KT... what can I say I have an inclination to the east, and I have loved being surrounded by it (through you), you are all such hard workers. Jeroen and KT, I know you were a little bit surprised that I asked you to be my paranymphs.... But I was not, perhaps I failed to show you more often, but I am a great FAN of you!!!, I think you are brilliant, I mean workwise, precise, creative, persistent... but also, I admire you as persons... I believe you are kind and honest, really transparent and that amuse me. Thanks for accepting and investing time on planning all this with me. I also want to recognize the people who are not that visible in our publications but whose work has been vital to our group, Mila, Pascal, Michael, Marijn, Joost, Anis, Eline, Annelies, Liz (again), Thanks.

As the work presented in this book started even before I started the PhD, embedded in the Generation R Study, my gratitude to all participants, doctors, technicians, nurses and sponsors, PIs, without you literally this work would not have been possible. Vincent, it has been excellent working within the Generation R which you lead pretty successfully, thanks for your feedback in all articles which you always had the time to read. Clau, I know I say this quite too often but I would not have known what to do without you... thanks for always being so keen to help me, although, I can imagine you had other hundred students asking you for variables, mismatches, etc. Denise, beauty you worked so hard and thanks to your

professionalism we got high-quality data. I also had the chance to get to know you outside the work-context and you are just as great!, working with you was a pleasure and you write so perfect, I would have liked you to coauthored the discussion of my thesis jajajjajaj.

Other people from the university regularly challenged my knowledge with really smart discussions or questions. So to: Abbas, Paul, Symen, Janine, Claire, Maryam, Romy, Luba, Marjolein, Hieab, Alex, Oscar, Irene, Ralf, Elisa, Najaf, Cornelia, Lennart, Prof. Tiemeier, Prof. Hofman. Thanks for preparing me for the day of the defense (I do not think even the committee will get to the level of some of your inquiries, or at least I hope). Also, I had the opportunity to participate in the endocrinology meetings, where I learned more about bone biology and relevant clinical pathways. This was a great complement to my education in ErasmusMC, so to the organizers and regular participants of these meetings, Jeroen, Bram, Marta, Hans, Marjolein and all the others, thanks.

If you ever look through this thesis you will notice, there is a name enriched in my list of publications: John Kemp. John, you have been the best *comrade-in-arms* during all these years, thanks for all the support, the analyses at the last minute, all the brainstorming, the long Skype calls even with our time difference, the never ending conversations in the different congresses. You are pretty talented and hard worker; I hope we will continue working together. I would also like to thank the whole ALSPAC team, Jon and Dave, thanks for all feedback; I know I write quite a lot of emails... is just that I value your concepts highly. Other people have also contributed their knowledge, time and energy to produce this dissertation, so to all coauthor in the publications not only in which my thesis is based on, but all the ones I have participated on, my sincerest gratitude. Oscar Lao, thanks for being always eager to run an extra mile... at the end we did like a marathon together, and regardless of the amount of lactic acid, I enjoyed it a lot. The GEoCoDE program enabled me to go to the Children Hospital of Philadelphia (CHOP) for a short stage, thanks for the support. Ale, it was awesome meeting you and sharing all my time in Philly working with you. Struan and Babette thank you for your time, for allowing me share ideas in your groups and learn in CHOP. Eskil, it was great having you here and working together... I have not heard from you for so long, hope everything is going fine with you.

I want to acknowledge all the GEFOS/GENOMOS community that has always shown respect for my work and are eager to help, to explore, to follow-up. I would like especially to thank Tom Beck, David Karasik, Evangelia, Vangelis and Cheryl, with who I have worked in close collaboration and always are keen to extend me a hand when I am drowning in my data, analysis or so ever. Other consortia also have brought me a position and I have learned a lot from their TCs or analyst meetings, so to all the participants of CHARGE, EGG and GIANT thanks.

All this Rotterdam adventure, would not have happened if Professor Emmanuel Lessafre, who I met during my Master thesis, would not have invited me to come to Rotterdam for my PhD and established the contact with Fernando. Emmanuel, I would always be in debt with you. Also, in the department of statistics, I had the luck of working with Prof. Paul Eilers, who I would like to thank. Few months before coming to Rotterdam I developed my MSc thesis. In LUMC, I worked with a great team Jeanine, Stefan, Hae-Won, Wendim, and Bru thanks for all the patience and hard work.

I was under a lot of pressure on my job, involved in a lot of projects, feeling homesick sometimes, with my heart half broken in other instances and the lunch or coffee breaks were my buoyancy aid, so to my girls (even if a couple of guy names will indeed be on the list, but with who I felt so free as with girls!). Dasha, Cris, Camila, Mariana, Meli, Ivancho, Aran, Veci, Princess, Tammy, Pau, Lady, Fabi, Sandrita, Giovanni, Monica, Adri and Pachito thanks for making my life lighter by carrying some of the weight of my concerns, for crying and laughing with me, and listen to all that crossed my mind to the limit of ... Lady, amiga gracias por enloquecerte conmigo cada vez... por esconderme el tornillo que me falta pa que nunca lo encuentre. Also, coming to the Netherlands gave me the opportunity of meeting great people who provide me good vibes everytime and have formed a circle around me full of care, thank to you all: Diego, Jessica, Alex, Mark, Juli, Bruna, Moe, Jonathan, Ana, Kristof, Favi, Fatima, Edu (supongo me echaras una porra de vez en cuando dese el cielo) and Marius (you know how much I appreciate the giant amounts of vitamine :D of these years...)

Even before I came here I had established longlife relationships with many people who has never, even in the distance, left me alone and who I try to express day by day my gratitude for their presence in my life, regardless of the place where I live. I know it is not fair that I just place your name in a list, please forgive me, just consider the extension of this book give me a call, and I will be precise in my acknowledgement... My primary- ,secondary- , high-school and still friends: Flaca, Lake, Lili, Sanchino, Cami, Angie, Mayis, Vero, Andre Cardenas, Cata Zapata, Caro Vergara, Meli, Pete, Pau Arevalo and Maria Paula. Mis parceros de la U, Cami, enana, Cata, enano, Germi, Andrecho, Alejo, Roni, Osqui, Malala, Mildi, Orlandito, Carino, Duque,Leo, Edgar, Keka, Magno, Churro, Mona, Adri Caro, Manu, Nico and Carlos Jaramillo (who convinced me in his first lecture that chemical engineering was not the path I wanted to follow). Los amigos de la segunda linea de la mano: Conde, Juanjo, Girata, Fasti, Chiquitico, Benavidez, Vic, Hugui, Carlos Andres, Clau, Mousa, Jan, Gabor, Alex, Cesitar, Juan C. Enciso, JuanDi, Alex, Ernestico. All the people who Belgium gave me the opportunity to meet and who have demonstrated me that the world is pretty small or I am damn lucky: Dona, Consu (who taught me that every cocomero is a good excuse to celebrate), Bea, Ambi, Pia, Ari, Sylvain, Kim, Tania, Martin, Renata, Eva, Jimmy, Kurt, Hellen,

Amparo, Migue, Ema, Urko, Sarita, Lidia, Lio, Laurent, Fortu (calvo, gracias por tanto, fue bello vivir in slow motion the first years of this journey). Surtout, je remercie à Gaby et Mark pour offrir moi a vraie famille donc loin de la Colombie.

After six hours and with the certainty I have forgotten half of the people I would have liked to acknowledge, I want to offer all this work to my family. A mis primos, tíos y abuelo que siempre están tan pendientes de mi y me llenan de amor en cada encuentro. A Ra, Janeth y mis niñas que son los protagonistas de cada salvapantalla en el compu, el cel, la tablet, para que la vida sea mas llevadera. A mi abuelita, que sé que aún hoy está sentada en primera fila con su dulce sonrisa y sus ojos fijos en los mios. To Filippo, who has not only sign the cover of this book but, is also the main author of my sporadic smiles, of the turbulence in my eyes, and who without hesitation enclose me in his arms and there, let me be. To his parents Chiarella and Franco, who have received me naturally in their nucleus and always support me.

Y a mis padres, los que siempre han creído, los que esperan, los que callan, los que cuidan, porque el mayor sacrificio para poder escribir este texto hoy, han sido todos estos años de distancia, de ausencia. Yo sé que ustedes se sienten muy orgullosos de mis logros y si este libro les hace latir fuerte el corazon... entonces valió la pena.