Dissection of the Complex Genetic Architecture of Human Stature and Osteoporosis

By Karol Estrada

DISSECTION OF THE COMPLEX GENETIC ARCHITECTURE OF HUMAN STATURE AND OSTEOPOROSIS

KAROL ESTRADA

Dissection of the Complex Genetic Architecture of Human Stature and Osteoporosis

Karol Estrada

The work presented in this thesis was conducted at the Department of Internal Medicine and the Department of Epidemiology, Erasmus MC University Medical Center Rotterdam. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. Support for genotyping was provided by the Netherlands Organization for Scientific Research (NWO) (175.010.2005.011, 911.03.012), and the NGI/NWO project (050-060-810). This work benefited from financial support of the European Union Seventh Framework Package funded GEnetic Factors for OSteoporosis (FP7-HEALTH- F2-2008-201865-GEFOS) project and the Netherlands Genomics Initiative (NGI)-sponsored (050060810) Netherlands Consortium for Healthy Aging (NCHA).

The publication of this thesis was financially supported by AMGEN B.V., the Dutch Society for Calcium and Bone metabolism (NVCB), the Netherlands Consortium for Healthy Aging (NCHA), the GEnetic Factors for OSteoporosis (GEFOS) and the Erasmus University Rotterdam.

Layout by Karol Estrada, 2012
Cover design by Agustin Quiñonez
Printed by Optima Grafiche Communicatie, Rotterdam

ISBN: 978-94-6169-246-7

Dissection of the Complex Genetic Architecture of Human Stature and Osteoporosis

Ontrafeling van de complexe genetische architectuur van menselijke lengte en osteoporose

Proefschrift

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

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 1 juni 2012 om 15:30 uur

door

Jesus Karol Estrada Gil

geboren te Mexico stad, Mexico

PROMOTIECOMMISSIE

Promotor:	Prof.dr. A.G. Uitterlinden
Overige leden:	Prof.dr. M.H. Kayser
	Prof.dr. R.M.W. Hofstra
	Prof.dr. T.M. Frayling
Copromotor:	Dr. F. Rivadeneira

“Even if it is jade it cracks, even if it is gold it breaks,
even if it is quetzal plumage it tears.
Not forever in the earth: just a little here”
Netzahualcoyotl – poet and ruler of Texcoco (1402-1472)

To Liz,
my son Oscar André,
my family,
and in memory of my father

PUBLICATIONS AND MANUSCRIPTS BASED ON THE STUDIES DESCRIBED IN THIS THESIS

Chapter 2.1
Estrada, K.*, Abuseiris, A.*, Grosveld, F.G., Uitterlinden, A.G., Knoch, T.A. & Rivadeneira, F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. ***Bioinformatics*** 25, 2750-2 (2009).

Chapter 3.1
Estrada, K., Krawczak, M., Schreiber, S., van Duijn, K., Stolk, L., van Meurs, J.B., Liu, F., Penninx, B.W., Smit, J.H., Vogelzangs, N., Hottenga, J.J., Willemsen, G., de Geus, E.J., Lorentzon, M., von Eller-Eberstein, H., Lips, P., Schoor, N., Pop, V., de Keijzer, J., Hofman, A., Aulchenko, Y.S., Oostra, B.A., Ohlsson, C., Boomsma, D.I., Uitterlinden, A.G., van Duijn, C.M., Rivadeneira, F. & Kayser, M. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. ***Hum Mol Genet*** 18, 3516-24 (2009).

Chapter 3.2
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., Ferreira, T., Wood, A.R., Weyant, R.J., Segre, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Magi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Hua Zhao, J., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R., Surakka, I., Tammesoo, M.L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., Konig, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Muller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J., Glorioso, N., Haiqing, S., Hartikainen, A.L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietilainen, K.H., Pouta, A., Ridderstrale, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Oliver Schmidt, C., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimaki, T., Melander, O., Mosley, T.H., Jr., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G.J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Gronberg,

H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Mark Lathrop, G., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Adrienne Cupples, L., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Volzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis, G.R.*, Stefansson, K.*, Frayling, T.M.* & Hirschhorn, J.N.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. ***Nature*** 467, 832-8 (2010).

Chapter 4.1

Rivadeneira, F.*, Styrkarsdottir, U.*, **Estrada, K.***, Halldorsson, B.V.*, Hsu, Y.H.*, Richards, J.B.*, Zillikens, M.C.*, Kavvoura, F.K.*, Amin, N., Aulchenko, Y.S., Cupples, L.A., Deloukas, P., Demissie, S., Grundberg, E., Hofman, A., Kong, A., Karasik, D., van Meurs, J.B., Oostra, B., Pastinen, T., Pols, H.A., Sigurdsson, G., Soranzo, N., Thorleifsson, G., Thorsteinsdottir, U., Williams, F.M., Wilson, S.G., Zhou, Y., Ralston, S.H., van Duijn, C.M.*, Spector, T.*, Kiel, D.P.*, Stefansson, K.*, Ioannidis, J.P.* & Uitterlinden, A.G.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. ***Nat Genet*** 41, 1199-206 (2009).

Chapter 4.2

Estrada, K.*, Styrkarsdottir, U.*, Evangelou, E.*, Hsu, Y*., Duncan, E.L.*, Ntzani, E.E.*, Oei, L.*, Albagha, O.M.E., Amin, N., Kemp, J.P., Koller, D.L., Li, G., Liu, C., Minster, R.L., Moayyeri, A., Vandenput, L., Willner, D., Xiao, S., Yerges-Armstrong, L.M., Zheng, H., Alonso, N., Eriksson, J., Kammerer, C.M., Kaptoge, S.K., Leo, P.J., Thorleifsson, G., Wilson, J.F., Wilson, S.G., Aalto, V., Alen, M., Aragaki, A.K., Aspelund, T., Center, J.R., Dailiana, Z., Duggan, D.J., Garcia, M., García-giralt, N., Giroux, S., Hallmans, G., Hocking, L.J., Husted, L.B., Jameson, K.A., Khusainova, R., Kim, G., Kooperberg, C., Koromila, T., Kruk, M., Laaksonen, M., Lacroix, A.Z., Lee, S., Leung, P.C., Lewis, J.R., Masi, L., Mencej-Bedrac, S., Nguyen, T.V., Nogues, X., Patel, M.S., Prezelj, J., Rose, L.M., Scollen, S., Siggeirsdottir, K., Smith, A.V., Svensson, O., Trompet, S., Trummer, O., van Schoor, N.M., Woo, J., Zhu, K., Balcells, S., Brandi, M., Buckley, B.M., Cheng, S., Christiansen, C., Cooper, C., Dedoussis, G., Ford, I., Frost, M., Goltzman, D., González-Macías, J., Kähönen, M., Karlsson, M., Khusnutdinova, E., Koh, J., Kollia, P., Langdahl, B.L., Leslie, W.D., Lips, P., Ljunggren, Ö., Lorenc, R.S., Marc, J., Mellström, D., Obermayer-Pietsch, B., Olmos, J.M., Pettersson, U., Reid, D.M., Riancho, J.A., Ridker, P.M., Rousseau, F., Slagboom, P.E., Tang, N.L.S., Urreizti, R., van Hul, W., Viikari, J., Zarrabeitia, M.T., Aulchenko , Y.S., Castano-Betancourt, M., Grundberg, E., Herrera, L., Ingvarsson, T., Johannsdottir, H., Kwan, T., Li, R., Luben, R., Medina-Gómez, C., Palsson, S.Th., Reppe, S., Rotter, J.I., Sigurdsson, G., van Meurs, J.B.J., Verlaan, D., Williams, F.M.K., Wood, A.R., Zhou, Y., Gautvik, K.M., Pastinen, T., Cauley, J.A., Chasman, D.I., Clark, G.R., Cummings, S.R., Danoy, P., Dennison, E.M., Eastell, R., Eisman, J.A., Gudnason, V., Hofman, A., Jackson, R.D., Jones, G., Jukema, J.W., Khaw, K., Lehtimäki, T., Liu, Y., Lorentzon, M., McCloskey, E., Mitchell, B.D., Nandakumar, K., Nicholson, G.C., Oostra, B.A., Peacock, M., Pols, H.A.P., Prince, R.L., Raitakari, O., Reid, I.R., Robbins, J., Sambrook, P.N., Sham, P.C., Shuldiner , A.R., Tylavsky, F.A., van Duijn, C.M., Wareham, N.J., Cupples, L.A., Econs, M.J., Evans, D.M., Harris, T.B., Kung, A.W.C., Psaty, B.M., Reeve, J., Spector, T.D., Streeten, E.A., Zillikens, M., Thorsteinsdottir, U.*, Ohlsson, C.*, Karasik, D.*, Richards, J.B.*, Brown, M.A.*, Stefansson, K.*, Uitterlinden, A.G.*, Ralston, S.H.*, Ioannidis, J.P.A.*, Kiel, D.P.*, Rivadeneira, F.*,Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. ***Nat Genet*** advance online publication, 15 April 2012; doi:10.1038/ng.2249.

Chapter 4.3

Liu, C-T.*, **Estrada, K.***, Yerges-Armstrong, L.M., Amin, N., Evangelou, E., Li, G., Minster, R.L., Carless, M.A., Kammerer, C.M., Oei, L., Zhou, Y., Alonso, N., Dailiana, Z., Eriksson, J., García-Giralt, N., Giroux, S., Husted, L.B., Khusainova, R.I., Koromila, T., Kung, A.W.C., Lewis, J.R., Masi, L., Mencej-Bedrac, S., Nogues, X., Patel, M.S., Prezelj, J., Richards, J.B., Sham, P.C., Spector, T., Vandenput, L., Xiao, S-M., Zheng, H-F., Zhu, K., Balcells, S., Brandi, M.L., Frost, M., Goltzman, D., González-Macías, J., Karlsson, M., Khusnutdinova, E.K., Kollia, P., Langdahl, B.L, Ljunggren, Ö., Lorentzon, M., Marc, J., Mellström, D., Ohlsson, C., Olmos, J.M., Ralston, S.H., Riancho, J.A., Rousseau, F., Urreizti, R., Hul, W.v., Zarrabeitia, M.T., Castano-Betancourt, M., Demissie, S., Grundberg, E., Herrera, L., Kwan, T., Medina-Gómez, C., Pastinen, T., Sigurdsson, G., Thorleifsson, G., vanMeurs, J.B.J., Blangero, J., Hofman, A., Liu, Y., Mitchell, B.D., O'Connell, J.R., Oostra, B.A., Rotter, J.I., Stefansson, K., Streeten, E.A., Styrkarsdottir, U., Thorsteinsdottir, U., Tylavsky, F.A., Uitterlinden, A., Cauley, J.A., Harris, T.B., Ioannidis, J.P.A., Psaty, B.M., Robbins, J.A., Zillikens, M.C., van Duijn, C.M., Prince, R.L., Karasik, D.*, Rivadeneira, F.*, Kiel, D.P.*, Cupples, L.A.*, Hsu, Y-H.*Assessment of Gene-by-Sex Interaction Effect on Bone Mineral Density (Submitted).

Chapter 4.4

Estrada, K., Hsu, Y-H., Styrkarsdottir, U., Nandakumar, K., Eussen, B.H., de Klein, A., Peters, M.J., Oei, L., Halldorsson, B., Liu, C-T., Alonso, N., Kaptoge, S.K., Thorleifsson, G., Hallmans, G., Hocking, L.J., Husted, L.B., Jameson, K.A., Kruk, M., Lewis, J.R., Patel, M.S., Scollen, S., Svensson, O., Trompet, S., van Schoor, N.M., Zhu, K., Buckley, B.M., Cooper, C., Ford, I., Goltzman, D., González-Macías, J., Langdahl, B.L., Leslie, W.D., Lips, P., Lorenc, R.S., Olmos, J.M., Pettersson, U., Reid, D.M., Riancho, J.A., Slagboom, P.E., Zarrabeitia, M.T., Ingvarsson, T., Johannsdottir, H., Luben, R., Medina-Gómez, C., Arp, P., Palsson, S.Th. , Sigurdsson, G., van Meurs, J.B.J., Zhou, Y., Hofman, A., Jukema, J.Wouter, Pols, H.A.P., Prince, R.L, Cupples, L.Adrienne, Marshall, C.R., Pinto, D., Sato, D., Scherer, S.W. , Reeve, J., Thorsteinsdottir, U., Karasik, D., Richards, J.Brent, Stefansson, K., Uitterlinden, A.G., Ralston, S.H, Ioannidis, J.P.A., Kiel, D.P., Rivadeneira, F. Genome-wide copy number association study identifies a rare deletion on chromosome 6p25.1 associated with osteoporotic fractures. (Submitted).

*These authors contributed equally

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Osteoporosis is a common complex disease characterized by decreased bone mineral density (BMD) and microarchitectural deterioration of bone tissue that eventually leads to increased risk of fracture. The most common type of fracture is hip fracture, a severe disabling event leading to decreased mobility and increased mortality. A women aged 65 years or older has a greater risk of dying after a hip fracture event than after the diagnosis of breast cancer (48.1% and 25.1% mortality rate respectively).[1] The annual cost of treating all types of fracture has been estimated to be $17 billion in the United States and is expected to increase 50% by 2025.[2] These ever increasing costs represent a burden for the already costly budget for public health in our graying society.

Epidemiological studies have shown that elderly women are in the highest risk category for fracture. But being young and/or men does not preclude from suffering a fracture. BMD, the best predictor of fracture risk, is used for the diagnosis of osteoporosis, defined as values at the femoral neck or lumbar spine (Fig. 1) below 2.5 standard deviations the population average in young (30 years) healthy individuals[3]. The recommended site for diagnosis is the proximal femur with dual energy X-ray absorptiometry (DXA)[3] . Height has also been established as an important risk factor for fracture. People in the highest quartile of stature have 2.4 higher risk of having a hip fracture as compared to the people in the lowest quartile.[4]

Another clinical factors influencing the risk of osteoporosis include medication use (such as glucocorticoids), diet, smoking, physical activity and low weight, but one of the most important risk factors is a positive family history[3], emphasizing the importance of genetics in the pathogenesis of the disease. Family and twin studies have estimated the heritability of BMD and height to be 66% and 80%, respectively.[5,6] On the other hand, fracture risk is a very heterogeneous phenotype with twin studies showing a large spread in the estimates of fracture heritability. The largest study (and probably the one providing the most reliable estimates) included 33,432 Swedish twins and reported a heritability of any type of fracture of 16%, increasing to 27% for osteoporotic fractures and 46% for hip fractures. Interestingly, this study also showed that those fractures occurring above the age of 79 years have a drastically lower heritability of only 3%.[7]

The study of genetic factors conferring risk to human diseases provides new biological insight on disease etiology. This new knowledge might translate into novel clinical applications such as detecting potential therapeutic targets or helping improve the diagnosis and the prognosis of disease.[8-10] Different strategies have been proposed for the detection of genetic variants associated with human traits. In this thesis, we have focused on the Genome-wide association study (GWAS) approach discussed below.

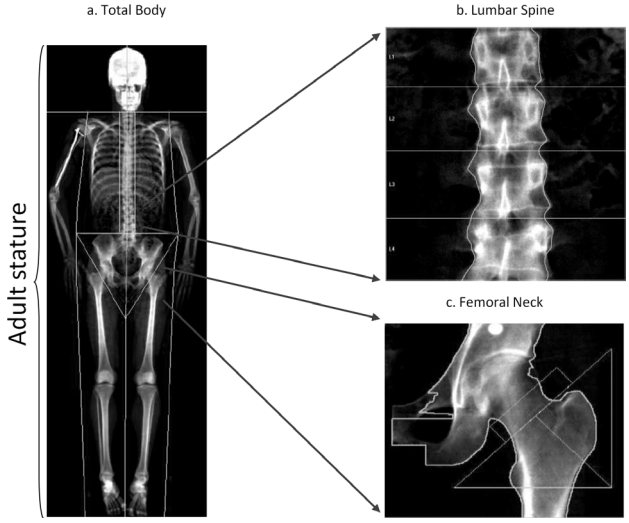


Figure 1. Dual energy X-ray absorptiometry (DXA) images of A) Total Body, B) Lumbar Spine and C) Femoral Neck. The latter two sites are usually studied for the diagnosis of osteoporosis. The measurement of stature (height) is depicted in A).

Genome-wide association studies (GWAS)

The GWAS approach combines the power of the association analysis with the advantage of being a hypothesis-free design. It uses SNP microarray technology, where hundreds of thousands to millions of variants are tested in one single experiment. Therefore, the advent of the GWAS approach was facilitated by the confluence of contributions arising from five main instances including:

1. Human Genome Project[11-13]
2. International SNP Map Working Group[14]
3. International HapMap Project[15]
4. Developments in high-throughput genotyping technologies
5. Availability of well characterized cohorts studies

The first three points above reflect projects that created the basis for the selection and design of assays to effectively genotype the most common genetic variants of the human genome. In 2005, the HapMap project provided a catalogue of human common variation and estimated that approximately 300,000 SNPs would be sufficient to "tag" or capture a large fraction of the genetic variation present in individuals of European descent. Shortly after, the first SNP microarrays

emerged and the first locus associated with common complex diseases (i.e. macular degeneration) was reported in the literature.[16] However, the GWAS approach has an important limitation: because of the large amount of genetic variants tested, a strict significance threshold has to be applied to avoid a high false positive rate. This threshold is estimated to be $5x10^{-8}$, which reflects a Bonferroni correction of testing one million variants (i.e., 0.05/1,000,000)[17].

Genetic studies of human stature

Human stature is an easily measured and highly heritable (estimated to approximately 80%) classic phenotype[18-20]. Rare mutations with severe effects have been described for either extremely short or extremely tall individuals (Fig. 2). However, before 2007 limited knowledge had been achieved regarding the genetics controlling normal height variation in humans. Linkage analyses in different sets of populations showed little overlap in loci nominally linked to normal height variation[21,22]. In 2007, a study selected 150 tag SNPs from eight candidate genes in the growth hormone/insulin-like growth factor-1 axis to evaluate their association with height in four cohorts (N=6,075)[23]. While being a well-powered setting and having selected SNPs based on Linkage Disequilibrium (LD) to increase coverage, this study did not find any of the 150 SNPs consistently associated with height, showing that prior biological knowledge does not determine having common genetic variants associated with this complex phenotype.

The first common variant associated with height (near the *HMGA2* gene) was reported in 2007 from a GWAS on 4,921 samples with a replication in 29,098 individuals[24]. A similar GWAS with 6,669 samples in the discovery and 28,801 in the replication found variants near the *GDF5* gene associated with height[25]. Based on this success, three simultaneous papers with sample sizes ranging from 15,000 to 27,000 samples in discovery plus additional 5,000 to 16,000 in the replication, identified in total 44 loci associated with height[26-28] (Fig. 2). However, even these three relatively large studies did not have a large overlap of associated loci, actually only four loci were significantly in the three studies. Considering the small effect size conferred by each variant, and the fact that these 44 variants together explained 5% of the variance, it was suggested that many more height associated variants were yet to be identified in larger studies[22]. This early success in identifying common genetic variants associated with this complex trait inspired similar studies on other quantitative traits such as bone mineral density.

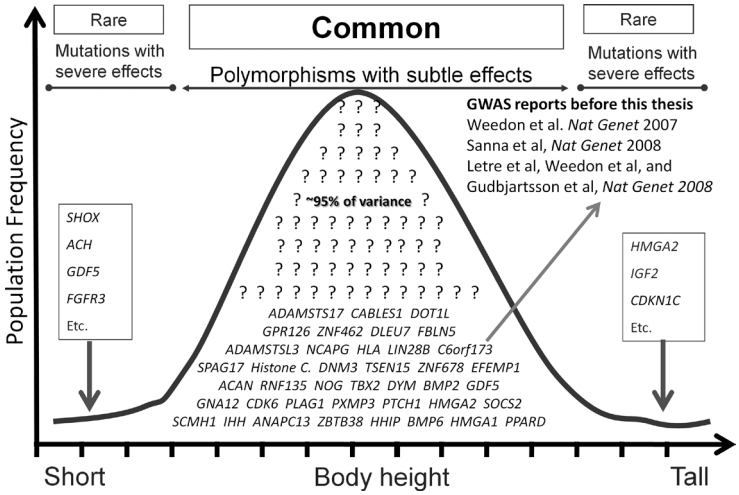


Figure 2. Genes associated with either extremely short, extremely tall or normal body height variation identified by different methods such as linkage, candidate gene or GWAS.

Genetic studies of osteoporosis

As in other disorders with Mendelian inheritance, linkage analysis has been successful identifying genes underlying monogenetic forms of bone diseases. These entities range from those presenting with low bone mass and fragility (Osteogenesis Imperfecta[29], Osteoporosis pseudoglioma syndrome[30]) to those characterized by high bone mass (Sclerosteosis[31] and van Buchem disease[32], Osteopetrosis[33], High Bone mass syndrome[34]) (Fig. 3). However, despite large sample sizes, linkage analysis has not been successful in identifying loci for the common form of osteoporosis.[35] This reflects that linkage analysis is a powerful approach for detecting rare genetic variants of large effect in monogenic syndromes, while it is not powered to identify genetic variants contributing to the etiology of common complex diseases.[36]

Several candidate gene studies were attempted for the identification of novel genes for osteoporosis, however, most of them were of small sample size and therefore underpowered, thus yielding contradictory results. Only when the right variants in the right gene are tested with enough sample size, a true association can be found. Several candidate genes such as *ESR1*[37], *COL1A1*[38], *VDR*[39] and *TGFB1*[40] were tested within the GENOMOS consortium. While these associations have been replicated at $P < 0.05$,[40-42] it has been *LRP5*[43] the only candidate gene associated at $P < 5x10^{-8}$. While mutations in LRP5 present with extreme (high and low) BMD values, two non-synonymous coding polymorphisms in exons 9 and 18 were shown to be

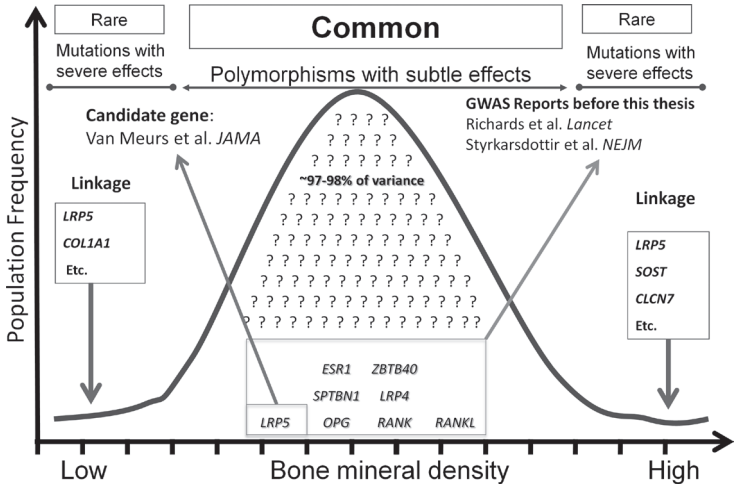


Figure 3. Genes associated with either extreme or normal bone mineral density variation identified by different methods such as linkage, candidate gene or GWAS.

influencing BMD variation at the population level (Fig. 3). The latter association was discovered using the candidate gene approach in an adequately powered setting and has been consistently and widely replicated. Thus, we can consider it as the most successful application of the candidate gene approach in the field of osteoporosis. In 2009, a comprehensive systematic assessment of 150 osteoporosis candidate genes in 19,000 individuals found variants in only nine genes as significantly associated with BMD.[44] These genes included SNPs within or close to *ITGA1*, *LRP5*, *SOST*, *SPP1*, *TNFRSF11A*, *TNFRSF11B*, and *TNFSF11*, confirming a fraction of the genes identified by a meta-analysis of genome-wide association studies done on the same set of individuals.

The first two successful GWAS papers for osteoporosis appeared early in 2008 almost simultaneously. The first one was based on GWAS data from the TwinsUK and Rotterdam Study cohorts summing to 8,558 samples after replication identifying variants near *OPG* and *LRP5* as associated at genome-wide significant level.[45] The second one was a study from DeCode[46] with a GWAS discovery set based on 5,861 samples from Iceland which reported variants mapping to RANK, RANKL, OPG, ESR1, ZBTB40, SPTBN1, LPR4 (Fig. 3). Associations with fracture risk were identified for ZBTB40[46], SPTBN1[46] and LRP5[45].

The variance explained by each of these variants was rather small (less than 0.6%), explaining in total less than 2% of the variance of BMD (Fig. 3). At that moment it was evident that many more variants were to be discovered by increasing the

sample size of the discovery. However, important methodological challenges had to be overcome in order to perform such studies.

Meta-analysis of imputation-driven GWAS data

One of the biggest challenge of doing a GWAS meta-analysis is that studies use different genotyping platforms which share only a limited subset of SNPs (i.e., Affymetrix 500K and Illumina 550K have only ~10% of their markers in common). The current and standard solution is to use a shared reference set of haplotypes (such as the HapMap CEU panel) to impute missing genotypes using different computational methods such as BIM-BAM[47] , MACH[48] or IMPUTE[49]. Once the data has been imputed, the analysis is performed for each SNP using either linear or logistic regression with the allele dosage as predictor and the phenotype as outcome. The computational time required to analyze one phenotype against ~ 2.5 million SNPs in 6,000 samples is approximately 4 hrs in a regular computer. Yet, it is usual that hundreds of phenotypes (and many potential covariates) are available for genetic studies. Also, larger reference sets are becoming available (such as the 1000 Genomes project) increasing the number of variants to be tested by more than one order of magnitude, currently adding to 40 million markers. Therefore, suitable analytical pipelines need to be implemented to overcome these constraints, including those allowing high-throughput parallel analysis of datasets. Once the data is analyzed, association results are shared between studies to perform a formal meta-analysis using either an inverse variance or a Z-score based meta-analysis. These and other best practices to avoid the presence of false positives derived from being underpowered, plating, genotyping, or imputation artifacts have been described elsehwere.[50]

Consortia of genetic research

As explained above, usually several thousand of samples are required in a discovery data set in order to identify genetic variants at the strict threshold of genome-wide significance. Most of the individual GWAS studies are currently composed of 500 to 5000 samples, therefore they need to join in consortia where researchers share a common goal. Most of the studies described in this thesis have been performed with the collaboration from members of the **G**enetic **I**nvestigation of **AN**thropometric **T**raits (GIANT)[51], the **N**etherlands **C**onsortium for **H**ealthy **A**geing (NCHA)[52], and the **GE**netic **F**actors for **OS**teoporosis (GEFOS)[53] consortia.

Aims and outline of this thesis

This thesis contains studies that aimed to identify novel genetic risk factors for adult human stature and osteoporosis. We have focused particularly on the study of two widely available and highly heritable traits: height and bone mineral density. These studies have been conducted applying the GWAS approach and used the meta-analysis of imputed datasets.

In **Chapter 2**, we focused on the implementation of analytical methods for GWAS in an interface which allows high-throughput analysis of phenotypes. We describe the implementation of a web-based pipeline allowing the use of GRid computing resources for the analysis of genome-wide association IMPuted datasets (GRIMP). GRIMP was a pillar for all the analyses described in this thesis. **Chapter 3**, consists of two studies focusing on the identification of common variants associated with adult height. The first is based on individuals of Northwestern European ancestry (**Chapter 3.1**), while the second represents the largest GWA meta-analysis done for adult height (**Chapter 3.2**) to date. **Chapter 4** contains studies on the genetics of osteoporosis. **Chapter 4.1** and **Chapter 4.2** focus on the identification of common variants associated with BMD and fracture risk using a GWAS approach. **Chapter 4.3** attempts to identify gene-by-sex interactions for BMD. In **Chapter 4.4** we assessed the effect of Copy Number Variants (CNVs) on fracture risk. Finally, **Chapter 5** provides a general discussion.

REFERENCES

1. Cauley, J.A. Abstract P28. in *American Geriatrics Society* (2007).
2. Burge, R., Dawson-Hughes, B., Solomon, D.H. *et al.* Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res* **22**, 465-75 (2007).
3. Kanis, J.A. Diagnosis of osteoporosis and assessment of fracture risk. *Lancet* **359**, 1929-36 (2002).
4. Hemenway, D., Feskanich, D. & Colditz, G.A. Body height and hip fracture: a cohort study of 90,000 women. *Int J Epidemiol* **24**, 783-6 (1995).
5. Sigurdsson, G., Halldorsson, B.V., Styrkarsdottir, U., Kristjansson, K. & Stefansson, K. Impact of genetics on low bone mass in adults. *J Bone Miner Res* **23**, 1584-90 (2008).
6. Visscher, P.M., Medland, S.E., Ferreira, M.A. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* **2**, e41 (2006).
7. Michaelsson, K., Melhus, H., Ferm, H., Ahlbom, A. & Pedersen, N.L. Genetic liability to fractures in the elderly. *Arch Intern Med* **165**, 1825-30 (2005).
8. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five Years of GWAS Discovery. *Am J Hum Genet* **90**, 7-24 (2012).
9. Potter, L.R. Regulation and therapeutic targeting of peptide-activated receptor guanylyl cyclases. *Pharmacol Ther* **130**, 71-82 (2011).
10. Prickett, T.C., Bothwell, J.C., Yandle, T.G., Richards, A.M. & Espiner, E.A. Pharmacodynamic responses of plasma and tissue C-type natriuretic peptide to GH: correlation with linear growth in GH-deficient rats. *J Endocrinol* **212**, 217-25 (2012).
11. Venter, J.C., Adams, M.D., Myers, E.W. *et al.* The sequence of the human genome. *Science* **291**, 1304-51 (2001).
12. Lander, E.S., Linton, L.M., Birren, B. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
13. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
14. Sachidanandam, R., Weissman, D., Schmidt, S.C. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-33 (2001).
15. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
16. Klein, R.J., Zeiss, C., Chew, E.Y. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-9 (2005).
17. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).
18. Fisher, R.A. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918).
19. Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* **5**, 329-348 (1885).
20. Silventoinen, K., Sammalisto, S., Perola, M. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* **6**, 399-408 (2003).

21. Perola, M., Sammalisto, S., Hiekkalinna, T. *et al.* Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet* **3**, e97 (2007).

22. Weedon, M.N. & Frayling, T.M. Reaching new heights: insights into the genetics of human stature. *Trends Genet* **24**, 595-603 (2008).

23. Lettre, G., Butler, J.L., Ardlie, K.G. & Hirschhorn, J.N. Common genetic variation in eight genes of the GH/IGF1 axis does not contribute to adult height variation. *Hum Genet* **122**, 129-39 (2007).

24. Weedon, M.N., Lettre, G., Freathy, R.M. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* **39**, 1245-1250 (2007).

25. Sanna, S., Jackson, A.U., Nagaraja, R. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* **40**, 198-203 (2008).

26. Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G. *et al.* Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**, 609-15 (2008).

27. Lettre, G., Jackson, A.U., Gieger, C. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* **40**, 584-91 (2008).

28. Weedon, M.N., Lango, H., Lindgren, C.M. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575-83 (2008).

29. Sykes, B., Ogilvie, D., Wordsworth, P. *et al.* Consistent linkage of dominantly inherited osteogenesis imperfecta to the type I collagen loci: COL1A1 and COL1A2. *Am J Hum Genet* **46**, 293-307 (1990).

30. Gong, Y., Vikkula, M., Boon, L. *et al.* Osteoporosis-pseudoglioma syndrome, a disorder affecting skeletal strength and vision, is assigned to chromosome region 11q12-13. *Am J Hum Genet* **59**, 146-51 (1996).

31. Balemans, W., Van Den Ende, J., Freire Paes-Alves, A. *et al.* Localization of the gene for sclerosteosis to the van Buchem disease-gene region on chromosome 17q12-q21. *Am J Hum Genet* **64**, 1661-9 (1999).

32. Van Hul, W., Balemans, W., Van Hul, E. *et al.* Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12-q21. *Am J Hum Genet* **62**, 391-9 (1998).

33. Benichou, O., Cleiren, E., Gram, J. *et al.* Mapping of autosomal dominant osteopetrosis type II (Albers-Schonberg disease) to chromosome 16p13.3. *Am J Hum Genet* **69**, 647-54 (2001).

34. Johnson, M.L., Gong, G., Kimberling, W. *et al.* Linkage of a gene causing high bone mass to human chromosome 11 (11q12-13). *Am J Hum Genet* **60**, 1326-32 (1997).

35. Ioannidis, J.P., Ng, M.Y., Sham, P.C. *et al.* Meta-analysis of genome-wide scans provides evidence for sex- and site-specific regulation of bone mass. *J Bone Miner Res* **22**, 173-83 (2007).

36. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-7 (1996).

37. Becherini, L., Gennari, L., Masi, L. *et al.* Evidence of a linkage disequilibrium between polymorphisms in the human estrogen receptor alpha gene and their relationship to bone mass variation in postmenopausal Italian women. *Human molecular genetics* **9**, 2043-50 (2000).

38. Grant, S.F., Reid, D.M., Blake, G. *et al.* Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I alpha 1 gene. *Nat Genet* **14**, 203-5 (1996).

39. Nejentsev, S., Godfrey, L., Snook, H. *et al.* Comparative high-resolution analysis of linkage disequilibrium and tag single nucleotide polymorphisms between populations in the vitamin D receptor gene. *Human molecular genetics* **13**, 1633-9 (2004).

40. Fang, Y., van Meurs, J.B., d'Alesio, A. *et al.* Promoter and 3'-untranslated-region haplotypes in the vitamin d receptor gene predispose to osteoporotic fracture: the rotterdam study. *Am J Hum Genet* **77**, 807-23 (2005).

41. Ioannidis, J.P., Ralston, S.H., Bennett, S.T. *et al.* Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *JAMA* **292**, 2105-14 (2004).

42. Uitterlinden, A.G., Burger, H., Huang, Q. *et al.* Relation of alleles of the collagen type Ialpha1 gene to bone density and the risk of osteoporotic fractures in postmenopausal women. *N Engl J Med* **338**, 1016-21 (1998).

43. van Meurs, J.B., Trikalinos, T.A., Ralston, S.H. *et al.* Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. *JAMA* **299**, 1277-90 (2008).

44. Richards, J.B., Kavvoura, F.K., Rivadeneira, F. *et al.* Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern Med* **151**, 528-37 (2009).

45. Richards, J.B., Rivadeneira, F., Inouye, M. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).

46. Styrkarsdottir, U., Halldorsson, B.V., Gretarsdottir, S. *et al.* Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* **358**, 2355-65 (2008).

47. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet* **3**, e114 (2007).

48. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).

49. 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* **39**, 906-13 (2007).

50. de Bakker, P.I., Ferreira, M.A., Jia, X. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human molecular genetics* **17**, R122-8 (2008).

51. Genetic Investigation of ANthropometric Traits. http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium. (2012).

52. Netherlands Consortium for Healthy Ageing, N. http://www.healthy-ageing.nl/. (2012).

53. Genetic Factors for Osteoporosis. http://www.gefos.org. (2012).

CHAPTER 2

ANALYTICAL METHODS FOR GWAS

CHAPTER 2.1

GRIMP: a tool for high-speed analysis of large-scale genome-wide association data

Bioinformatics 25, 2750-2 (2009)

ABSTRACT

The current fast growth of genome-wide association studies (GWAS) combined with now common computationally expensive imputation requires the online access of large user groups to high-performance computing resources capable of analyzing rapidly and efficiently millions of genetic markers for ten thousands of individuals. Here, we present a web-based interface—called GRIMP—to run publicly available genetic software for extremely large GWAS on scalable super-computing grid infrastructures. This is of major importance for the enlargement of GWAS with the availability of whole-genome sequence data from the 1000 Genomes Project and for future whole-population efforts.

INTRODUCTION

By 2008 more than 150 associations between common genetic variants and human complex traits and disease have been successfully identified through the use of GWAS[1]. It rapidly became evident that very large samples sizes are required to detect variants with modest genetic effects (e.g. a study requires ~8,600 samples to have 90% of power to find genetic variants with a frequency of 0.20, an odds ratio of 1.2 and a genome wide significance of 10^{-8}). Such study sizes are achieved by meta-analysis of data shared collaboratively in consortia analyzing 100s of traits in >40000 individuals[2]. Since they use different genotyping platforms (e.g Affymetrix, Illumina), imputation of millions of markers from a reference (e.g. a HapMap population) is required[3-4]. Statistical methods as linear or logistic regressions measure marker-wise the actual association of the genetic variants with quantitative and binary diseases and traits. Freely-available software like MACH2QTL/MACH2DAT[5], SNPTEST[6], or ProbABEL[7] perform similarly well for these analyses and allow trivial parallelization for distributed computing: The computation time on a regular computer for one continuous trait (~2.5 million markers, ~6,000 samples) is currently ~6h. Assuming linear scaling future studies with ~50 million markers from genome sequencing in 10^5-10^6 samples and even low (1%) allele frequencies can result in ~85-850 days of analysis. Thus, secure, fast accessible web services and scalable high-performance computing grid infrastructures as the Erasmus Computing Grid[8] or the German MediGRID[9] are required to make this analysis feasible.

Here, we present a web-based interface and application to run publicly-available genetic software for extremely large GWAS on such super-computing grid infrastructures. Consequently, we provide a solution to analyze GWAS in very large populations.

IMPLEMENTATION

To achieve high-speed result delivery, the work is split and distributed on different grid processors by trivial parallelization depending on the total data amount. The complete system consists of i) the user remote access computer, ii) a webserver with user webservices and a data/application database, iii) a submit machine with a job handler and a grid resource database, and iv) grid resources with head nodes and execution nodes. The implementation consists of a hardened linux system, which has a hardened apache2 webserver and an PostgreSQL database. Php is used for the website and the job-handler is scripted in Perl. Concerning security, data

transmission is encrypted and complete user separation is applied. Currently, the system administrator manages user accounts and monitors user access, job status and statistics. He also uploads the GWA imputed data to all available grid head nodes for each genotyped cohort, since it is of large size and the same for all cohort phenotypes. Thus, only the phenotype information has to be uploaded by the GRIMP user to the system, which controls the detailed workflow (Fig. 1).

2.1 User package submission: After logging into the system the users manually specify the analysis details: They label the analysis and select a regression model (currently linear and logistic models), dataset and optionally a gender stratified or combined analysis. Additional individual-phenotype links and phenotype specifying annotation files can now be uploaded to the database as well. Further covariates (specified in the phenotype file) can also be annotated. After choosing the progress notification scheme, the user submits the process package.

2.2 Package pre-processing: The phenotype file is transformed to fit the format required by the analytical application implemented (currently mach2qtl and mach2dat for linear and logistic regression, MACH). In principle any GWA analysis software can be used here and installed in the application database.

2.3 Job submission to the grid infrastructure: An implemented job handler periodically checks the database for newly submitted packages and also checks for the workload on the grid head nodes for available capacity to split the packages properly into jobs to be distributed to an individual grid part. To avoid queue overflow, each head node has a predefined amount of jobs that can be queued. Thereafter, the job handler creates a submit file and packages to be uploaded to the individual grid head node. The local respective grid middleware will handle the jobs of the package for these specific grid infrastructures. Currently, we use here the Globus toolkit, but in principle any grid driving middleware can be used here. For high-speed delivery the individual jobs have highest priority compared to other and filler jobs.

2.4 Job/package monitoring: The job handler checks every five min the database for sent jobs and verifies the current status of the individual jobs distributed to a CPU through the middleware on the specific grid head node. An individual failed job is resubmitted up to three times. After all individual jobs of a package are completed, the results are uploaded to the database and the package on the head node is removed. In case of complete failure, the job handler will remove all jobs of the package on the head node including the uploaded package and a failure notification is sent.

2.5 Package post-processing and notification: Once all jobs of a package were finished, all individual result files are combined into one file together with additional

marker annotations such as chromosome, position, allele frequency, sample size and quality of the imputed markers. The results are archived in the database for later analysis and the result files are compressed to save disk space. Depending on the choice of notification the user is now informed - e.g. by email.

RESULTS AND CONCLUSIONS

Via a web-based interface the successful implementation of GRIMP allows to use publicly available genetic software for very large GWAS on scalable super-computing grid infrastructures such as the Erasmus Computing Grid or the German MediGRID within the hour. The analysis of ~2.5 million markers and ~6,000 samples takes now ~12 min in contrast to ~6h. For ~10^7 markers and ~10^5 samples we achieve ~10 to 20 min, in contrast to ~400 h, i.e. ~17 days for a single CPU. Thus, GRIMP will improve the learning curve for new users and will reduce human errors involved in the management of large databases. Consequently, researchers and other users with little experience will largely benefit from the use of high-performance grid computing infrastructures. Since each Grid infrastructure has different middleware setups, adjustments might be needed for each particular GRIMP implementation.

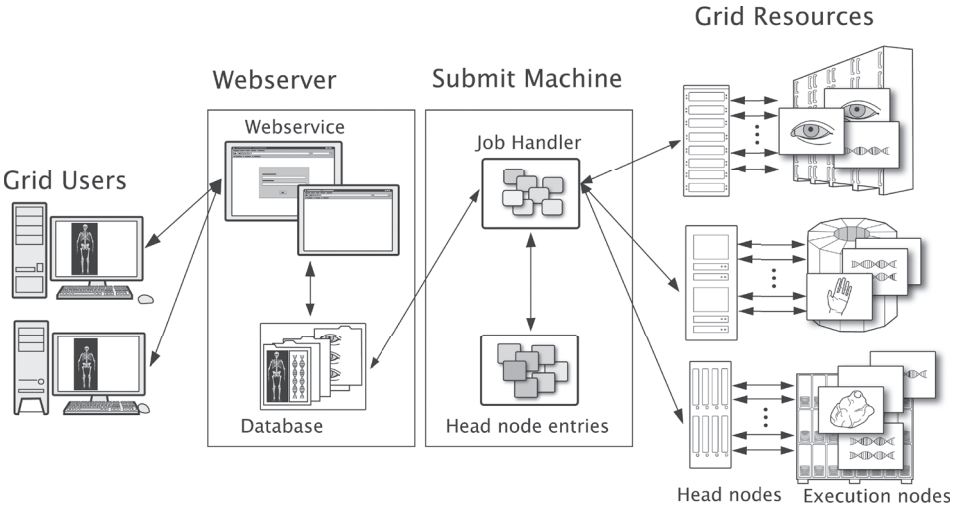


Figure 1. Structure of the work-flow of GRIMP consisting of i) remote user access, ii) a webserver with webservices and a data/application database, iii) a submit machine with job handler and grid resource database, and iv) grid resources with head nodes and execution nodes.

ACKNOWLEDGEMENTS

We thank Luc V. de Zeeuw, Rob de Graaf (Erasmus Computing Grid, Rotterdam, The Netherlands), the national German MediGRID and Services@MediGRID, German D-Grid, funded by the German Bundesministerium für Forschung und Technology (#01 AK 803 A-H, # 01 IG 07015 G) for access to their grid resources. We also thank J.C. van Dongen for the art design of figure 1.

WEB-RESOURCES

Mach, Mach2Qtl and Mach2Dat are freely available at:
http://www.sph.umich.edu/csg/abecasis/MACH/download/

REFERENCES

1. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *Science* **322**, 881-8 (2008).
2. Psaty, B.M., O'Donnell, C.J., Gudnason, V. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009).
3. de Bakker, P.I., Ferreira, M.A., Jia, X. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
4. International HapMap, C., Frazer, K.A., Ballinger, D.G. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
5. Li, Y. & Abecasis, G.R. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*, 2290 (2006).
6. 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* **39**, 906-13 (2007).
7. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
8. de Zeeuw, L.V. Het bouwen van een 20 TeraFLOP virtuelle supercomputer. *NIOC proceedings*, 52–59 (2007).
9. Krefting, D. MediGRID – Towards a user friendly secured grid infrastructure. *Future Generation Computers Systems* **25**, 326-336 (2008).
10. Hofman, A., Breteler, M.M., van Duijn, C.M. *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
11. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* **7**, 403-22 (1991).

CHAPTER 3

GENETICS OF HUMAN STATURE

CHAPTER 3.1

A GWAS of northwestern europeans involves the CNP signaling pathway in the etiology of human height

Hum Mol Genet 18, 3516-24 (2009).

ABSTRACT

Northwestern Europeans are among the tallest of human populations. The increase in body height in these people appears to have reached a plateau, suggesting the ubiquitous presence of an optimal environment in which genetic factors may have exerted a particularly strong influence on human growth. Therefore, we performed a genome-wide association study (GWAS) of body height using 2.2 million markers in 10,074 individuals from three Dutch and one German population-based cohorts. Upon genotyping the 12 most significantly height-associated single nucleotide polymorphisms (SNPs) from this GWAS in 6912 additional individuals of Dutch and Swedish origin, a genetic variant (rs6717918) on chromosome 2q37.1 was found to be associated with height at a genome-wide significance level ($P_{combined}$ = $3.4x10^{-9}$). Notably, a second SNP (rs6718438) located ~450 bp away and in strong LD (r^2 = 0.77) with rs6717918 was previously found to be suggestive of a height association in 29,820 individuals of mainly northwestern European ancestry, and the over-expression of a nearby natriuretic peptide precursor type C (*NPPC)* gene, has been associated with overgrowth and skeletal anomalies. We also found a SNP (rs10472828) located on 5p14 near the natriuretic peptide receptor 3 (*NPR3*) gene, encoding a receptor of the NPPC ligand, to be associated with body height ($P_{combined}$ = $2.1x10^{-7}$). Taken together, these results suggest that variation in the C-type natriuretic peptide (CNP) signaling pathway, involving the *NPPC* and *NPR3* genes, plays an important role in determining human body height.

INTRODUCTION

Human body height has a heritability of at least 80%[1] but, in terms of its genetic complexity, it may still serve as a model for the architecture of human complex traits in general. Thus, recent genome-wide association studies (GWAS) have revealed that tens to hundreds of loci with small individual effects are likely to underlie the observed population variation in body height[1-9]. What is more, taken together, the 44 height loci identified in the five first GWAS[2,3,5,8,9] were found to explain only 5% of this variation, with the most strongly associated single variant accounting for not more than 0.3%[1]. The Genetic Investigation of Anthropometric Traits (GIANT) consortium is currently assembling a collection of at least 100,000 individuals from different GWAS worldwide, trying to achieve the power necessary to identify the hundreds or even thousands of genetic variants expected to influence human height, and to jointly explain ~15-20% of its population variation[1]. However, it must be kept in mind that the genetic basis of body height may show regional differences owing to, for example, genetic heterogeneity or variable patterns of gene-environment interaction[7,10]. Genetic associations with small effects therefore may have been obscured in instances where individuals of different origin were pooled in meta-analyses, or used for the confirmation of findings made in other populations[2,3,5,6,11]. Finally, different environments may result in different levels of heritability for complex traits, including height[12]. We therefore decided to perform a GWAS of human stature that focused upon northwestern European individuals. In these people, adult height has potentially stabilized at a biologically determined maximum, suggesting environmental conditions that are more homogenous than in populations with an ongoing secular trend in height, such as the southern Europeans[13,14]. Consequently, our study employed Dutch and German individuals in the stage one meta-analysis and Dutch and Swedish samples in the stage two meta-analysis. In addition to corroborating height associations from previous multi-regional studies, our data also revealed a region of height association that achieved genome-wide statistical significance for the first time.

RESULTS

Meta-analysis of height GWAS in the stage one data set (n=10,074)

Genome-wide SNP data from three Dutch and one northern German cohort, comprising a total of 10,074 individuals, were used to search for genetic variants associated with human height in stage one of the analysis (Table 1). These samples included individuals from the initial Rotterdam Study cohort (RS-I; n=5746), from an extension of the Rotterdam Study (RS-II; n=1891), the Erasmus Rucphen Family study (ERF; n=1473) and from the Kiel PopGen biobank study (PopGen-KIEL; n=964). In each study, we performed an association analysis of up to 2,543,888 SNPs, including genotypes that were imputed with reference to the International HapMap Project CEU panel release 22. Genotype imputation is an approach to overcome the missing data problem in the analysis of data from different genotyping platforms. A recent pan-European study[15], including a subset of the samples used here, revealed that the HapMap CEU samples are genetically closest to western and northern Europeans. Therefore, the CEU data can be assumed to represent an appropriate basis for genotype imputation for the present study. Following quality-control (QC), summary statistics from each study for the 2,228,850 remaining SNPs were subjected to the stage one meta-analysis. In agreement with previous recommendations[16,17], results were deemed significant at a genome-wide level if the locus-specific unadjusted P value was smaller than $5x10^{-8}$.

In the stage one data set, the strongest association with height was observed for rs3118905 ($P = 3.1x10^{-10}$), thereby confirming a previously reported association of a locus near the *DLEU7* gene[5]. In addition, the previously identified associations with height in or around *HMGA1*, *ZBTB38*, *CDK6*, *HGMA2*, *CABLES1, NPR3, GPR126* and *GDF5* [2,3,5,18] were also replicated with $P < 5x10^{-6}$ (Figure 1). In total, 29 of the 48 independent height loci totally known so far[2,3,5,6,8,9,18], signified by 38 of 57 SNPs in our study, attained nominal statistical significance ($P < 0.05$) in the stage one data set (Supplementary Table 1). Inspection of the Quantile-quantile (Q-Q) plot (Figure 2) of all SNPs included in the meta-analysis indicated an excess of significantly associated markers. Upon exclusion of all markers in LD with the most significant SNPs from the 48 known height loci[2,3,5,6,18], a deviation from the expected plot under the null hypothesis was still prevalent, suggesting the presence of additional, significantly associated loci in our data. After excluding all loci previously reported to be associated with human height, we ascertained 12 genetic variants with $P < 5x10^{-6}$ (Figure 2 and Supplementary Table 2) for further analysis.

Table 1. Characteristics of the study samples used for stage one and stage two meta-analyses

Study	Population	Gender	Average age in years (sd)	Average height in cm (sd)
RS-I	Dutch	Male (n= 2372)	68.13 (8.16)	174.85 (6.76)
		Female (n= 3374)	70.32 (9.60)	161.35 (6.57)
RS-II	Dutch	Male (n= 862)	64.68 (7.82)	175.49 (6.59)
		Female (n= 1029)	65.65 (8.86)	162.36 (6.27)
PopGen-KIEL	German	Male (n= 506)	51.27 (14.24)	180.31 (7.45)
		Female (n= 458)	50.79 (14.92)	167.16 (6.73)
ERF	Dutch	Male (n= 557)	51.12 (15.73)	174.00 (7.73)
		Female (n=916)	49.99 (15.82)	161.50 (7.01)
Stage One Data Set		**n=10,074**		
LASA	Dutch	Male (n=371)	72.25 (6.49)	173.35 (6.64)
		Female (n= 381)	72.56 (6.51)	160.44 (6.27)
EPOS	Dutch	Female (n= 1698)	50.01 (2.14)	164.67 (6.10)
NTR/NESDA	Dutch	Male (n= 1211)	46.09 (13.42)	181.59 (7.17)
		Female (n= 2311)	42.65 (13.25)	169.06 (6.37)
GOOD	Swedish	Male (n=940)	18.90 (0.56)	181.38 (6.74)
Stage Two Data Set		**n=6,912**		
Combined Stage One and Stage Two Data Set		**n=16,986**		

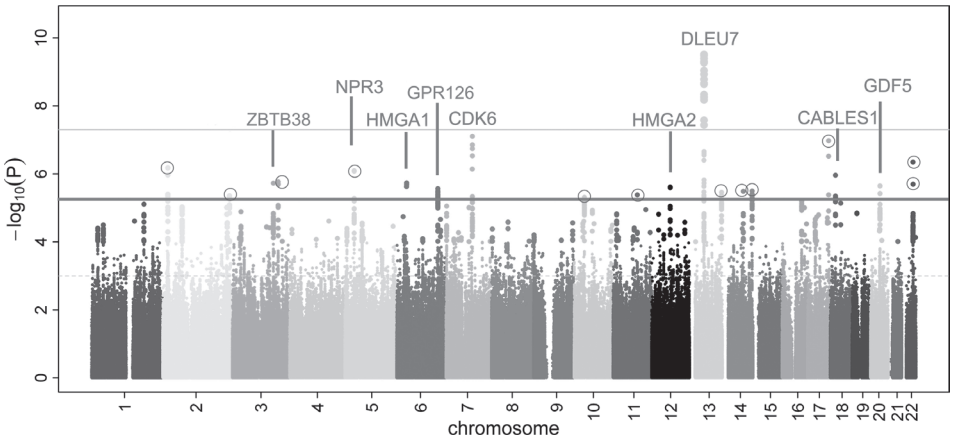


Figure 1. Manhattan plot of the height association test results (log_{10}(P)) for all SNPs in the stage one data set (n=10,074). Horizontal lines are the suggestive ($5x10^{-6}$) and genome-wide significance ($5x10^{-8}$) P value thresholds, respectively. Signals passing the suggestive threshold, and with the respective gene name, are previously known regions of height association. Twelve loci for which SNPs were selected for the stage two analysis are emphazised by a circle.

Combined meta-analysis of the 12 novel height-associated SNPs from stage one and stage two (n=16.986)

The 12 putative height-associated SNPs newly identified in the stage one data set were next scrutinized in 4,462 additional individuals of northern European ancestry (Table 1), namely 3,522 Dutch from the Netherlands Twin Register study and the Netherlands Study of Depression and Anxiety (NTR/NESDA) and 940 Swedes from the Gothenburg Osteoporosis and Obesity Determinants (GOOD) study. Furthermore, de-novo genotyping of the 12 SNPs was carried out in another 2,450 individuals from two additional Dutch population-based studies: 752 participants of the Longitudinal Aging Study Amsterdam study (LASA), and 1,698 participants of the European Prospective Osteoporosis Study (EPOS), thereby bringing the total number of Dutch, German and Swedish individuals in the combined stage one and stage two data set to 16,986.

The combined analysis of all samples identified SNP rs6717918 on chromosome *2q37.1* as being associated with body height at a genome-wide significant level ($P=3.4x10^{-9}$). In the stage one data set, several SNPs in this region showed suggestive association with height and were only in relatively weak LD ($0.2<r^2<0.5$) with rs6717918 (Figure 3). The direction of the association between rs6717918 and height was consistent across all sub-samples, without significant evidence for any inter-study heterogeneity (Q-statistic P=0.74) (Figure 4). The T-allele of rs6717918 was associated with an increase in height by 0.44 cm per allele copy (Table 2). At the genome-wide significance level, two more loci, namely rs139909 mapping to the trinucleotide repeat containing 6B (*TNRC6B*) gene on chromosome 22 (Supplementary Figure 1) and rs10472828 near the natriuretic peptide receptor 3 (*NPR3*) gene on chromosome 5 (Supplementary Figure 2), showed suggestive evidence for an association with height in the combined data set ($P< 5x10^{-7}$; Table 2).

Next, we aimed at replicating the putative association of one or the other of the 12 SNPs in recently published GWAS results[2,3,5,6,8,9,18]. We found additional evidence for an association with height only for the *2q37.1* region, using the Illumina data provided by Gudbjartsson et al.[2] for 25,174 Icelanders, 2876 Dutch, and 1770 European Americans. However, all nine SNPs typed in the *2q37.1* region in the original study failed to attain genome-wide significance there. The strongest evidence for an association was obtained for rs749052 ($P = 1.4x10^{-6}$). In a regional meta-analysis of their and our data SNP rs6718438, a proxy for rs6717918 in strong LD ($r^2 = 0.77$) and only 456 bp away from it, was found to be the most significantly height-associated SNP ($P = 8.4x10^{-12}$). Furthermore, all nine SNPs in the *2q27.1*

region attained genome-wide significance ($P < 4x10^{-8}$; Supplementary Table 3 and Supplementary Figure 3). To determine whether the association with rs749052 observed by Gudbjartsson *et al.*[2] was independent of rs6718438 or whether it was due to LD between the two SNPs (r^2=0.17), we conducted conditional association analyses of rs749052 controlling for the effect of rs6718438. A meta-analysis of the respective results for RS-I, RS-II and GOOD revealed that the height association of rs749052 remained nominally significant (P = 0.01). Taken together, the available data therefore provide conclusive evidence for an association between human body height and genetic variation at *2q37.1*.

DISCUSSION

Northwestern Europeans are among the tallest of human populations[13]. The average Dutch male, for example, is currently almost 20 cm taller than 150 years ago[13]. Over the last 50 years, however, average human body height has increased much less in northern than in southern Europe[13,14]. This suggests that height has approached a biologically determined maximum in the north, and that this leveling off has occurred against an optimal and comparatively homogeneous environmental background as regards growth-relevant factors[13,14]. With an aim to identify new genetic variants determining human body height, we therefore conducted a two-stage meta-analysis of genome-wide association studies (GWAS) of 16,986

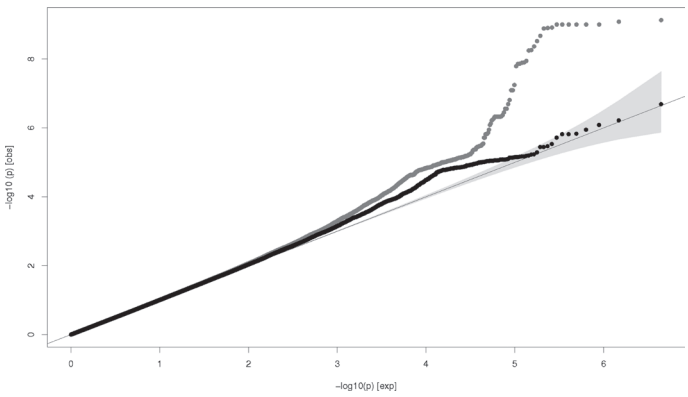


Figure 2. Quantile-quantile (Q-Q) plot of the height association test results (log_{10}(P)) for all SNPs passing quality-control (gray line) in the stage one analysis including 10,074 subjects, excluding variants in 48 independent loci previously associated with height on a genome-wide level (black line). Depicted P values were corrected for population stratification using the over inflation factor λ_{GC} = 1.049.

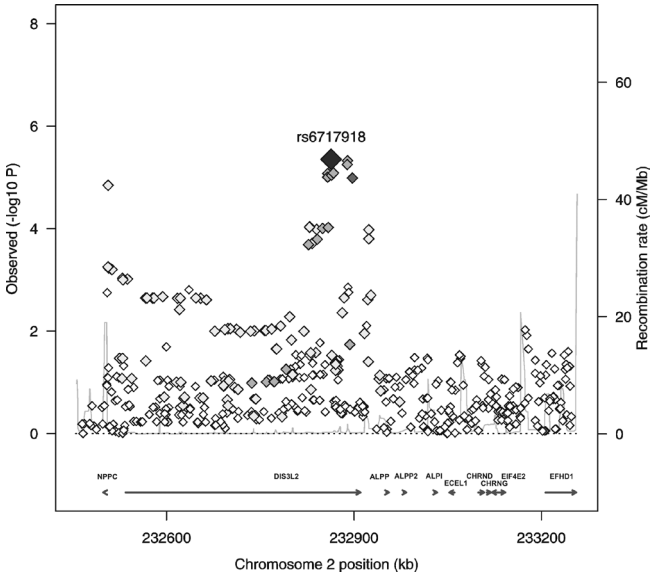


Figure 3. Local plot of the height association test results (log_{10}(P)) around SNP rs6717918. P values are as obtained from the meta-analysis of the stage one data set. The combined P value from the meta-analysis of stage one and stage two data sets equals $3.4x10^{-9}$ for rs6717918. Darker diamonds indicates higher LD with rs6717918. The vertical lines depicts local recombination rates.

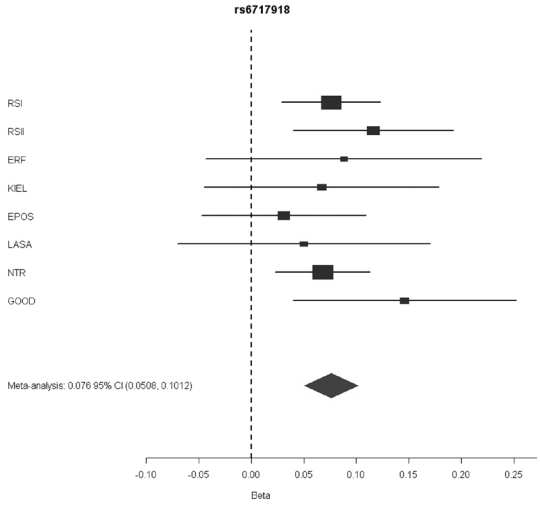


Figure 4. Forrest plot for the most significant SNP (rs6717918) in the combined meta-analysis of both the stage one and the stage two data sets. Beta: increase in height per SNP allele. Squares represent effect size estimates (measured in standard deviations of height) and 95% CI for each study. The diamond represents the summary effect size estimate.

Table 2. Most significant height associations in the stage one, stage two and combined stage one and stage two meta-analyses[1]

SNP	Chr	Pos (bp)[2]	Genes	A[3]	Freq[3]	Effect (SE)[4]	Stage One P-value n=10,074	Stage Two P-value n=6,912	Combined P-value n=16,986
rs6717918	2	232863344	*DISC3L2, ALPP, NPPC*	T	0.78	0.44 (0.12)	$4.4x\ 10^{-6}$	$1.4x10^{-4}$	$3.4x10^{-9}$
rs139909	22	39027527	*TNRC6B*	T	0.68	0.25 (0.11)	$4.5x10^{-7}$	$2.2x10^{-2}$	$1.7x10^{-7}$
rs10472828	5	32924575	*NPR3*	C	0.56	0.22 (0.09)	$8.1x10^{-7}$	$2.0x10^{-2}$	$3.4x10^{-7}$

[1]Only associations with an overall P < $5x10^{-7}$ are shown. [2]position relative to Build 36.2. [3]Effect allele and its frequency. [4]Per-allele Change in Height in cm from the stage two data of the study only

northwestern Europeans, comprising individuals of Dutch, German and Swedish origin. The first stage of this study not only confirmed the phenotype association of common genetic variants previously described as determinants of human stature in populations of various (mostly European) origins (Figure 2, Supplementary Table 1), but also pointed to additional putative height loci followed-up in the second stage analysis using independent samples. Since GWAS are only suited to find phenotype associations with common genetic variants, however, we cannot exclude that rare alleles with large effects may also have contributed to the height variation observed in our samples. Similarly, environmental effects (e.g., diet) and gene-environment interaction effects may have contributed as well, but such considerations fell outside the scope of the current project.

In the combined stage one and stage two meta-analysis, we found for the first time that a locus at *2q37.1* is associated with human height at the genome-wide significance level. The strongest association was observed with intronic SNP rs6717918 in the DIS3 mitotic control homolog (S. cerevisiae)-like 2 (*DIS3L2*) gene which, until very recently, had been considered a hypothetical protein-coding sequence only (MGC42174). Since there is no LD-based evidence for extensive recombination in the surrounding 500 kb region, however, it would still appear plausible that other genes in the vicinity of *DIS3L2* may have contributed to the observed height association (Figure 3), most notably the natriuretic peptide precursor type C (*NPPC*) gene ~350 kb away from rs6717918 and only 6 kb away from another height-associated SNP, rs749052. In fact, we regard variation in the *NPPC* gene as the most likely cause of the height association observed with *2q37.1*, represented by both rs6717918 and rs749052. The *NPPC* gene encodes the C-type natriuretic peptide (CNP), a

molecule that regulates endochondral ossification of the cartilaginous growth plate and, hence, influences longitudinal bone growth[19,20]. Recently, a balanced t(2;7) translocation has been reported in a patient with unusually high stature (>97th percentile), Marfanoid habitus, and skeletal anomalies[19]. The respective breakpoint on chromosome 2 was located halfway between rs6717918 and the *NPPC* gene, and was shown to induce over-expression of CNP and consequent skeletal overgrowth. Furthermore, transgenic mice with CNP over-expression in osteoblasts exhibit a phenotype similar to the skeletal abnormalities of that patient[19]. It has also been demonstrated that over-expression of CNP in chondrocytes can counteract dwarfism in a mouse model of achondroplasia[21]. Taken together, these findings suggests that the association between human height and variation at *2q37.1*, as observed in our study, reflects differential regulation of the *NPPC* gene expression with an impact on bone growth regulation and consequent body height.

Although not statistically significant at a genome-wide level in our study, the pronounced height association observed with rs10472828 on chromosome *5p14* deserves further attention. This is because rs10472828 is located only 100 kb upstream of *NPR3*, a gene that encodes a receptor of the CNP ligand. Indeed, Soranzo *et al.*[18] recently found that this SNP is significantly associated with human height ($P = 3.0x10^{-7}$) in a collection of British and Dutch individuals (which included the subset of the RS-I participants). When combining all our data with those provided by Soranzo *et al.*[18] in their Supplementary Table 2, while excluding samples from the Rotterdam study (remaining n=14,052), the combined P value of $3.5x10^{-11}$ attains genome-wide significance. Furthermore, two SNPs near *NPR3* (rs3811958 and rs13154066) showed a suggestive height association ($P < 5x10^{-6}$) in the study by Gudbjartsson *et al*[2], but these polymorphisms were not in strong LD with rs10472828. The *NPR3* gene encodes one of three CNP receptors (NPR-C), and knock-out of NPR-C was found to result in significant skeletal overgrowth in mice[22]. It has also been suggested that NPR-C may act as a clearance receptor modulating the effect of CNP[22,23]. Recently, CNP-induced differentiation of osteoblasts was found to switch from NPR-B to NPR-C with aging in rat cells[24], thus implying an important role of NPR-C in the late stages of bone formation. The identification of SNPs near both the *NPPC* and the *NPR3* gene as being strongly associated with human height clearly points to a prominent role of the CNP signaling pathway in the etiology of body height variation, at least in northwestern Europeans.

Meta-analyses of GWAS are not without limitations. False positive associations due to multiple hypothesis testing or population stratification are inherent possibilities. Here, we minimized the impact of multiple testing by adopting a stringent genome-

wide significance level. To alleviate the possible effects of population stratification, we adjusted all relevant test statistics by the inflation factor λ_{GC}[17] and by principal components derived from the multidimensional scaling analysis of identity-by-state distances between individuals[25]. Furthermore, all studies included in our meta-analyses were confined to individuals of northwestern European descent and, consequently, the overall inflation factor of the stage one data set (λ_{GC}= 1.049) was low for a study of this size[26]. This notwithstanding, the ERF samples exhibited a relative high inflation factor (λ_{GC} = 1.950) most likely due to intricate family relationships. We therefore conducted a sensitivity analysis, excluding the ERF samples, which revealed that the association between rs6717918 and height remained significant at the genome-wide level (P = 7.8×10^{-9}). Taken together, multiple testing, population stratification and cryptic relatedness are therefore unlikely to have confounded our association findings.

In conclusion, we have unequivocally identified variation at *2q37.1* as being associated with human body height in northwestern Europeans. The fact that this locus has not been found in previous studies may either be due to chance (i.e., sampling variation, power differences etc.) or may be explicable in terms of a higher level of genetic and environmental heterogeneity in the other samples, compared to ours. Thus, sufficiently powered studies of additional, geographically confined populations are needed to clarify whether the observed height association of the *2q37.1* region we observed represents a region-specific effect or not. We further propose that the observed association is due to variation in the *NPPC* gene (encoding the CNP ligand), the most plausible functional candidate in the *2q37.1* region. The strong association observed with SNPs in the vicinity of the *NPR3* gene on chromosome *5p14* (encoding the CNP receptor) lends additional support to the view that common variants in the CNP signaling pathway play a prominent role in the regulation of normal height variation in humans.

METHODS

Subjects

All studies were approved by the institutional ethics review committees of the respective organizations and all participants provided written informed consent. The initial Rotterdam Study (**RS-I**) is a prospective population-based cohort study of chronic disabling conditions in Dutch individuals aged 55 years or above (http://www.epib.nl/ergo.htm)[27,28]. The Rotterdam Study II (**RS-II**) is an extension of

the initial Rotterdam Study, which started in 1999 and used the same inclusion criteria and design as the original cohort. In short, 3011 individuals (response rate 67%) who had turned 55 years of age or had moved into the study district of Ommoord, Rotterdam, since the start of the original study in 1990 were included in the extension cohort. The Erasmus Rucphen Family (**ERF**) study is a family-based study of a genetic isolate in the southwestern Netherlands to identify genetic risk factors for complex disorders[29]. The Kiel PopGen biobank (**PopGen-KIEL)** study is a centralized platform for the recruitment and follow-up of probands for genetic epidemiological studies in Schleswig-Holstein, the most northern part of Germany. Since its establishment in 2003, PopGen-Kiel has assembled a collection of biomaterials, phenotypic and genotypic data from more than 60,000 individuals. This includes nearly 1000 controls for which height and genome-wide SNP genotype data were available for inclusion into the present study. The European Prospective Osteoporosis Study (**EPOS)**[30] is a cross-sectional study of 5896 women born between 1941 and 1947, and currently living in the city of Eindhoven, the Netherlands. DNA was available for 1798 of these women. The Longitudinal Aging Study Amsterdam (**LASA**) study[31] is a population-based cohort study, including 919 individuals for whom DNA was available. **NTR/NESDA**: The two parent projects that supplied data are the Netherlands Study of Depression and Anxiety (NESDA)[32] and the Netherlands Twin Registry (NTR)[33]. NESDA and NTR studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam (IRB number IRB-2991 under Federal wide Assurance-3703; IRB/ institute codes, NESDA 03-183; NTR 03-180). The sample consisted of 1777 NTR and 1763 NESDA participants. For NTR participants, longitudinal (1991-2004) survey and data on height were combined. Only one subject per family was selected. For NESDA participants, height was assessed during a visit to the clinic. The Gothenburg Osteoporosis and Obesity Determinants (**GOOD**) study was initiated to determine both environmental and genetic factors involved in the regulation of bone and fat mass. Male study subjects were randomly identified in the greater Gothenburg area in Sweden using national population registers, contacted by telephone, and invited to participate[34]. To be enrolled in the GOOD study, subjects had to be between 18 and 20 years of age. There were no other exclusion criteria, and 49% of the study candidates agreed to participate.

Genotyping and quality control

The six GWAS were carried out using either the Illumina Infinium HumanHap550 Beadchip (RS-I and RS-II), the Illumina Infinium HumanHap610 (GOOD), the Illumina

Infinium HumanHap300 (ERF), the Perlegen 600K (NTR/NESDA), or the Affymetrix Dual NspI/StyI GeneChip 2x250K (PopGen-KIEL). *De novo* genotyping of 12 SNPs in the LASA and EPOS samples was performed using Taqman allelic discrimination (Applied Biosystems Inc., Foster City, CA, USA) according to the manufacturer's protocols and quality control standards (assay numbers and primer designs can be found in the Supplementary Materials). The following sample quality control criteria were applied in the GWAS of RS-I, RS-II, PopGen-KIEL, ERF and GOOD: sample call rate ≥ 97.5%, gender mismatch with typed X-linked markers, evidence for DNA contamination in the samples using the mean of the autosomal heterozygosity > 0.33, exclusion of duplicates or 1st degree relatives identified using IBS probabilities, exclusion of outliers (3 SD away from the population mean) using Multi-Dimensional Scaling (MDS) analysis with four principal components, and exclusion of samples with missing height measurements. Complete information on genotyping protocols and quality control measures for NTR/NESDA cohorts have been described elsewhere[35]. The exclusion/filtering criteria for SNPs are described in Supplementary Table 4.

Genotype imputation

Genotype imputation was used to evaluate the height association of one and the same SNP across samples typed on different genotyping platforms. Genotypes were imputed for all polymorphic SNPs (minor allele frequency > 0.01) using either the MACH[36] or the IMPUTE[37] software, based upon phased autosomal chromosomes of the HapMap CEU Phase II panel (release 22, build 36), orientated on the positive strand. Imputation quality control metrics from MACH and IMPUTE were used for filtering out SNPs with low-quality data. Detailed descriptions of the quality control and imputation procedures are provided in Supplementary Table 4.

GWAS

In each GWAS, the association between a SNP and height was assessed using sex-specific, age-standardized residuals that were analyzed under an additive (per allele) genetic model. To adjust for population substructure, we included as covariates in the regression analysis of RS-I, RS-II, and PopGen-KIEL the four most important principal components (PC), derived from a Multi-Dimensional Scaling analysis of IBS distances using the PLINK[38] software. In the analysis of imputed genotypes, uncertainty in genotype prediction was accounted for by utilizing either the dosage information from MACH[36] or the genotype probabilities from IMPUTE[37]. We carried

out association testing for imputed SNPs using a linear regression framework as implemented in MACH2QTL[36], SNPTEST[37], and ProbABEL[39] (Supplementary Table 4). A linear regression analysis (1df) as implemented in PLINK[38] was performed for the genotype data in EPOS and LASA.

Meta-analysis

The genomic control method[17], as implemented in METAL, was used to correct for any residual population stratification or relatedness not accounted for by the four most important PC. The estimated inflation factors were 1.089, 1.006, 1.000, 1.950, 1.086 and 1.030 for RS-I, RS-II, PopGen-KIEL, ERF, NTR-NESDA, GOOD, respectively. SNPs with a minor allele frequency < 0.05, a MACH observed/expected allele dosage variance < 0.05, or a SNPTEST proper_info < 0.4 were excluded from the meta-analysis. A detailed description of each study is provided in Supplementary Table 4. We obtained the combined results of 2,228,850 SNPs, pooling effect sizes by means of a fixed effects inverse variance meta-analysis as implemented in METAL. Estimated heterogeneity variance and forest plots were generated using the Rmeta R package. Regional association plots of the meta-analysis results were obtained with SNAP[40].

AKNOWLEDGEMENTS

The generation and management of GWAS genotype data for the Rotterdam Study **(RS-I, RS-II)** was supported by a Netherlands Organization for Scientific Research (NWO) Investments grant (nr. 175.010.2005.011, 911-03-012). This study was also supported by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI) / NWO project no. 050-060-810, the Netherlands Forensic Institute, and by a grant from the NGI / NWO within the framework of the Forensic Genomics Consortium Netherlands. The Rotterdam Study is funded by the Erasmus University Medical Center, the Erasmus University Rotterdam, the Netherlands Organization for the Health Research and Development (ZonMw), the RIDE, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports of the Netherlands, the European Commission (DG XII), and the Municipality of Rotterdam. The **ERF** study was supported by grants from the NWO, Erasmus MC and the Centre for Medical Systems Biology (CMSB). The Kiel PopGen biobank study (**PopGen-KIEL)** is directly funded by the Ministry of Science, Economy and Transport of Schleswig-Holstein, Germany. The **GOOD** study is funded by: The Swedish Research Council, The ALF grant at the Sahlgrenska University Hospital, Gothenburg, Sweden, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, and the Novo Nordisk Foundation. For **NTR/NESDA**, we acknowledge financial support from the NWO: twin-family database for behavior genetics and genomic studies (480-04-004), genetic basis of anxiety and depression (904-61-090); resolving cause and effect in the association between exercise and well-being (904-61-193); CMSB (NWO Genomics); Spinozapremie (SPI 56-464-14192); Centre for Neurogenomics and Cognitive Research (CNCR-VU); genome-wide analyses of European twin and population cohorts (EU/QLRT-2001-01254); Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ

Friesland, GGZ Drenthe). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and the analysis was supported by grants from GAIN and the NIMH (MH081802). We thank Pascal Arp, Mila P. Jhamai, Dr Michael J. Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the RS-I and RS-II GWAS database. We are grateful to all general practitioners for their support, to Petra Veraart for her help in verifying genealogy information of ERF, Jeannette Vergeer for the supervision of the ERF laboratory work, and Peter Snijders for his help with the ERF data collection. The authors are grateful to all study participants, staff and the participating general practitioners and pharmacists.

Note: Supplementary information is available on the Human Molecular Genetics website:

http://hmg.oxfordjournals.org/content/suppl/2009/07/01/ddp296.DC1/ddp296supp.pdf

REFERENCES

1. Weedon, M.N. and Frayling, T.M. (2008) Reaching new heights: insights into the genetics of human stature. *Trends Genet.*, 24, 595-603.
2. Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S. *et al.* (2008) Many sequence variants affecting diversity of adult human height. *Nat. Genet.*, 40, 609-615.
3. Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guiducci, C. *et al.* (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.*, 40, 584-591.
4. Visscher, P.M. (2008) Sizing up human height variation. *Nat. Genet.*, 40, 489-490.
5. Weedon, M.N., Lango, H., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., Freathy, R.M., Perry, J.R., Stevens, S., Hall, A.S. *et al.* (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.*, 40, 575-583.
6. Johansson, A., Marroni, F., Hayward, C., Franklin, C.S., Kirichenko, A.V., Jonasson, I., Hicks, A.A., Vitart, V., Isaacs, A., Axenovich, T. *et al.* (2009) Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.*, 18, 373-380.
7. Lei, S.F., Yang, T.L., Tan, L.J., Chen, X.D., Guo, Y., Guo, Y.F., Zhang, L., Liu, X.G., Yan, H., Pan, F. *et al.* (2009) Genome-wide association scan for stature in Chinese: evidence for ethnic specific loci. *Hum. Genet.*, 125, 1-9.
8. Sanna, S., Jackson, A.U., Nagaraja, R., Willer, C.J., Chen, W.M., Bonnycastle, L.L., Shen, H., Timpson, N., Lettre, G., Usala, G. *et al.* (2008) Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat. Genet.*, 40, 198-203.
9. Weedon, M.N., Lettre, G., Freathy, R.M., Lindgren, C.M., Voight, B.F., Perry, J.R., Elliott, K.S., Hackett, R., Guiducci, C., Shields, B. *et al.* (2007) A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat. Genet.*, 39, 1245-1250.
10. Sammalisto, S., Hiekkalinna, T., Schwander, K., Kardia, S., Weder, A.B., Rodriguez, B.L., Doria, A., Kelly, J.A., Bruner, G.R., Harley, J.B. *et al.* (2009) Genome-wide linkage screen for stature and body mass index in 3.032 families: evidence for sex- and population-specific genetic effects. *Eur. J. Hum. Genet.*, 17, 258-266.
11. Lei, S.F., Tan, L.J., Liu, X.G., Wang, L., Yan, H., Guo, Y.F., Liu, Y.Z., Xiong, D.H., Li, J., Yang, T.L. *et al.* (2008) Genome-wide association study identifies two novel loci containing FLNB and SBF2 genes underlying stature variation. *Hum. Mol. Genet.*, 10.1093/hmg/

ddn1405.

12. Silventoinen, K., Kaprio, J., Lahelma, E. and Koskenvuo, M. (2000) Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *Am. J. Public Health.*, 90, 627-630.
13. Cole, T.J. (2003) The secular trend in human physical growth: a biological view. *Econ. Hum. Biol.*, 1, 161-168.
14. Garcia, J. and Quintana-Domeque, C. (2007) The evolution of adult height in Europe: a brief note. *Econ. Hum. Biol.*, 5, 340-349.
15. Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D. *et al.* (2008) Correlation between genetic and geographic structure in Europe. *Curr. Biol.*, 18, 1241-1248.
16. Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.*, 32, 381-385.
17. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, 55, 997-1004.
18. Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A. *et al.* (2009) Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet.*, 5, e1000445.
19. Bocciardi, R., Giorda, R., Buttgereit, J., Gimelli, S., Divizia, M.T., Beri, S., Garofalo, S., Tavella, S., Lerone, M., Zuffardi, O. *et al.* (2007) Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. *Hum. Mutat.*, 28, 724-731.
20. Gilbert, S. (2003) *Paraxial and intermediate mesoderm.* Sinauer Associates, Sunderland, MA.
21. Yasoda, A., Komatsu, Y., Chusho, H., Miyazawa, T., Ozasa, A., Miura, M., Kurihara, T., Rogi, T., Tanaka, S., Suda, M. *et al.* (2004) Overexpression of CNP in chondrocytes rescues achondroplasia through a MAPK-dependent pathway. *Nat. Med.*, 10, 80-86.
22. Matsukawa, N., Grzesik, W.J., Takahashi, N., Pandey, K.N., Pang, S., Yamauchi, M. and Smithies, O. (1999) The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc. Natl. Acad. Sci. U S A*, 96, 7403-7408.
23. Pejchalova, K., Krejci, P. and Wilcox, W.R. (2007) C-natriuretic peptide: an important regulator of cartilage. *Mol. Genet. Metab.*, 92, 210-215.
24. Kaneki, H., Kurokawa, M. and Ide, H. (2008) The receptor attributable to C-type natriuretic peptide-induced differentiation of osteoblasts is switched from type B- to type C-natriuretic peptide receptor with aging. *J. Cell. Biochem.*, 103, 753-764.
25. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, 38, 904-909.
26. de Bakker, P.I., Ferreira, M.A., 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.
27. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.*, 22, 819-829.
28. Hofman, A., Grobbee, D.E., de Jong, P.T. and van den Ouweland, F.A. (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study.

Eur. J. Epidemiol., 7, 403-422.

29. Aulchenko, Y.S., Heutink, P., Mackay, I., Bertoli-Avella, A.M., Pullen, J., Vaessen, N., Rademaker, T.A., Sandkuijl, L.A., Cardon, L., Oostra, B. *et al.* (2004) Linkage disequilibrium in young genetically isolated Dutch population. *Eur. J. Hum. Genet.*, 12, 527-534.

30. Smeets-Goevaers, C.G., Lesusink, G.L., Papapoulos, S.E., Maartens, L.W., Keyzer, J.J., Weerdenburg, J.P., Beijers, L.M., Zwinderman, A.H., Knottnerus, J.A., Pols, H.A. *et al.* (1998) The prevalence of low bone mineral density in Dutch perimenopausal women: the Eindhoven perimenopausal osteoporosis study. *Osteoporos. Int.*, 8, 404-409.

31. Knipscheer, C.P., Dykstra, P.A., van Tilburg, T.G. and de Jong-Gierveld, J. (1998) [Living arrangements and social networks of elders. A selection of findings from a NESTOR-Study] Leefvormen en sociale netwerken van ouderen. Een selectie van bevindingen uit een NESTOR-Studie. *Tijdschr. Gerontol. Geriatr.*, 29, 110-119.

32. Penninx, B.W., Beekman, A.T., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W., Assendelft, W.J. *et al.* (2008) The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.*, 17, 121-140.

33. Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M. *et al.* (2006) Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.*, 9, 849-857.

34. Lorentzon, M., Swanson, C., Andersson, N., Mellstrom, D. and Ohlsson, C. (2005) 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.*, 20, 1334-1341.

35. Sullivan, P.F., de Geus, E.J., Willemsen, G., James, M.R., Smit, J.H., Zandbelt, T., Arolt, V., Baune, B.T., Blackwood, D., Cichon, S. *et al.* (2008) Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol. Psychiatry*, 10.1038/mp.2008.1125.

36. Li, Y. and Abecasis, G.R. (2006) Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.*, S79, 2290.

37. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, 39, 906-913.

38. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., 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.

39. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)*, 23, 1294-1296.

40. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics (Oxford, England)*, 24, 2938-2939.

CHAPTER 3.2

Hundreds of variants clustered in genomic loci and biological pathways affect human height

Nature 467, 832-8 (2010)

ABSTRACT

Most common human traits and diseases have a polygenic pattern of inheritance: DNA sequence variants at many genetic loci influence phenotype. Genome-wide association (GWA) studies have identified >600 variants associated with human traits1, but these typically explain small fractions of phenotypic variation, raising questions about the utility of further studies. Here, using 183,727 individuals, we show that hundreds of genetic variants, in at least 180 loci, influence adult height, a highly heritable and classic polygenic trait2,3. The large number of loci reveals patterns with important implications for genetic studies of common human diseases and traits. First, the 180 loci are not random, but instead are enriched for genes that are connected in biological pathways ($P=0.016$), and that underlie skeletal growth defects ($P<0.001$). Second, the likely causal gene is often located near the most strongly associated variant: in 13 of 21 loci containing a known skeletal growth gene, that gene was closest to the associated variant. Third, at least 19 loci have multiple independently associated variants, suggesting that allelic heterogeneity is a frequent feature of polygenic traits, that comprehensive explorations of already-discovered loci should discover additional variants, and that an appreciable fraction of associated loci may have been identified. Fourth, associated variants are enriched for likely functional effects on genes, being over-represented amongst variants that alter amino acid structure of proteins and expression levels of nearby genes. Our data explain ~10% of the phenotypic variation in height, and we estimate that unidentified common variants of similar effect sizes would increase this figure to ~16% of phenotypic variation (~20% of heritable variation). Although additional approaches are needed to fully dissect the genetic architecture of polygenic human traits, our findings indicate that GWA studies can identify large numbers of loci that implicate biologically relevant genes.

RESULTS

In Stage 1 of our study, we performed a meta-analysis of GWA data from 46 studies, comprising 133,653 individuals of recent European ancestry, to identify common genetic variation associated with adult height. To enable meta-analysis of studies across different genotyping platforms, we performed imputation of 2,834,208 single nucleotide polymorphisms (SNPs) present in the HapMap Phase 2 European-American reference panel[4]. After applying quality control filters, each individual study tested the association of adult height with each SNP using an additive model (**Supplementary Methods**). The individual study statistics were corrected using the genomic control (GC) method[5,6] and then combined in a fixed effects based meta-analysis. We then applied a second GC correction on the meta-analysis statistics, although this approach may be overly conservative when there are many real signals of association (**Supplementary Methods**). We detected 207 loci (defined as 1Mb on either side of the most strongly associated SNP) as potentially associated with adult height ($P<5x10^{-6}$).

To identify loci robustly associated with adult height, we took forward at least one SNP (**Supplementary Methods**) from each of the 207 loci reaching $P<5x10^{-6}$ into an additional 50,074 samples (Stage 2) that became available after completion of our initial meta-analysis. In the joint analysis of our Stage 1 and Stage 2 studies, SNPs representing 180 loci reached genome-wide significance ($P<5x10^{-8}$; **Supplementary Figures 1 and 2, Supplementary Table 1**). Additional tests, including genotyping of a randomly-selected subset of 33 SNPs in an independent sample of individuals from the 5^{th}-10^{th} and 90^{th}-95^{th} percentiles of the height distribution (N=3,190)[7], provided further validation of our results, with all but two SNPs showing consistent direction of effect (sign test $P<7x10^{-8}$) (**Supplementary Methods, Supplementary Table 2**).

Genome wide association studies can be susceptible to false positive associations from population stratification[7]. We therefore performed a family-based analysis, which is immune to population stratification in 7,336 individuals from two cohorts with pedigree information. Alleles representing 150 of the 180 genome-wide significant loci were associated in the expected direction (sign test $P<6x10^{-20}$; **Supplementary Table 3**). The estimated effects on height were essentially identical in the overall meta-analysis and the family-based sample. Together with several other lines of evidence (**Supplementary Methods**), this indicates that stratification is not substantially inflating the test statistics in our meta-analysis.

Common genetic variants have typically explained only a small proportion of the heritable component of phenotypic variation[8]. This is particularly true for height,

where >80% of the variation within a given population is estimated to be attributable to additive genetic factors[9], but over 40 previously published variants explain <5% of the variance[10-17]. One possible explanation is that many common variants of small effects contribute to phenotypic variation, and current GWA studies remain underpowered to detect the majority of common variants. Using five studies not included in Stage 1, we found that the 180 associated SNPs explained on average 10.5% (range 7.9-11.2%) of the variance in adult height (**Supplementary Methods**). Including SNPs associated with height at lower significance levels ($0.05 > P > 5x10^{-8}$) increased the variance explained to 13.3% (range 9.7-16.8%) (**Figure 1a**)[18]. In addition, we found no evidence that non-additive effects including gene-gene interaction would increase the proportion of the phenotypic variance explained (**Supplementary Methods, Supplementary Tables 5 and 6**).

As a separate approach, we used a recently developed method[19] to estimate the total number of independent height-associated variants with effect sizes similar to the ones identified. We obtained this estimate using the distribution of effect sizes observed in Stage 2 and the power to detect an association in Stage 1, given these effect sizes (**Supplementary Methods**). The cumulative distribution of height loci, including those we identified and others as yet undetected, is shown in **Figure 1b**. We estimate that there are 697 loci (95% confidence interval (CI): 483-1040) with effects equal or greater than those identified, which together would explain approximately 15.7% of the phenotypic variation in height or 19.6% (95% CI: 16.2-25.6) of height heritability (**Supplementary Table 4**). We estimated that a sample size of 500,000 would detect 99.6% of these loci at $P<5x10^{-8}$. This figure does not account for variants that have effect sizes smaller than those observed in the current study and, therefore, underestimates the contribution of undiscovered common loci to phenotypic variation.

A further possible source of missing heritability is allelic heterogeneity – the presence of multiple, independent variants influencing a trait at the same locus. We performed genome-wide conditional analyses in a subset of Stage 1 studies, including a total of 106,336 individuals. Each study repeated the primary GWA analysis but additionally adjusted for SNPs representing the 180 loci associated at $P<5x10^{-6}$ (**Supplementary Methods**). We then meta-analysed these studies in the same way as for the primary GWA study meta-analysis. Nineteen SNPs within the 180 loci were associated with height at $P<3.3x10^{-7}$ (a Bonferroni-corrected significance threshold calculated from the ~15% of the genome covered by the conditioned 2Mb loci; **Supplementary Methods**, **Table 1**, **Figure 2, Supplementary Figure 3**). The distances of the second signals to the lead SNPs suggested that both are likely

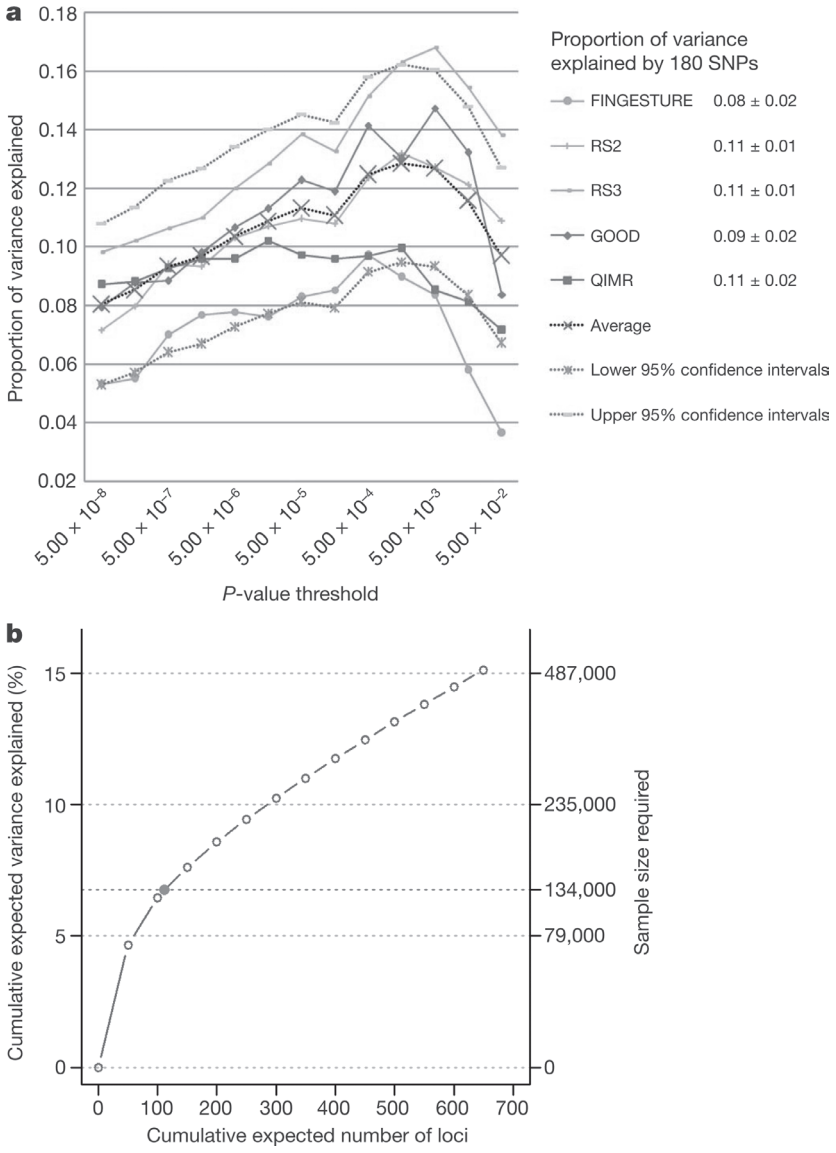


Figure 1. Phenotypic variance explained by common variants. (a) Variance explained is higher when SNPs not reaching genome-wide significance are included in the prediction model. The y-axis represents the proportion of variance explained at different P-value thresholds from Stage 1. Results are given for five studies that were not part of Stage 1. The proportion of variation explained by the 180 SNPs is shown in the column to the right of the graph. (b) Cumulative number of susceptibility loci expected to be discovered, including already identified loci and as yet undetected loci. Projections are based on loci that achieved a significance level of $P < 5 \times 10^{-8}$ in the initial scan and the distribution of their effect sizes in Stage 2. The dotted line with the filled bullet corresponds to expected phenotypic variance explained by the 110 loci that reached genome-wide significance in Stage 1, were replicated in Stage 2 and had at least 1% power.

to be affecting the same gene, rather than being coincidentally in close proximity. At 17 of 17 loci (excluding two contiguous loci in the *HMGA1* region), the second signal occurred within 500kb, rather than between 500kb and 1 Mb, of this lead SNP (binomial test $P=2x10^{-5}$). Further analyses of allelic heterogeneity may identify additional variants that increase the proportion of variance explained. For example, within the 180 2Mb loci, a total of 45 independent SNPs reached $P<1x10^{-5}$ when we would expect <2 by chance.

While GWA studies have identified many variants robustly associated with common human diseases and traits, the biological significance of these variants, and the genes on which they act, is often unclear. We first tested the overlap between the 180 height-associated variants and two types of putatively functional variants, nonsynonymous (ns) SNPs and cis-eQTLs (variants strongly associated with expression of nearby genes). Height variants were 2.4-fold more likely to overlap

Table 1: Secondary signals at associated loci after conditional analysis..

Second signal SNP	Index SNP	Chr	Second signal SNP position	Dist.[a] (bp)	r^2 [(b)]	Second signal *P*-value after cond.	Second signal *P*-value pre-cond.	Gene[c]
rs2280470	rs16942341	15	87196630	6721	0.009	$1x10^{-14}$	$1x10^{-15}$	***ACAN***
rs10859563	rs11107116	12	92644470	141835	0.003	$3x10^{-12}$	$8x10^{-10}$	*SOCS2*
rs750460	rs5742915	15	72028559	95127	0.004	$4x10^{-12}$	$7x10^{-08}$	*PML*
rs6938239	rs2780226[d]	6	34791613	484583	0.019	$6x10^{-12}$	$9x10^{-14}$	*HMGA1*
rs7652177	rs572169	3	173451771	196650	0.006	$7x10^{-11}$	$1x10^{-11}$	***GHSR***
rs7916441	rs2145998	10	80595583	196119	0.112	$6x10^{-10}$	$3x10^{-07}$	*PPIF*
rs3792752	rs1173727	5	32804391	61887	0.020	$7x10^{-10}$	$4x10^{-08}$	*NPR3*
rs10958476	rs7460090	8	57258362	98355	0.020	$1x10^{-09}$	$5x10^{-13}$	*SDR16C5*
rs2353398	rs7689420	4	145742208	45594	0.022	$2x10^{-09}$	$1x10^{-10}$	*HHIP*
rs2724475	rs6449353	4	17555530	87056	0.098	$2x10^{-09}$	$8x10^{-16}$	*LCORL*
rs2070776	rs2665838	17	59361230	41033	0.150	$9x10^{-09}$	$1x10^{-14}$	*GH region*
rs1401796	rs227724	17	52194758	60942	0.005	$2x10^{-08}$	$7x10^{-07}$	***NOG***
rs4711336	rs2780226[d]	6	33767024	540046	0.111	$3x10^{-08}$	$5x10^{-08}$	*HMGA1*
rs6892884	rs12153391	5	170948228	187815	0.000	$4x10^{-08}$	$2x10^{-05}$	*FBXW11*
rs1367226	rs3791675	2	55943044	21769	0.204	$4x10^{-08}$	0.1245	*EFEMP1*
rs2421992	rs17346452	1	170507874	187964	0.019	$5x10^{-08}$	$1x10^{-05}$	*DNM3*
rs225694	rs7763064	6	142568835	270147	0.001	$1x10^{-07}$	$2x10^{-06}$	*GPR126*
rs10187066	rs12470505	2	219223003	393610	0.022	$2x10^{-07}$	$5x10^{-08}$	***IHH***
rs879882	rs2256183	6	31247431	241077	0.016	$2x10^{-07}$	$8x10^{-08}$	*MICA*

[a]Distance of conditioned SNP from index SNP. [b]Linkage disequilibrium measured as r^2 in the HapMap CEU phase II release 22. [c]Nearest gene unless there is a known skeletal growth disorder gene in the locus (in bold). Positions are based on NCBI build 36. [d]Nearest conditioned SNP where second signal occurs within 1Mb of two conditioned SNPs.

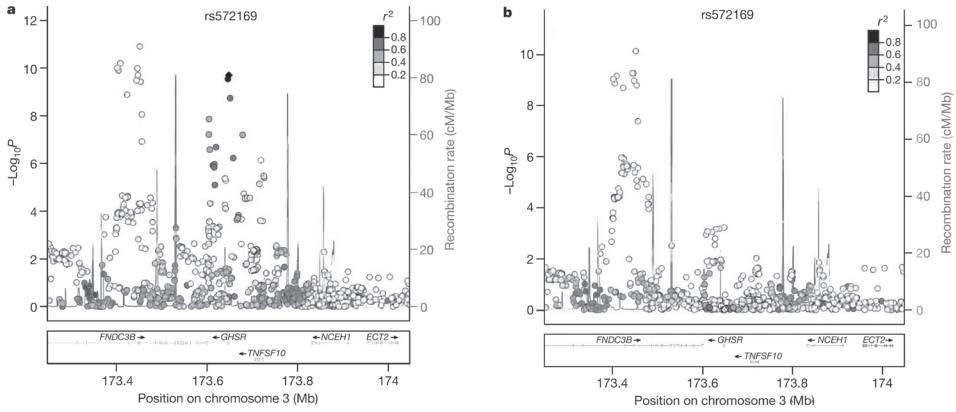


Figure 2. Example of a locus with a secondary signal before (a) and after (b) conditioning. The plot is centered on the conditioned SNP (black diamond) at the locus. r^2 is based on the CEU HapMap II samples. The vertical lines and right hand Y axis represent CEU HapMap II recombination rates. Created using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

with cis-eQTLs in lymphocytes than expected by chance (47 variants: P=4.7x10^{-11}) (**Supplementary Table 7**) and 1.7-fold more likely to be closely correlated (r^2>0.8 in HapMap CEU) with nsSNPs (24 variants P=0.004) (**Supplementary Methods**, **Supplementary Table 8**). Although the presence of a correlated eQTL or nsSNP at an individual locus does not establish the causality of any particular variant, this enrichment shows that common functional variants contribute to the causal variants at height-associated loci. We also noted five loci where the height associated variant was strongly correlated (r^2>0.8) with variants associated with other traits and diseases (P<5x10^{-8}), including bone mineral density, rheumatoid arthritis, type 1 diabetes, psoriasis and obesity, suggesting that these variants have pleiotropic effects on human phenotypes (**Supplementary Methods**; **Supplementary Table 9**).

We next addressed the extent to which height variants cluster near biologically relevant genes; specifically, genes mutated in human syndromes characterized by abnormal skeletal growth. We limited this analysis to the 652 genes occurring within the recombination hotspot-bounded regions surrounding each of the 180 index SNPs. We showed that the 180 loci associated with variation in normal height contained 21 of 241 genes (8.7%) found to underlie such syndromes (**Supplementary Table 10**), compared to a median of 8 (range 1-19) genes identified in 1,000 matched control sets of regions (P<0.001: 0 observations of 21 or more skeletal growth genes among 1,000 sets of matched SNPs). In 13 of these 21 loci the closest gene to the most associated height SNP in the region is the growth disorder gene, and in 9 of

these cases, the most strongly associated height SNP is located within the growth disorder gene itself (**Supplementary Methods, Supplementary Table 11**). These results suggest that GWA studies may provide more clues about the identity of the functional genes at each locus than previously suspected.

We also investigated whether significant and relevant biological connections exist between the genes within the 180 loci, using two different computational approaches. We used the GRAIL text-mining algorithm to search for connectivity between genes near the associated SNPs, based on existing literature[20]. Of the 180 loci, 42 contained genes that were connected by existing literature to genes in the other associated loci (the pair of connected genes appear in articles that share scientific terms more often than expected at $P<0.01$). For comparison, when we used GRAIL to score 1,000 sets of 180 SNPs not associated with height (but matched for number of nearby genes, gene proximity, and allele frequency), we only observed 16 sets with 42 or more loci with a connectivity $P<0.01$, thus providing strong statistical evidence that the height loci are functionally related ($P=0.016$) (**Figure 3a**). For the 42 regions with GRAIL connectivity $P<0.01$, the implicated genes and SNPs are highlighted in **Figure 3b**. The most strongly connected genes include those in the Hedgehog, TGF-beta, and growth hormone pathways.

As a second approach to find biological connections, we applied a novel implementation of gene set enrichment analysis (GSEA) (Meta-Analysis Gene-set Enrichment of variaNT Associations, MAGENTA[21]) to perform pathway analysis (**Supplementary Methods**). This analysis revealed 17 different biological pathways and 14 molecular functions nominally enriched ($P<0.05$) for associated genes, many of which lie within the validated height loci. These gene-sets include previously reported[11,13] (e.g. Hedgehog signaling) and novel (e.g. TGF-beta signaling, histones, and growth and development-related) pathways and molecular functions (**Supplementary Table 12**). Several SNPs near genes in these pathways narrowly missed genome-wide significance, suggesting that these pathways likely contain additional associated variants. These results provide complementary evidence for some of the genes and pathways highlighted in the GRAIL analysis. For instance, genes such as *TGFB2* and *LTBP1-3* highlight a role for the TGF-beta signaling pathway in regulating human height, consistent with the implication of this pathway in Marfan syndrome[22].

Finally, to examine the evidence for the potential involvement of specific genes at individual loci, we aggregated evidence from our data (eQTLs, proximity to the associated variant, pathway-based analyses), and human and mouse genetic databases (**Supplementary Table 13**). Of 32 genes with highly correlated ($r^2>0.8$) nsSNPs, several are newly identified strong candidates for playing a role in human

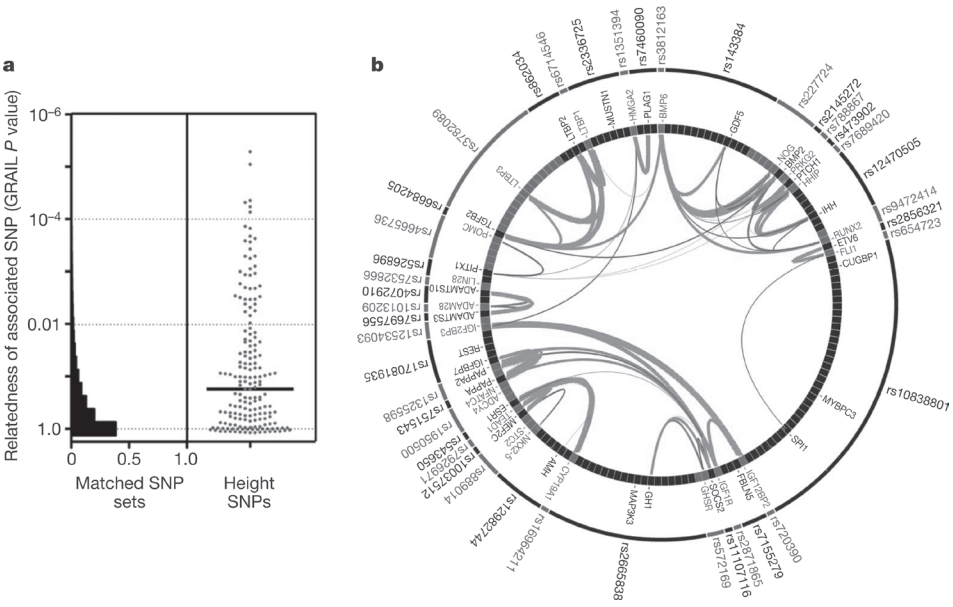


Figure 3. Loci associated with height contain genes related to each other. (a) 180 height-associated SNPs. The y-axis plots GRAIL P-values on a log scale. The histogram corresponds to the distribution of GRAIL P-values for 1,000 sets of 180 matched SNPs. The scatter plot represents GRAIL results for the 180 height SNPs (dots). The black horizontal line marks the median of the GRAIL P-values (P=0.14). The top 10 keywords linking the genes were: 'growth', 'kinase', 'factor', 'transcription', 'signaling', 'binding', 'differentiation', 'development', 'insulin', 'bone'. (b) Graphical representation of the connections between SNPs and corresponding genes for the 42 SNPs with GRAIL P<0.01. Thicker lines imply stronger literature-based connectivity.

growth. Some are in pathways enriched in our study (such as *ECM2*, implicated in extracellular matrix), while others have similar functions to known growth-related genes, including *FGFR4* (*FGFR3* underlies several classic skeletal dysplasias[23]) and *STAT2* (*STAT5B* mutations cause growth defects in humans[24]). Interestingly, *Fgfr4-/- Fgfr3-/-* mice show severe growth retardation not seen in either single mutant[25], suggesting that the *FGFR4* variant might modify *FGFR3*-mediated skeletal dysplasias. Other genes at associated loci, such as *NPPC* and *NPR3* (encoding the C-type natriuretic peptide and its receptor), influence skeletal growth in mice and will likely also influence human growth[17]. Many of the remaining 180 loci have no genes with obvious connections to growth biology, but at some our data provide modest supporting evidence for particular genes, including *C3orf63*, *PML*, *CCDC91*, *ZNFX1, ID4*, *RYBP, SEPT2*, *ANKRD13B*, *FOLH1*, *LRRC37B*, *MFAP2*, *SLBP*, *SOCS5*, and *ZBTB24* (**Supplementary Table 13**).

DISCUSSION

We have identified >100 novel loci that influence the classic polygenic trait of normal variation in human height, bringing the total to 180. Our results have potential general implications for genetic studies of complex traits. We show that loci identified by GWA studies highlight relevant genes: the 180 loci associated with height are non-randomly clustered within biologically relevant pathways and are enriched for genes that are involved in growth-related processes, that underlie syndromes of abnormal skeletal growth, and that are directly relevant to growth-modulating therapies (*GH1*, *IGF1R*, *CYP19A1*, *ESR1*). The large number of loci with clearly relevant genes suggests that the remaining loci could provide potential clues to important and novel biology.

We provide the strongest evidence yet that the causal gene will often be located near the most strongly associated DNA sequence variant. At the 21 loci containing a known growth disorder gene, that gene was on average 81 kb from the associated variant, and in over half of the loci it was the closest gene to the associated variant. Despite recent doubts about the benefits of GWA studies[26], this finding suggests that GWA studies are useful mapping tools to highlight genes that merit further study. The presence of multiple variants within associated loci could help localize the relevant genes within these loci.

By increasing our sample size to >100,000 individuals, we identified common variants that account for 10.5% of phenotypic variation. Although larger than predicted by some models[26], this figure suggests that GWA studies, as currently implemented, will not explain a majority of the estimated 80% contribution of genetic factors to variation in height. This conclusion supports the idea that biological insights, rather than predictive power, will be the main outcome of this initial wave of GWA studies, and that new approaches, which could include sequencing studies or GWA studies targeting variants of lower frequency, will be needed to account for more of the "missing" heritability. Our finding that many loci exhibit allelic heterogeneity suggests that many as yet unidentified causal variants, including common variants, will map to the loci already identified in GWA studies, and that the fraction of causal *loci* that have been identified could be substantially greater than the fraction of causal *variants* that have been identified.

In our study, many associated variants are tightly correlated with common nsSNPs, which would not be expected if these associated common variants were proxies for collections of rare causal variants, as has been proposed[27]. Although a substantial contribution to heritability by less common and/or quite rare variants may be more

plausible, our data are not inconsistent with the recent suggestion[28] that a large number of common variants of very small effect mostly explain the regulation of height.

In summary, our findings indicate that additional approaches, including those aimed at less common variants, will likely be needed to dissect more completely the genetic component to complex human traits. Our results also strongly demonstrate that GWA studies can identify large numbers of loci that together implicate biologically relevant pathways and mechanisms. We envision that thorough exploration of the genes at associated loci through additional genetic, functional, and computational studies will lead to novel insights into human height and other polygenic traits and diseases.

METHODS

The primary meta-analysis (Stage 1) included 46 GWA studies of 133,653 individuals. The *in-silico* follow up (Stage 2) included 15 studies of 50,074 individuals. All individuals were of European ancestry and >99.8% were adults. Details of genotyping, quality control, and imputation methods of each study are given in **Supplementary Methods Table 1-2**. Each study provided summary results of a linear regression of age-adjusted, within-sex Z scores of height against the imputed SNPs, and an inverse-variance meta-analysis was performed in METAL (http://www.sph.umich.edu/csg/abecasis/METAL/). Validation of selected SNPs was performed through direct genotyping in an extreme height panel (N=3,190) using Sequenom iPLeX, and in 492 Stage 1 samples using the KASPar SNP System. Family-based testing was performed using QFAM, a linear regression-based approach that uses permutation to account for dependency between related individuals[29], and FBAT, which uses a linear combination of offspring genotypes and traits to determine the test statistic[30]. We used a previously described method to estimate the amount of genetic variance explained by the nominally associated loci (using significance threshold increments from $P<5\times10^{-8}$ to $P<0.05$)[18]. To predict the number of height susceptibility loci, we took the height loci that reached a significance level of $P<5\text{x}10^{-8}$ in Stage 1 and estimated the number of height loci that are likely to exist based on the distribution of their effect sizes observed in Stage 2 and the power to detect their association in Stage 1. Gene-by-gene interaction, dominant, recessive and conditional analyses are described in **Supplementary Methods**. Empirical assessment of enrichment for coding SNPs used permutations of random sets of SNPs matched to the 180 height-associated SNPs on the number of nearby genes,

gene proximity, and minor allele frequency. GRAIL and GSEA methods have been described previously[20,21]. To assess possible enrichment for genes known to be mutated in severe growth defects, we identified such genes in the OMIM database (**Supplementary Table 10**), and evaluated the extent of their overlap with the 180 height-associated regions through comparisons with 1000 random sets of regions with similar gene content (±10%).

ACKNOWLEDGMENTS

A number of participating studies are members of CHARGE and ENGAGE consortia. We acknowledge funding from the Academy of Finland (104781, 117797, 120315, 121584, 126925, 129269, 129494, 129680, 213506); Affymetrix, Inc for genotyping services (N02-HL-6-4278); Agency for Science, Technology and Research of Singapore (A*STAR); ALF/LUA Gothenburg; Althingi (the Icelandic Parliament); Amgen; AstraZeneca AB; Australian National Health and Medical Research Council (241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 496688, 552485, 613672); Australian Research Council (DP0770096); Biocentrum Helsinki; Boston Obesity Nutrition Research Center (DK46200); British Diabetes Association; British Heart Foundation (PG/02/128); British Heart Foundation Centre for Research Excellence, Oxford; CamStrad; Cancer Research UK; Centre for Neurogenomics and Cognitive Research (CNCR-VU); Chief Scientist Office of the Scottish Government (CSO) (CZB/4/279); Council of Health of the Academy of Finland; DIAB Core project of the German Network of Diabetes; Diabetes UK; Donald W. Reynolds Foundation; Emil and Vera Cornell Foundation; Erasmus MC; Estonian Government (SF0180142s08); European Commission (201413, ECOGENE:205419, BBMRI:212111, OPENGENE:245536, ENGAGE:HEALTH-F4-2007-201413, EURODIA:LSHG-CT-2004-518153, EU/WLRT-2001-01254, HEALTH-F2-2008-ENGAGE, HEALTH-F4-2007-201550, LSH-2006-037593, LSHG-CT-2006-018947, LSHG-CT-2006-01947, Procardis:LSHM-CT-2007-037273, POLYGENE:LSHC-CT-2005, QLG1-CT-2000-01643, QLG2-CT-2002-01254, DG XII, Marie Curie Intra-European Fellowship); Eve Appeal; Finish Ministry of Education; Finnish Diabetes Research Foundation; Finnish Diabetes Research Society; Finnish Foundation for Cardiovascular Research; Finnish Medical Society; Finska Läkaresällskapet; Folkhälsan Research Foundation; Fondation LeDucq; Foundation for Life and Health in Finland; Foundation for Strategic Research (SSF); GEN-AU-Programme "GOLD"; Genetic Association InformationNetwork (GAIN); German Bundesministerium fuer Forschung und Technology (01 AK 803 A-H, 01 IG 07015 G); German Federal Ministry of Education and Research (BMBF) (01GS0831); German Ministry for Health, Welfare and Sports; German Ministry of Cultural Affairs; German Ministry of Education, Culture and Science; German National Genome Research Net (NGFN2 and NGFNplus) (01GS0823, 01ZZ0103, 01ZZ0403, 01ZZ9603, 03ZIK012); German Research Council (KFO-152); GlaxoSmithKline; Göteborg Medical Society; Gyllenberg Foundation; Helmholtz Center Munich; Juvenile Diabetes Research Foundation International (JDRF) (U01 DK062418); Karolinska Institute; Knut and Alice Wallenberg Foundation; Lundberg Foundation; March of Dimes (6-FY-09-507); MC Health; Medical Research Council UK (G0000649, G0000934, G0500539, G0600331, G0601261, G9521010D, PrevMetSyn); Microarray Core Facility of the Interdisciplinary Centre for Clinical Research (IZKF) (B27); Mid-Atlantic Nutrition and Obesity Research Center of Maryland (P30 DK072488); Ministry of Health and Department of Educational Assistance (South Tyrol, Italy); Ministry of Science, Education and Sport of the Republic of Croatia (216-1080315-0302); Montreal Heart Institute Foundation; Närpes Health Care Foundation; National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre; NIHR Oxford Biomedical Research Centre; NIHR comprehensive Biomedical Research Centre; National Institutes of Health (263-MA-410953, AA014041, AA07535, AA10248, AA13320, AA13321, AA13326, CA047988, CA49449, CA50385, CA65725, CA67262, CA87969, DA12854, DK062370, DK063491, DK072193,

DK079466, DK080145, DK58845, HG002651, HG005214, HG005581, HL043851, HL084729, HL69757, HL71981, K08-AR055688, K23-DK080145, K99-HL094535, M01-RR00425, MH084698, N01-AG12100, N01-AG12109, N01-HC15103, N01-HC25195, N01-HC35129, N01-HC45133, N01-HC55015, N01-HC55016, N01-HC55018 through N01-HC55022, N01-HC55222, N01-HC75150, N01-HC85079 through N01-HC85086, N01-HG65403, R01-AG031890, R01 CA104021, R01-DK068336, R01-DK073490, R01-DK075681, R01-DK075787, R01-HL086694, R01-HL087641, R01-HL087647, R01-HL087652, R01-HL087676, R01-HL087679, R01-HL087700, R01-HL088119, R01-HL59367, R01-MH059160, R01-MH59565, R01-MH59566, R01-MH59571, R01-MH59586, R01-MH59587, R01-MH59588, R01-MH60870, R01-MH60879, R01-MH61675, R01-MH63706, R01-MH67257, R01-MH79469, R01-MH81800, RL1-MH083268, T32-HG00040, U01-CA098233, U01-GM074518, U01-HG004399, U01-HG004402, U01-HL080295, U01-HL084756, U01-HL72515, U01-MH79469, U01-MH79470, U54-RR020278, UL1-RR025005, Z01-AG00675, Z01-AG007380, Z01-HG000024; contract HHSN268200625226C; ADA Mentor-Based Postdoctoral Fellowship; Pew Scholarship for the Biomedical Sciences); Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) (050-060-810); Netherlands Organisation for Scientific Research (NWO) (Investments nr. 175.010.2005.011, 911-03-012); Netherlands Organization for the Health Research and Development (ZonMw) (10-000-1002); Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, Center for Medical Systems Biology (NOW Genomics), SPI 56-464-1419) ; NIA Intramural Research Program; Nordic Center of Excellence in Disease Genetics; Novo Nordisk Foundation; Ollqvist Foundation; Oxford NIHR Biomedical Research Centre; Paavo Nurmi Foundation; Perklén Foundation; Petrus and Augusta Hedlunds Foundation; Queensland Institute of Medical Research; Radboud University Nijmegen Medical Centre; Research Institute for Diseases in the Elderly (014-93-015); Royal Swedish Academy of Science; Sahlgrenska Center for Cardiovascular and Metabolic Research (A305:188); Siemens Healthcare, Erlangen, Germany; Signe and Ane Gyllenberg Foundation; Sigrid Juselius Foundation; Social Insurance Institution of Finland; Social Ministry of the Federal State of Mecklenburg-West Pomerania; South Tyrolean Sparkasse Foundation; Stockholm County Council (560183); Support for Science Funding programme; Susan G. Komen Breast Cancer Foundation; Swedish Cancer Society; Swedish Cultural Foundation in Finland; Swedish Foundation for Strategic Research; Swedish Heart-Lung Foundation; Swedish Medical Research Council (K2007-66X-20270-01-3, 8691); Swedish National Cancer Institute; Swedish Research Council; Swedish Society of Medicine; Swiss National Science Foundation (33CSCO-122661); Torsten and Ragnar Söderberg's Foundation; Vandervell Foundation; Västra Götaland Foundation; Wellcome Trust (072960, 075491, 079557, 079895, 083270, 068545/Z/02, 076113/B/04/Z, 076113/C/04/Z, 076113/C/04/Z, 077016/Z/05/Z, 081682/Z/06/Z, 084183/Z/07/Z, 085301/Z/08/Z, 086596/Z/08/Z, 091746/Z/10/Z; WT Research Career Development Fellowship); Western Australian Genetic Epidemiology Resource and the Western Australian DNA Bank (both National Health and Medical Research Council of Australia Enabling Facilities). Detailed list of acknowledgments is given in the Supplementary Information.

AUTHOR CONTRIBUTIONS

Full author contributions and roles are listed in the Supplementary Information.

Note: Supplementary information is available on the Nature website**:**
http://www.nature.com/nature/journal/v467/n7317/extref/nature09410-s1.pdf

REFERENCES

1. Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
2. Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* **5**, 329-348 (1885).
3. Fisher, R.A. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918).
4. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
5. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004. (1999).
6. Reich, D.E. & Goldstein, D.B. Detecting association in a case-control study while correcting for population stratification. **20**, 4-16 (2001).
7. Campbell, C.D. et al. Demonstrating stratification in a European-American population. *Nature Genet*, 37:868-872 (2005).
8. Manolio, T.A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
9. Visscher, P.M. et al. Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLoS Genet* **2**, e41 (2006).
10. Weedon, M.N. et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* **39**, 1245-1250 (2007).
11. Weedon, M.N. et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575-83 (2008).
12. Sanna, S. et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* **40**, 198-203 (2008).
13. Lettre, G. et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* **40**, 584-91 (2008).
14. Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U. & Malkina, I. Meta-analysis of genome-wide scans for human adult stature in humans identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet* **5**, e1000445 (2009).
15. Gudbjartsson, D.F. et al. Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**, 609-15 (2008).
16. Johansson, A. et al. Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum Mol Genet* **18**, 373-80 (2009).
17. Estrada, K. et al. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum Mol Genet* **18**, 3516-24 (2009).
18. Purcell, S.M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
19. Park, J.H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42**, 570-5 (2010).
20. Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).

21. Segrè, A.V. et al. Common Inherited Variation in Mitochondrial Genes is not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS Genet*, In press (2010).

22. Neptune, E.R. et al. Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* **33**, 407-11 (2003).

23. Superti-Furga, A. & Unger, S. Nosology and classification of genetic skeletal disorders: 2006 revision. *Am J Med Genet A* **143**, 1-18 (2007).

24. Kofoed, E.M. et al. Growth hormone insensitivity associated with a STAT5b mutation. *N Engl J Med* **349**, 1139-47 (2003).

25. Weinstein, M., Xu, X., Ohyama, K. & Deng, C.X. FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125**, 3615-23 (1998).

26. Goldstein, D.B. Common genetic variation and human traits. *N Engl J Med* **360**, 1696-8 (2009).

27. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol* **8**, e1000294 (2010).

28. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).

29. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

30. Laird, N.M., Horvath, S. & Xu, X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* **19 Suppl 1**, S36-42 (2000).

CHAPTER 4

GENETICS OF OSTEOPOROSIS

CHAPTER 4.1

Twenty bone-mineral-density loci identified by large-scale meta-analysis of GWAS

Nat Genet 41, 1199-206 (2009)

ABSTRACT

Bone mineral density (BMD) is a heritable complex trait used in the clinical diagnosis of osteoporosis and the assessment of fracture risk. We performed meta-analysis of five genome-wide association studies of femoral neck and lumbar spine BMD in 19,195 subjects of Northern European descent. We identified 20 loci reaching genome-wide significance (GWS; $P<5x10^{-8}$), of which 13 map to new regions including 1p31.3 (*GPR177*), 2p21 (*SPTBN1*), 3p22 (*CTNNB1*), 4q21.1 (*MEPE*), 5q14 (*MEF2C*), 7p14 (*STARD3NL*), 7q21.3 (*FLJ42280*), 11p11.2 (*LRP4*; *ARHGAP1*; *F2*), 11p14.1 (*DCDC5*), 11p15 (*SOX6*), 16q24 (*FOXL1*), 17q21 (*HDAC5*) and 17q12 (*CRHR1*). The meta-analysis also confirmed at GWS level, seven known BMD loci on 1p36 (*ZBTB40*), 6q25 (*ESR1*), 8q24 (*TNFRSF11B*), 11q13.4 (*LRP5*), 12q13 (*SP7*), 13q14 (*TNFSF11*), and 18q21 (*TNFRSF11A*). The numerous SNPs associated with BMD map to genes in signaling pathways with relevance to bone metabolism, and highlight the complex genetic architecture underlying osteoporosis and BMD variation.

INTRODUCTION

Osteoporosis is a condition characterized by reduced bone mass and microarchitectural deterioration of bone tissue, leading to loss of bone strength and increased risk of fracture. Osteoporosis increases in incidence with age, and is a major health threat to hundreds of millions of elderly individuals worldwide. Linkage analysis in monogenic bone disorders like Osteoporosis-Pseudoglioma syndrome, sclerosteosis, high bone mass syndrome and Paget's disease, have yielded major advances in recent years and highlighted the importance of the *Wnt* signaling[1] and the RANK/RANKL/Opg pathways in the regulation of bone mass and bone turnover[2]. Linkage studies for the common form of osteoporosis have had limited success however[3]. As with other complex diseases, most of the candidate gene association studies in osteoporosis have produced conflicting results with limited replication[4] mostly due to small sample sizes[5]. However, large-studies of major candidate gene polymorphisms within the sufficiently powered setting of the GENOMOS consortium, have been successful in obtaining consistent evidence of association between some genetic variants, BMD and fracture[6-10]. Concurrently, genome-wide association studies (GWAS), which perform a hypothesis-free search for genetic determinants,[11] have already been successful in identifying 10 loci associated at a genome-wide significance (GWS) level with BMD[12,13]. Four of these loci involve members of the *Wnt* and *RANK*/*RANKL*/*Opg* signaling pathways highlighting their role in monogenic forms as well as the common form of osteoporosis.

BMD is used in clinical practice for the diagnosis of osteoporosis and in the assessment of fracture risk[14]. BMD is usually measured at the hip (femoral neck) and lumbar spine, which are common sites of fracture. However, BMD measurements at different skeletal sites are predictive of fracture in general because of their high correlation (r^2~0.60).[15] From a genetic perspective, BMD at both spine and hip is a complex, but highly heritable trait (h^2~0.60-0.80)[16]. As shown for human height[17-19], dozens and possibly hundreds of loci with small effects can be expected to influence the variation in BMD. Detection and reliable documentation of these loci of weak effects requires studies with comprehensive coverage of the genome and very large sample sizes. Here, we report the findings of a large-scale meta-analysis of 19,195 adult individuals from five GWAS on BMD of the lumbar spine (LS) and the femoral neck (FN), within the setting of the *GEnetic Factors of OSteoporosis* (GEFOS) consortium.

RESULTS

Samples, genotyping and (meta) analysis of genome-wide scans

The five study populations included the Rotterdam Study (RS, n=4,987), Erasmus Rucphen Family Study (ERF, n=1,228), Twins UK Study (TUK, n=2,734), deCODE Genetics Study (dCG, n=6,743) and the Framingham Osteoporosis Study (FOS, n=3,503). The age of participants ranged from 18 to 96 years. All studies had a majority of women (range 57-88%) in their samples with TUK including women only. Additional characteristics of the study populations and subject exclusion criteria are presented in Supplementary Tables 1 and 2, respectively.

BMD loci identification

Association results (corrected by the genomic control method[20]) of all HapMap CEU imputed SNPs passing quality-control (QC) criteria in each study (Supplementary Table 3), were meta-analyzed using METAL[21]. We declared results genome-wide significant at $\alpha=5x10^{-8}$ after adjusting for all common variant tests in the human genome[22,23]. We investigated if there was an excess of significant associations by comparing the test statistics to those expected under the null distribution using inter quantile-quantile (QQ) plots (corrected for overall meta-analysis genomic control $\lambda_{LS\text{-}BMD}=1.09$ and $\lambda_{FN\text{-}BMD}=1.08$). As observed in Figure 1, strong (and not early) deviation of the observed statistics from the null distribution was observed for both BMD traits, corresponding to an excess of significant and likely true associations. Excluding all SNPs within 500 Kb of the SNPs associated at a GWS level and correction for overall meta-analysis genomic control, still left many SNPs associated with BMD more than expected by chance alone. This suggests that among many false positives appearing at less stringent statistical thresholds, additional truly associated BMD variants may exist.

The meta-analysis identified 467 SNPs from 20 genomic loci exceeding the GWS threshold of association with the BMD traits (Figure 2). Of these, 15 loci associated with LS-BMD (Supplementary Tables 4A and 5A) and 10 with FN-BMD (Supplementary Tables 4B and 5B); five of these loci were associated with both skeletal sites. The effect sizes and significance of the top SNPs from the 20 regions containing markers associated with LS- and FN-BMD at GWS are presented in Table 1. Applying correction for overall- meta-analysis genomic control resulted in seven of the 20 loci not to be GWS (with $7.1x10^{-7}>P>5.0x10^{-8}$). For most markers heterogeneity was not very large

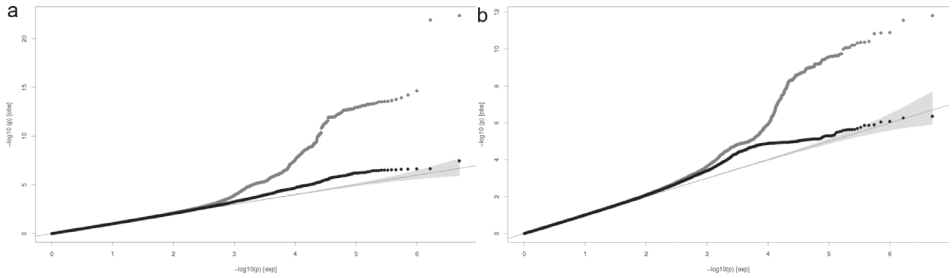


Figure 1. Quantile-Quantile (Q-Q) plots. **a)** Lumbar Spine BMD. **b)** Femoral Neck BMD. The plots compare additive model statistics to those expected under the null distribution using fixed-effects for all analyzed HapMap CEU imputed SNPs passing quality-control (QC) criteria in the studies (upper lines), and after exclusion of all genome-wide significant and correlated (r^2>0.1) SNPs (lower lines).

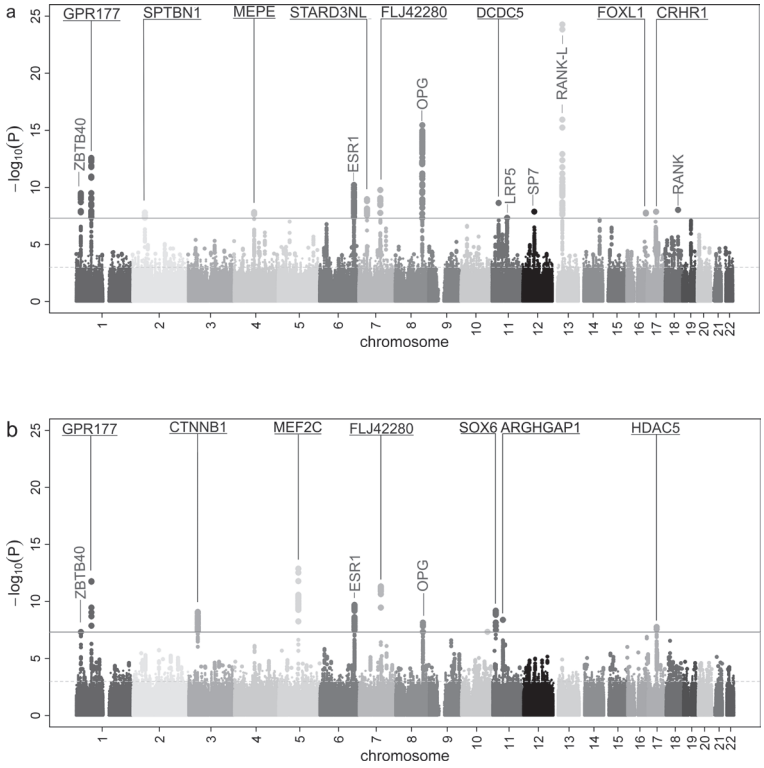


Figure 2. Manhattan plots. (a,b) Plots displaying novel and previously reported (known) loci associated at genome-wide significant level (GWS) with Lumbar Spine BMD (a) and Femoral Neck BMD (b) for all 2,543,686 HapMap CEU-imputed SNPs analyzed using fixed-effects. The 13 new GWS loci are in bold and underlined type.

Table 1. Top genome-wide significant markers of the 20 loci associated at GWS with lumbar spine and/or femoral neck BMD

	Marker Information						Lumbar spine BMD Effect estimate (in SD)		
Locus	SNP	A1/A2	FREQ	Type[a]	Closest Gene	Distance[b] (kb)	Beta (β_{LS})	SE	P-value
Novel loci associated with BMD at GWS level									
1p31.3	rs1430742	C/T	0.21	I	GPR177	44.0	0.105	0.014	**2.6E-13**
	rs2566755	C/T	0.21	G	GPR177	44.3	0.104	0.014	**3.3E-13**
3p22	rs87938	A/G	0.45	G	CTNNB1	103.3	-0.043	0.012	1.7E-04
5q14	rs1366594	C/A	0.45	G	MEF2C	197.0	0.005	0.012	0.65
7p14	rs1524058	T/C	0.40	I	STARD3NL	81.7	-0.070	0.012	**1.1E-09**
7q21.3	rs4729260	G/C	0.32	I	FLJ42280	14.9	-0.081	0.013	**1.7E-10**
	rs7781370	T/C	0.34	I	FLJ42280	0.7	-0.074	0.012	**1.1E-09**
11p14.1	rs16921914	A/G	0.27	G	DCDC5	61.6	0.077	0.013	**2.3E-09**
11p15	rs7117858	G/A	0.20	G	SOX6	297.3	-0.042	0.014	0.004
16q24	rs10048146	G/A	0.19	G	FOXL1	95.4	-0.093	0.016	**1.7E-08**
17q12	rs9303521	T/G	0.46	G	CRHR1	56.5	-0.068	0.012	**1.4E-08**
Suggestive loci now associated with BMD at GWS level									
2p21	rs11898505	A/G	0.34	G	SPTBN1	1.1	0.067	0.012	**1.6E-08**
4q21.1	rs1471403	T/C	0.34	G	MEPE	7.3	0.068	0.012	**1.5E-08**
11p11.2	rs7932354	T/C	0.29	I	ARHGAP1	0.1	0.056	0.013	1.1E-05
17q21	rs228769	G/C	0.20	I	HDAC5	7.8	0.067	0.014	4.0E-06
Known loci associated with BMD at GWS level									
1p36	rs7524102	G/A	0.17	G	ZBTB40	79.9	0.094	0.015	**3.2E-10**
	rs6426749	C/G	0.17	I	ZBTB40	66.9	0.107	0.017	**7.6E-10**
6q25	rs2504063	A/G	0.40	G	ESR1	38.0	-0.078	0.012	**6.1E-11**
	rs2941740	G/A	0.43	I	C6orf97	67.3	0.070	0.012	**2.0E-09**
8q24	rs2062377	T/A	0.44	I	TNFRSF11B	43.0	0.094	0.012	**3.5E-16**
	rs11995824	G/C	0.55	I	TNFRSF11B	48.3	-0.093	0.012	**1.1E-15**
11q13.4	rs599083	G/T	0.31	G	LRP5	24.4	-0.067	0.012	**4.7E-08**
12q13	rs2016266	G/A	0.32	G	SP7	1.6	0.070	0.012	**1.3E-08**
13q14	rs9533090	T/C	0.50	I	AKAP11	54.0	-0.120	0.012	**5.4E-25**
18q21	rs884205	A/C	0.27	I	TNFRSF11A	1.4	-0.078	0.014	**9.4E-09**

Bold face indicates *$P<5x10^{-8}$. NC, not calculated. [a]Type of SNP: G=Genotyped (at least in 1 study) I=Imputed. [b]Distance to coding region. [c]OMA-GC, overall meta-analysis genomic control. [d]Q P value, Q-statistic P value. [e]Site-spec. H0, site-specificity null hypothesis, i.e., $\beta_{LS} = \beta_{FN}$.*

Lumbar spine BMD						Femoral neck BMD			Site Specificity H_o:β_{LS}=β_{FN}
	Heterogeneity		Effect estimate (in SD)				Heterogeneity		
OMA-GC P-value[c]	Q P-value[d]	I^2	Beta	SE	P-value	OMA-GC P-value[c]	Q P-value[d]	I^2	P-value[e]
2.5E-12	0.26	21	0.100	0.014	**1.8E-12**	**1.2E-11**	0.75	0	0.82
3.1E-12	0.27	20	0.100	0.014	**1.7E-12**	**1.1E-11**	0.76	0	0.83
3.1E-04	0.24	23	-0.070	0.011	**8.1E-10**	**3.4E-09**	0.14	34	0.10
0.66	***NC***	***NC***	-0.085	0.011	**1.3E-13**	**1.1E-12**	0.62	0	3.2E-08
5.2E-09	0.18	30	-0.038	0.011	8.9E-04	1.4E-03	0.05	48	0.05
9.5E-10	0.14	35	-0.085	0.012	**9.4E-12**	**5.4E-11**	0.77	0	0.82
5.5E-09	0.12	37	-0.083	0.012	**4.7E-12**	**2.9E-11**	0.68	0	0.60
1.0E-08	0.52	0	0.038	0.013	0.003	0.005	***NC***	***NC***	0.03
0.005	***NC***	***NC***	0.088	0.014	**6.4E-10**	**2.7E-09**	0.70	0	1.5E-10
6.0E-08	0.21	28	-0.085	0.016	1.7E-07	4.9E-07	0.96	0	0.73
5.0E-08	0.05	49	-0.055	0.012	3.6E-06	8.3E-06	0.07	45	0.46
6.3E-08	0.39	6	0.027	0.012	0.02	0.03	***NC***	***NC***	0.02
5.7E-08	0.18	30	0.059	0.012	7.8E-07	2.0E-06	0.56	0	0.58
2.4E-05	0.61	0	0.073	0.012	**4.0E-09**	**1.5E-08**	0.68	0	0.32
1.0E-05	0.87	0	0.081	0.014	**1.7E-08**	5.8E-08	0.94	0	0.49
1.7E-09	0.27	19	0.079	0.015	**8.8E-08**	2.6E-07	0.64	0	0.48
3.8E-09	0.74	0	0.082	0.015	**4.8E-08**	1.5E-07	0.64	0	0.26
3.7E-10	0.03	53	-0.066	0.012	**3.0E-08**	9.6E-08	0.84	0	0.45
9.3E-09	0.01	59	0.073	0.012	**2.0E-10**	**9.1E-10**	0.02	57	0.84
5.7E-15	0.31	15	0.062	0.011	5.4E-08	1.7E-07	0.61	0	0.05
1.6E-14	0.24	22	-0.066	0.011	**7.1E-09**	**2.6E-08**	0.48	0	0.10
1.7E-07	0.50	0	-0.047	0.012	9.7E-05	0.0002	0.76	0	0.25
5.2E-08	0.93	0	0.046	0.012	1.9E-04	0.0003	0.81	0	0.16
4.6E-23	0.02	57	-0.041	0.011	3.9E-04	0.0006	0.39	6	1.1E-06
3.8E-08	0.90	0	-0.039	0.013	0.004	0.005	***NC***	***NC***	0.04

or statistically significant.

Despite the correlation between LS- and FN-BMD measurements, site-specific effects were observed. Seven of the 20 loci showed evidence for skeletal site-specificity (P<0.05), three of which displayed strong evidence for site-specificity ($P<1x10^{-6}$). This site-specificity is to be expected given the differences in heritability and that the genetic correlation (or fraction of "shared" heritability) between the measurements is considerably less than 1 (Supplementary Table 6).

Genes in associated regions and their function

Of the 20 BMD loci identified in this genome scan at a GWS level, seven have been reported previously as GWS[13,24], whereas the remaining 13 have not. Of these 13 loci not previously associated with BMD at a GWS level, four were suggestively associated in previous reports ([12,13]), whereas nine are novel loci. In Supplementary Table 7 we present a summary of relevant gene annotations including related pathways, monogenic syndromes, knockout mouse models, and additional functional details of the genes most likely to be underlying the associated signals in these 13 loci.

Novel loci associated with BMD at GWS level

There are nine loci displaying novel associations with BMD for which we present Forest plots of effects (Figure 3) and regional association plots (Supplementary Figure 1) from the top SNPs.

***1p31.3* locus [*GRP177*]**.Two common SNPs (minor allele frequency (MAF)=0.21) in complete pair wise LD (rs1430742 and rs2566755) were associated at a GWS level with both FN- and LS-BMD. Both top SNPs are located within an intronic region of the *G protein-coupled receptor 177* (*GPR177*, also named WNTLESS homologue) gene. *GPR177* is part of the highly evolutionary conserved *Wnt* signaling pathway,[25] involved in bone cell differentiation and development. The gene has been shown to be positive regulator of the *I-kappaB kinase*/*NF-kappaB* cascade, part of the RANK system. Cross-talk between the NF-*k*B and the mitogen-activated protein kinase (*MAPK*) pathway has been indicated by the identification of several overlapping genes expressed in both pathways, including several G protein-coupled receptors[26].

***3p22* locus [*CTNNB1*].** The rs87939 SNP (MAF=0.45) located 103 Kb upstream of the *catenin (cadherin-associated protein), beta 1* (*CTNNB1*) gene on chromosome 3 was associated at a GWS level with FN-BMD. *CTNNB1* is integral to the *Wnt* signaling pathway, and as such, is an excellent candidate for BMD regulation, considering that

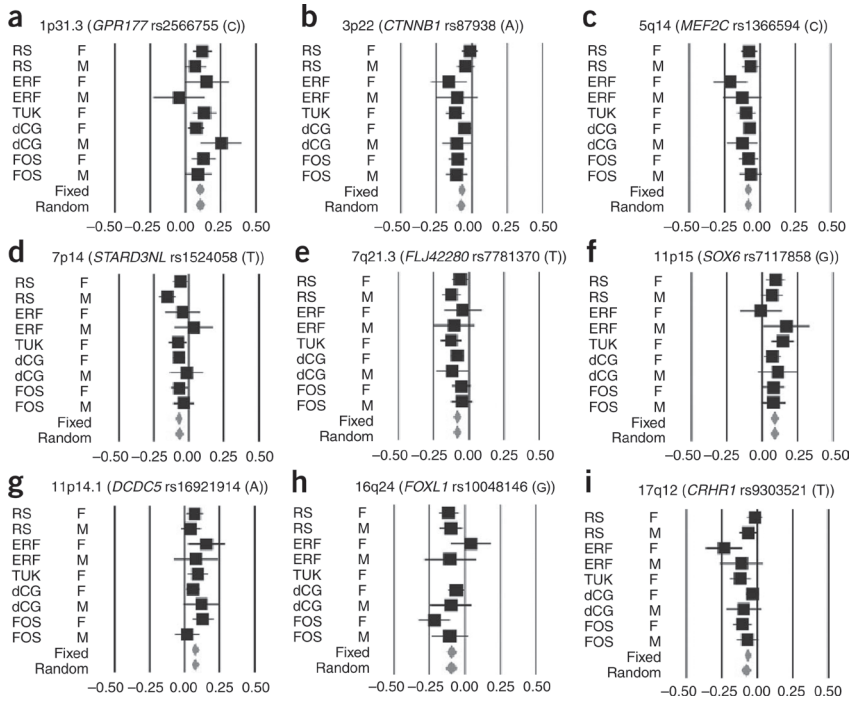


Figure 3. Forest Plots for the top SNPs for each of the 9 novel loci: **a)** *1p31.3*, **b)** *3p22*, **c)** *5q14*, ***d)*** *7p14*, *e) 7q21.3*, f) *11p15*, g) *11p14.1*, ***h)*** *16q24*, **i)** *17q12*. Black squares represent effect estimate and 95%CI for each study, and the diamonds are summary effect. Measurements units are in BMD standard deviations (SD).

Wnt signaling controls the process of bone resorption by (negative) regulation of *Opg* (*TNFRS11B*) expression in osteoblasts[27].

***5q14* locus [*MEF2C*].**The rs1366594 top SNP (MAF=0.45) displayed skeletal site-specificity being associated with FN-BMD only, together with other 60 GWS markers at this locus. The SNP is located 197 Kb upstream of the *MADS box transcription enhancer factor 2, polypeptide C* (*MEF2C*) gene with no other known annotation within the large LD block. *MEF2C* is a transcription factor highly expressed in muscle which allows transcriptional cross-talk between the Ca2+/calmodulin-dependent kinase (*CaMK*) and mitogen-activated protein kinase (*MAPK)* signaling pathways by signal-dependent dissociation from histone deacetylases (HDACs). *MEF2C* interacts with *HDAC4* and *HDAC5* (see below 17q21 region), resulting in repression of the transcriptional activity of *MEF2C*. Potthoff et al. showed in mice that HDACs selectively degraded by the proteasome enables *MEF2C* to activate the slow myofiber gene program, resulting in enhanced endurance during physical exercise[28].

***7p14-p13* locus [*STARD3NL*].** The rs1524058 SNP (MAF=0.40) located on the short arm of chromosome 7 is 81 Kb upstream of the *STARD3 n-terminal like* (*STARD3NL*) gene, a cholesterol endosomal transporter, was associated at a GWS level with LS-BMD but less strongly with FN-BMD. Very recently, a study on Asian individuals[29] identified a SNP in this 7p14 region consistently associated with BMD measurements of the radius, tibia and heel (lowest $P=1.4x10^{-7}$). Yet, the rs1721400 SNP was not associated with either LS- or FN- BMD in our study (P>0.05). The rs1721400 marker (mapping close to *SFRP4*) is in very low LD with our top SNP rs1524058 in HapMap individuals of European descent ($r^2=0.015$, D'=0.194), while LD between markers is considerably higher in Chinese or Japanese individuals ($r^2=0.230$, D'=0.931). Thus, an underlying signal common to both Europeans and Asians may be still be captured by these SNPs considering the differences in LD across populations.

***7q21.3* locus [*FLJ42280*].** Several SNPs are GWS in this region, with the two top SNPs being in moderate pair wise LD ($r^2=0.36$; D'=0.84) and associated at GWS with both LS- and FN-BMD. The SNPs are located within (rs4729260) and just upstream (rs7781370) of *FLJ42280*, a gene of unknown function. There are several genes within the ~480 kb LD stretch region, of which the *split hand/foot malformation (ectrodactyly) type 1* (*SHFM1)* is the closest gene (87-185 kb away).

***11p15* locus [*SOX6*].**The rs7117858 SNP (MAF=0.20) was associated at GWS level with FN-BMD only, displaying strong evidence for skeletal site-specificity. The SNP is located 297kb upstream from the *SRY (sex determining region Y)-box* (*SOX6*) gene, which is a transcription factor of the SOX gene family defined by a conserved high mobility group (*HMG*) DNA-binding domain. The gene is expressed in a wide variety of tissues, most abundantly in skeletal muscle. *Sox6* knock-out mice exhibit early lethality due to cardiac insufficiency and present with mild skeletal abnormalities affecting size and mineralization of endochondral elements[30]. Other SOX-family genes regulate *RUNX2*-mediated differentiation of mesenchymal cells during endochondral ossification (skeletogenesis)[31].

***11p14.1* locus [*DCDC5*; *DCDC1*]**. The rs16921914 SNP (MAF=0.27) is the only marker in the 11p14.1 region associated with LS-BMD at a GWS level. All other associated SNPs are in moderate LD with rs16921914 ($r^2<0.70$) and display less strong associations ($P>1x10^{-7}$). The SNP is located 62 kb downstream of the *doublecortin domain containing 1* (*DCDC1*) and 73 kb upstream of the *DCDC5* genes. Doublecortin domains are highly conserved elements which serve as protein-interaction platforms[32]. Mutations in members of this protein superfamily are linked to several neurogenetic diseases and to our knowledge are not expressed in bone.

***16q24.3* locus [*FOXC2*; FOXL1].** The rs10048146 SNP (MAF=0.19) was associated

with LS-BMD (-0.09 and is located on the subtelomeric region of chromosome 16, about 95Kb downstream from a cluster of small (1 Kb) genes of the "forkhead" (or winged helix) gene family. The genes are mainly expressed in the gastrointestinal mucosa (*FOXL1*) or are involved in adipocyte metabolism and early stage chondrogenic differentiation (*FOXC2*). *FOXC2* stimulates osteoblast differentiation of mesenchymal cells through activation of canonical *Wnt-beta-catenin* signals[33] while *FOXC2* expression has been shown to occur via bone morphogenetic proteins[34]. Skeletal defects of the spine have been reported in *FOXC2* mouse knockout models[35] and recently, deletions and inactivating mutations affecting the *FOX* gene cluster have been identified as causing severe malformations of the VACTER type in humans, which include vertebral malformations.[36]

***17q12-q22* [*CRHR1*].** The rs9303521 SNP (MAF=0.46) on chromosome 17 was associated with LS-BMD and is located 56 Kb from the *corticotrophin-releasing factor receptor* (*CRHR1*) gene. Among other genes in this LD region, MAP3K14 is a candidate potentially involved in bone-active pathways, particularly through the activation of NF-kappa-B.

Suggestive loci now associated with BMD at GWS level

Four BMD loci which now reach for the first time a GWS level were "suggestively" associated with BMD in previous reports[12,13]. We present Forest plots of effects (Figure 4) and regional association plots (Supplementary Figure 2) from these SNPs.

***2p21 locus* [*SPTBN1*].** The rs11898505 (MAF=0.34) on chromosome 2 is intronic to the spectrin, beta, non-erythrocytic 1 (*SPTBN1*) gene which encodes a major cytoskeletal scaffolding protein, and was associated with LS-BMD. The same SNP was previously shown to be associated with BMD and fractures, even though it did not reach GWS level for BMD in the same study.[13] In mice, disruption of *beta-spectrin* isoforms (Elf) leads to the disruption of TGF-beta signaling by Smad proteins[37].

***4q21.1* locus [*MEPE*].** The 4q21.1 region contains a cluster of structurally and phylogenetically related genes encoding matricellular phosphoglycoproteins with function in bone formation and growth[38]. The top associated rs1471403 SNP (MAF=0.34) is located 7 Kb 3' to the matrix, extracellular, phosphoglycoprotein (*MEPE*) gene (also known as osteoblast/osteocyte factor 45), 42 Kb to the integrin-binding sialoprotein (*IBSP*) gene and 122 Kb 5' to the secreted phosphoprotein 1 (*SPP1*) gene, also known as osteopontin. *IBSP* and *SPP1* are highly expressed in osteoblasts, osteoclasts and hypertrophic chondrocytes. *MEPE* is predominantly expressed by osteocytes in human bone, playing an inhibitory role in bone formation.

All three genes display diverse skeletal phenotypes in mice knock out (KO) models. *MEPE* (Of45) KO show increased bone mass and inhibition of age-related bone loss[39], *IBSP* KO show high trabecular bone density with low bone turnover but respond to bone loss caused by disuse[40] and the *SPP1* KO have high trabecular bone mass and is resistant to bone loss[41]. Previously, a non-synonymous SNP in the IBSP gene (rs1054627, G195E) in moderate LD with (r2=0.2, D'=0.8) was reported as suggestively associated with hip BMD.[12]

***11p11.2* locus [*ARHGAP1*;*LRP4*]**

At the 11p11.2 region a large LD block extends the region withholding several genes, including *C11orf49*, *LRP4*, *CKAP5*, *F2*, *ZN408* and *ARHGAP1* among others. Two fully correlated SNPs (r^2=1;MAF=0.29) including rs7932354, which lies in the promoter region of the *Rho GTPase activating protein 1* (*ARHGAP1*) gene; and rs2070852, located in intron 5 of the coagulation factor II (*F2*) gene were associated with FN-BMD at GWS level. Other correlated SNPs in the region (r^2 between 0.2 and 0.8) were previously suggestively associated with hip BMD and attributed to the *LRP4* gene[13]. *ARHGAP1*, a ubiquitous factor composing one of the GTPase activating proteins that represses RhoA, is another good candidate. RhoA is a small G-protein of the Rho family that regulates cell morphology via actin- cytoskeleton reorganization and which is thought to be a potential commitment switch in the differentiation of mesenchymal stem cells to osteoblasts[42]. In addition, data from *ARHGAP1* KO mouse models show a strong skeletal phenotype, including a 3-fold reduction in BMD, decreased cortical thickness and bone fragility in older animals[43].

***17q21* locus [*HDAC5;C17orf53*].** The 17q21 region contains more than 30 genes in 1 Mb surrounding the top rs228769 SNP (MAF=0.20) which is located 8 Kb upstream of the *histone deacetylase 5* (*HDAC5*) and 26 kb upstream of the *C17orf53* genes. A non synonymous SNP in the *C17orf53* gene, rs227584 (T126P), moderately correlated to rs228769 (r^2=0.64) was found associated with hip BMD, albeit not at GWS level[12]. These SNPs likely represent the same signal. *HDAC5* is a class histone deacetylase II (homologous to yeast Hda1), ubiquitously expressed and responsible for the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4). Histone acetylation and deacetylation plays an important role in transcriptional regulation, cell cycle progression and developmental events, particularly for myocyte differentiation[44]. In undifferentiated myoblasts, *HDAC5* is present in the nucleus where it binds to the myocyte enhancer *MEF2C* (see above) to repress transcription and detain muscle maturation. In bone, recruitment of class II histone deacytylases like *HDAC5*, is needed for TGF-β mediated osteoblast differentiation[45], which occurs through inhibition of Runx2 function by Smad3[46].

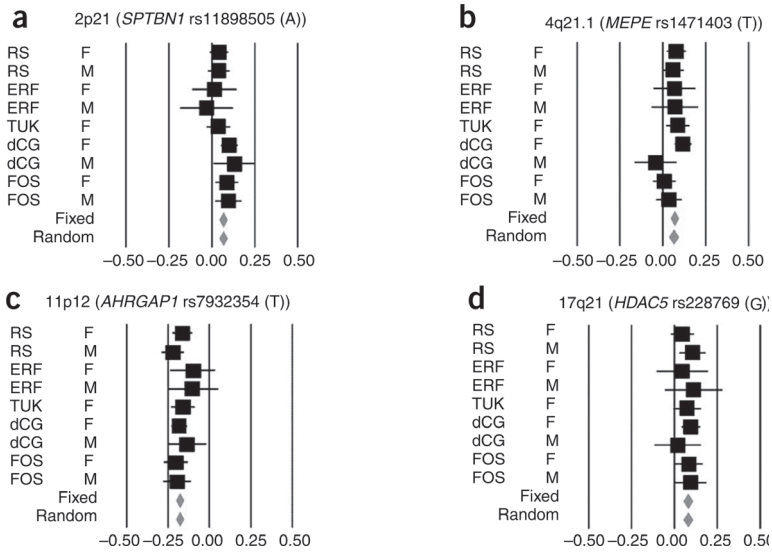


Figure 4. Forest Plots for the top SNPs for each of the 4 loci attaining GWS for the first time in this study: **A.** *2p21*, **B**. *4q21.1*, **C.** *11p11.2*, **D.** *17q21*. Black squares represent effect estimate and 95%CI for each study, and the red diamonds are summary effect estimates. Measurements units are in BMD standard deviations (SD).

Known loci associated with BMD at GWS level

We have replicated at GWS level 7 loci associated with BMD in previous GWA studies[12,13,24]. They include the 1p36 (*ZBTB40*), 6q25 (*ESR1*), 8q24 (*TNFRSF11B;*Opg), 11q13.4 (*LRP5*), 12q13 (*SP7*;Osterix), 13q14 (*TNFSF11*;RANKL), and 18q21 (*TNFRSF11A;*RANK) regions. Regional association plots for all seven regions are included in Supplementary Figure 3. In the 6q25 (*ESR1*) region at least two independent GWS signals are seen at the sides of a high recombination rate peak (Supplementary Figure 3B). This locus showed a complex pattern of association in a previous study indicating three independent signals at the locus[13]. In Supplementary Table 8 we report the associations observed in this study for all loci reported previously as attaining GWS or "suggestive" association[12,13,24]. SNPs from the 6p21.32 (*MHC*), 14q32 (*MARK3*) and 17q21 (*SOST*) regions described previously to be GWS[13] were still significantly associated to BMD in our study, but not at a GWS level. It should be noted that in the previous study the 14q32 and the 17q21 regions were associated with total hip BMD, which differs from the femoral neck BMD phenotype used in the current study.

Gene expression (eQTL) associations

We tested the association of SNPs (or proxies) from the 13 newly GWS associated regions with cis-allelic expression of gene transcripts in primary human osteoblasts. All SNPs associated with gene expression at P<0.05 and located within the same LD block of the strongest associated variants (D'≥0.8) are presented in Supplemental Table 9. Associations were seen for transcripts of *GPR177*, *MEF2C* and *FOXC2*. Similarly, for variants in (or in LD with variants in) *MEPE* the most significant correlation with expression in osteoblasts was seen with the integrin bone sialoprotein (*IBSP)* gene, while *MEPE* seems not highly expressed in osteoblasts. Yet, the statistical evidence is not fully conclusive since only subtle overrepresentation of the associated loci was observed (10.5% vs 7% for non-associated control SNPs, Chi^2=8.9 and P=0.003). The small overrepresentation of the associated loci suggests several of the associated genes may be expressed in cell & tissues other than osteoblast lineages.

Combined effect of the 20 GWS BMD loci and fracture risk

We examined the combined effect of the top SNPs arising from the 20 associated BMD loci in 4,983 individuals from the prospective population-based Rotterdam study. Risk allele counts were derived from the top associated SNPs from the 15 LS-BMD and 10 FN-BMD loci, all of which followed a normal distribution of their frequency in the study population (Figure 5). The 15 LS-SNPs combined explained ~2.9% of the variance in LS-BMD and the 10 FN-SNPs combined explained ~1.9% of the variance in FN-BMD. A highly significant linear decrease in the mean LS-BMD and FN-BMD of individuals was seen with increasing numbers of 'low BMD' risk alleles: those carrying 20 or more alleles that associated with low LS-BMD ('low BMD' alleles, n = 245) had 0.74 s.d (~0.13 g/cm2) lower BMD ($P = 2 \times 10^{-18}$) than those carrying 11 or fewer 'low-BMD' alleles (n = 291). A similar (yet less pronounced) trend was seen for FN-BMD. The association between the compound allelic scores and the risk of fracture were assessed in 2727 radiographically screened individuals (302 vertebral fracture cases) and in 4865 individuals followed-up 8.2 years on average (672 non-vertebral fractures). The compound FN-BMD allelic score was significantly associated with the risk of incident nonvertebral fracture in the Rotterdam Study dataset (HR = 1.042, 95% CI [1.003, 1.084]; P = 0.04), whereas it was borderline significant for association with the risk of vertebral fracture (OR = 1.061, 95% CI [0.997, 1.129]; P=0.06). In contrast, the compound LS-BMD allelic score was significantly associated with the risk of vertebral fracture (OR = 1.061, 95% CI [1.009, 1.116]; P = 0.02),

whereas it was not significant for association with the risk of incident nonvertebral fracture (HR = 1.025, 95% CI [0.993, 1.058]; P = 0.13). Adjustment for BMD showed that at least 54% of the genetic effect on incident nonvertebralfracture could be explained by FN-BMD ($HR_{adjusted}$ = 1.019, 95% CI [0.980,1.060]; P = 0.35), whereas 78% of the genetic effect on vertebral fracture was through LS-BMD ($OR_{adjusted}$ = 1.013, 95 %CI [0.962,1.068]; P = 0.62). Despite power limitations, consistent significant associations between the compound BMD allelic scores were observed with the risk of both vertebral and nonvertebral fracture.

DISCUSSION

The GEFOS consortium has been assembled to identify the genetic determinants of osteoporosis and fracture. This study represents the first step in a collaborative effort and expands the current knowledge on the underlying genetics of BMD -a clinical measurement used for the diagnosis of osteoporosis and the assessment of fracture risk. BMD measures of the lumbar spine and femoral neck were analyzed independently, because despite a relatively high phenotypic correlation, the genetic correlation is not perfect. This is illustrated by the site-specificity detected in some of the associations probably reflecting genuine biological mechanisms (i.e. differences in cortical vs trabecular bone content), but also intrinsic measurement differences (i.e. artifacts influencing BMD values like osteophytes of the lumbar spine or aortic

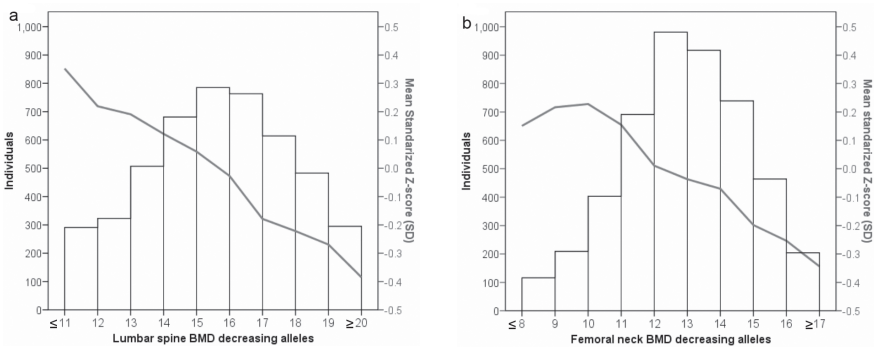


Figure 5. Histogram and line plot modeling in the Rotterdam Study the combined allelic effect across all genome-wide significant associated loci for a) Lumbar Spine BMD and b) Femoral neck BMD. Subjects were classified according to the number of BMD decreasing (risk) alleles at the lumbar spine BMD (20 SNPs) and the femoral neck (15 SNPs). The mean for each risk allele-count group was determined and the extremes of the distribution counts were pooled into the nearest risk allele-count group of size >100 individuals.

calcifications).

Performing meta-analysis of GWAS has limitations. False positive associations generated from multiple hypothesis testing and population stratification are inherent possibilities. Nevertheless, we applied well-established methods to minimize the impact of multiple testing by applying stringent GWS thresholds for determining significance. Also, our findings are constrained by the power of our current sample suited to identify effect sizes explaining ~0.2% of the variance of the trait. This means we are not powered to detect real effects of the same (small) magnitude arising within specific sex and/or age groups. Similarly, due to power limitations we cannot address potential gene-gene and gene-environment interactions, the effects of rare alleles that are not captured by the haplotype tagging approach employed in GWAS nor determine the effect on fracture risk. Despite being underpowered to assess heterogeneity of effects, some markers displayed significant heterogeneity and were not GWS when analyzed under the random effects model. Further evaluation of such markers in larger populations is needed to determine the source(s) of heterogeneity across datasets, like inaccuracies in the imputations, subtle differences in phenotype ascertainment, differences in linkage to the culprit, differences in environmental modifying factors, or genuinely different genetic effects across populations.[47] Achieving such sufficiently-powered setting is the target of the expanding GEFOS consortium. All our top associated markers displayed high quality imputation scores, high correlation after de-novo genotyping and/or at least one genotyped proxy in complete LD which was also associated at GWS level (Supplementary Table 10). Thus, we can exclude imputation inaccuracies as a source of heterogeneity and/or false-positive associations. In addition, we excluded individuals with non-European profiles strategy which confines our current findings to the context of populations of Northern European-descent. Similarly, test statistics corrected by the genomic inflation factors affecting each study, makes unlikely that population stratification or cryptic relatedness (like those observed in the ERF, TUK, and dCG populations) play an important role in our associations. In addition, we examined the effect of applying a second correction for overall meta-analysis genomic control in our results. Only one locus (*LRP5*) drifted away importantly from the GWS threshold after correction. This approach is likely to be over-conservative considering that the association of variants in LRP5 has been consistently replicated at GWS level in at least two previous efforts[10,24].

In summary, we identified and/or confirmed at least 20 loci associated with lumbar spine and femoral neck BMD, highlighting the complex genetic architecture underlying the variation in BMD. Yet, these loci explain only a minor fraction of the

variance in BMD, and hence, an even smaller fraction of the heritability for fracture risk. Nevertheless, these findings underscore molecules within novel and key-known biological pathways influencing BMD variation, particularly the *Wnt* and *NF-kappa-B* signaling pathways (including about half of the identified loci GPR177, *CTNNB1, FOXC2, LRP5, SPTBN1, HDAC5, TNFSF11, TNFRSF11A* and *TNFRSF11B)*. None of the SNPs we have identified can be unequivocally designated as the underlying "true" variants driving the associations. Additional efforts to identify such variants are warranted to maximize the application of this genetic knowledge towards the prediction of risk in individuals and the translation into new pharmacological agents for the treatment of osteoporosis. Increasing further the sample size will aid the identification of additional loci associated not only with BMD, but more importantly will allow focusing on the risk of fracture, the ultimate consequence of osteoporosis.

METHODS

Study population. The GEnetic Factors of Osteoporosis (GEFOS) consortium is a coalition of teams of investigators working on the genetics of osteoporosis. The current meta-analysis incorporated 19,195 individuals of Northern European ancestry derived from five GWAS on BMD of the lumbar spine (LS-BMD) and the femoral neck (FN-BMD) including: the *Rotterdam Study* (RS, n=4,987), *Erasmus Rucphen Family Study* (ERF, n=1,228), *Twins UK Study* (TUK, n=2,734), *deCODE Genetics Study* (dCG, n=6,743) and the *Framingham Osteoporosis Study* (FOS, n=3503). All studies were approved by institutional ethics review committees at the relevant organizations and all participants provided written informed consent. The Rotterdam Study (RS) is a prospective population-based cohort study of chronic disabling conditions in Dutch elderly individuals age 55 years and over (http://www.epib.nl/ergo.htm).[50,51] The Erasmus Rucphen Family (ERF) study is a family-based study of a genetic isolate in the South West Netherlands to identify genetic risk factors for complex disorders.[52] The Twins UK (TUK) study is a population-based sample of twins from the UK studying the hereditary basis of a wide variety of age-related traits and diseases (http:/www.twinsUK.ac.uk).[53] The Icelandic deCODE Genetics (dCG) study comprises a population-based sample to identify the genetic basis of complex diseases.[13] The Framingham Osteoporosis Study (FOS) is embedded in the Framingham Heart Study (FHS), a community-based, longitudinal, prospective cohort comprising three generations of individuals in multigenerational pedigrees and additional unrelated individuals (http://www.framinghamheartstudy.org/). Individuals of "Generation 1" include those first examined in 1948[54], "Generation 2" includes those examined at

the first cycle from 1971 to 1975[55], and "Generation 3" includes those examined at the first cycle beginning in 2002-2005.[56] For these analyses, 812 members of the Generation 1 cohort (22.6% of the sample) and 2783 (77.4%) of the Generation 2 cohort who had BMD measured as part of the FOS were included.

Bone mineral density and anthropometric measurements. BMD was measured in all cohorts at the lumbar spine (either at L1–L4 or L2-L4) and femoral neck using dual-energy X-ray absorptiometry following standard manufacturer protocols (GE-Lunar Corporation, Madison, WI or Hologic Incorporated, Bedford, MA) see Supplementary Table 1 for details. All DXA and anthropometric measurements were performed in the RS at the baseline visit baseline between 1991-1992, in ERF between 2002-2003, in TUK at the latest follow-up, dCG at the baseline visit and in FOS Generation 1 between 1992-1997, and Generation 2 between 1996-2001.

Phenotype modeling. The overall strategy involved linear regression to adjust BMD measurements for effects of age, weight, sex and study using standardized residuals with mean 0 and standard deviation 1 in the genotype-phenotype association testing. Such residuals were obtained by regressing within each study the raw BMD measurements on age and weight (and principal components in FOS to adjust for population substructure using Eigenstrat[57]) in sex-specific models. Thus, in studies including both men and women the data for each gender are included as separate estimates in the meta-analysis.

Genotyping. The five GWAS were genotyped using the Illumina Infinium HumanHap550 Beadchip (RS), the Illumina Infinium HumanHap300 or HumanCNV370 Beadchip (ERF, TUK & dCG) or the Affymetrix Dual NspI/StyI GeneChip 2x250K with 50K gene-centered MIP set (FOS), all according to manufacturer's protocols and quality control standards. The exclusion/filtering criteria for individuals and SNPs are described in Supplementary Tables 2-3.

Genotype imputation. Imputation was used to evaluate associations for the same SNPs across study populations using scans from different genotyping platforms. Genotypes were imputed for all polymorphic SNPs oriented to the positive strand from phased autosomal chromosomes of the HapMap CEU Phase II panel (release 22, build 36)[58]. Hidden Markov Model-based algorithms were used to infer unobserved genotypes probabilistically as implemented in either MACH[59] or IMPUTE[60]. Imputation quality control metrics from MACH and IMPUTE were used. Detailed descriptions of quality control and imputation procedures are summarized for all studies in Supplementary Table 3. We performed technical validation of the imputed genotypes in an independent set of 880 individuals of Icelandic origin using Centaurus (Nanogen)[61] discrimination assays, for the top associated hits that

reached GWS for the first time in this study (Supplementary Table 10).

Genotype-phenotype association testing[62]. Each study performed genome-wide association for BMD using sex-specific, age- and weight-adjusted standardized residuals analyzed under an additive (per allele) genetic model. Analysis of imputed genotype data accounted for uncertainty in each genotype prediction by utilizing either the dosage information from MACH or the genotype probabilities from IMPUTE. Studies used MACH2QTL[59], which uses genotype dosage value (0 – 2, continuous) as a predictor in a linear regression framework, SNPTEST[60], Merlin[63] or the linear mixed effects model of the Kinship[64] and ProABEL[65] packages in R[66] to account for relatedness (Supplementary Table 3). The genomic control method[20] was used to correct the standard error by the square root of the genomic inflation factor (lambda): $SE_{corrected} = SE * \sqrt{\lambda}$, which is equivalent to the proposed correction of the Chi^2-statistic by lambda. Genomic inflation factors for the studies are presented on Supplementary Table 3. Overall meta-analysis genomic control inflation factors were calculated as described previously.[67] Genomic inflation factors scaled to a standard size (1000 individuals) to calibrate for the effect of sample size on λ[67], showed residual genomic inflation was negligible ($\lambda_{LSBMD1000}$=1.005 and $\lambda_{FN\text{-}BMD1000}$=1.004).

Meta-analysis. The minor allele from HapMap CEU genotypes was used to define the coded allele in all analyses, regardless of frequency in individual cohorts. All meta-analysis calculations were done using the METAL[21] software package applying inverse-variance methodology assuming fixed effects with Cochran's Q and I^2 metrics used to quantify between-study heterogeneity. We also calculated the summary results by random effects using STATA software[68] for those markers associated at GWS level with Cochran's Q P-value < 0.05 and/or I^2 estimates > 50%, reflecting large heterogeneity beyond chance. Random effects models also incorporate in the calculations the between-study heterogeneity, and estimate the average genetic effect from the population of genetic effects that may differ in different studies. In the absence of between-study heterogeneity fixed and random effects calculations give identical results. We declared results genome-wide significant at $\alpha=5x10^{-8}$ after adjusting for all common variant tests in the human genome[22,23]. To test for BMD site specificity we estimated the effect difference ($\Delta\beta$) as $\beta_{femoralneck} - \beta_{lumbarspine}$, the SE of the mean difference (ΔSEM) as sqrt($SE_{femoralneck}^2 + SE_{lumbarspine}^2$) and the Z-statistic as $\Delta\beta/\Delta SEM$ from which the P-values were computed.

eQTL analysis in human osteoblast (HOb). SNPs from new loci associated with BMD at the genome-wide significance (GWS) level were tested for association with *cis*-allelic expression of neighboring gene transcripts, in primary human osteoblasts (HOb) derived from 95 unrelated Swedish donors. Detailed cell culture methods

have been described previously[69]. Expression profiling was performed using the Illumina HumRef-8 BeadChips according to the manufacturer protocol. Genotyping for genotype-expression association was performed using the Illumina HumanHap 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. The association of the expression levels was focused on *cis*-acting genetic variants, defined as being within 250kb window flanking the gene, using a linear regression model implemented in the PLINK[70] software package with age and sex as covariates. SNPs included on the Illumina 550K chip were assessed for expression *cis*-associations directly. In addition, all genotyped SNPs included on the Illumina chip that were in strong LD (defined as D´≥0.8) and mapping ± 50kb from the GWS hit were included in the association study. To test for a significant enrichment of functional SNPs (i.e., SNPs associated with gene expression in HObs at p<0.05) among the candidate SNPs, a Chi^2- statistic was obtained to test whether observed associations were different from expected associations in the expression data set beyond chance. Expected values (7.1%) were based on the proportion of SNPs with p<0.05 seen in the HOb gene expression data set and in a random selection of 1200 SNPs associated with both LS- and FN-BMD at P>0.90, MAF>0.20, and present in the Illumina HumanHap550 array (assessed as negative controls for association with HOb expression).

Combined effect of associated loci. Within the setting of the prospective population-based Rotterdam Study the combined effect of all 20 BMD loci was studied by classifying subjects according to the number of BMD decreasing (risk) alleles. This was based on 15 lumbar spine and 10 femoral neck BMD loci as follows: Lumbar spine (15 SNPs) => rs7524102 [*ZBTB40*], rs1430742 [*GPR177*], rs11898505 [*SPTBN1*], rs1471403 [*MEPE*], rs2504063 [*ESR1*], rs1524058 [*STARD3NL*], rs4729260 [*FLJ42280*], rs2062377 [*TNFRSF11B*], rs16921914 [*DCDC5*], rs599083 [*LRP5*], rs2016266 [*SP7*], rs9533090 [*TNFSF11*], rs10048146 [*FOXC2*], rs9303521 [*CRHR1*] and rs884205 [*TNFRSF11A*]. Femoral neck (10 SNPs) => (rs6426749 [*ZBTB40*], rs2566755 [*GPR177*], rs87938 [*CTNNB1*], rs1366594 [*MEF2C*], rs2941740 [*ESR1*], rs7781370 [*FLJ42280*], rs11995824 [*TNFRSF11B*], rs7117858 [*SOX6*], rs7932354 [*ARHGAP1*], and rs228769 [*HDAC5*]). The mean BMD for each risk allele-count group was determined, and at the extremes of the distribution counts were pooled into the nearest risk allele-count group of size >100 individuals. The approximate BMD difference in g/cm^2 was obtained by multiplying in each group the mean Z-score LS-BMD by 0.18 g/cm^2 and the mean Z-score FN-BMD by 0.13 g/cm^2 (the SDs of BMD in the Rotterdam study).

The allele score was obtained by dividing the number of "BMD decreasing alleles" by the total number of alleles. Also within the setting of the prospective population-based Rotterdam Study**,** we determined the risk for vertebral and non-vertebral fracture for the combined allelic scores constructed for all the top hits associated at a genome-wide significantly level with BMD. Risk ratio (RR) estimates were obtained from logistic regression (vertebral fractures) and Cox-proportional hazards (incident non-vertebral fractures) models adjusted for sex, age and weight. To determine the fraction of fracture risk explained by BMD, we applied the following formula: $[\ln RR_{unadjusted} - \ln RR_{BMDadjusted}] / \ln RR_{unadjusted}$. Methods describing the fracture datasets have been published previously[71,72]. In summary, thoracolumbar radiographs of the spine were obtained in 3308 (genotyped) individuals who survived on average 6.4 ± 0.4 (SD) years and were scored for presence of vertebral fractures (n=329) using the McCloskey/Kanis method[73]. Record of the incident non-vertebral fractures (n=900) occurring between the baseline visit from 1990 through 1993 until January 1, 2002, was obtained from the computerized records of general practitioners and hospital registries for 5974 genotyped individuals followed on average 8.2 ± 2.7 (SD) years after the baseline visit.

ACKNOWLEDGEMENTS

Above all, we thank all study participants for making this work possible. This research and the GEnetic Factors of Osteoporosis (GEFOS) consortium (http://www.gefos.org) have been funded by the European Commission (HEALTH-F2-2008-201865-GEFOS). Rotterdam Study (RS): This study was funded by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the staff from the Rotterdam Study, particularly Lydia Buist and J. Hannie van den Boogert and also with the participating general practitioners and pharmacists. Erasmus Rucphen Family (ERF): The study was supported by grants from The Netherlands Organization for Scientific Research (NWO), Erasmus MC and the Centre for Medical Systems Biology (CMSB). We are grateful to all general practitioners for their contributions, to Petra Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and Peter Snijders for his help in data collection. Twins UK (TUK): The study was funded by the Wellcome Trust; the Arthritis Research Campaign; the Chronic Disease Research Foundation; the Canadian Institutes of Health Research (JBR); European Society for Clinical and Economic Aspects of Osteoporosis (JBR); the European Union FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. We thank the staff from the TwinsUK, the DNA Collections and Genotyping Facilities at the Wellcome Trust

Sanger Institute for sample preparation; Quality Control of the Twins UK cohort for genotyping (in particular Amy Chaney, Radhi Ravindrarajah, Douglas Simpkin, Cliff Hinds, and Thomas Dibling); Paul Martin and Simon Potter of the DNA and Genotyping Informatics teams for data handling; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie Icelandic deCODE Study (dCG): We thank the staff of the deCODE core facilities and recruitment centre for their important contributions to this work. Framingham Osteoporosis Study (FOS): The study was funded by grants from the National Institute for Arthritis, Musculoskeletal and Skin Diseases and the National Institute on Aging (R01 AR/AG 41398; DPK and R01 AR 050066; DK). The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine were supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. eQTL HOb Study: The study was supported by Genome Quebec, Genome Canada, and the Canadian Institutes of Health Research (CIHR). T.P. holds a Canada Research Chair. We thank Profs Olof Nilsson, Hans Mallmin and Östen Ljunggren at the Departments of Surgical and Medical Sciences, Uppsala University Hospital, Sweden for large-scale collection of primary bone samples.

AUTHOR CONTRIBUTIONS

FR, KE, BVH, FKK, JPAI, ran meta-analysis; FR, KE, BVH, YHH, JBR, MCZ, NA, YAA, LAC, SD, NS, GT, YZ, ran statistical analysis in studies; FR, US, PD, JBvM, UT, AGU, coordinated GWA genotyping of studies; EG, TP, did expression studies; FR, US, MCZ, AH, BO, HAPP, GS, GT, FMKW, SGW, CMvD, TS, DPK, AGU, coordinated/collected phenotype information of studies; US, LAC, AH, AK, DK, BO, HAPP, UT, CMvD, TS, DPK, KS, AGU, designed studies; FR, US, JBvM, TS, UT, SHR, JPAI, AGU, established consortium and UT, SHR, JPAI, AGU obtained funding; all authors interpreted results; all authors read critically the manuscript; FR, wrote manuscript draft.

Note: Supplementary information is available on the Nature Genetics website**:**
http://www.nature.com/ng/journal/v41/n11/extref/ng.446-S1.pdf

REFERENCES

1. Ellies, D.L. et al. Bone density ligand, Sclerostin, directly interacts with LRP5 but not LRP5G171V to modulate Wnt activity. *J Bone Miner Res* **21**, 1738-49 (2006).
2. Theoleyre, S. et al. The molecular triad Opg/RANK/RANKL: involvement in the orchestration of pathophysiological bone remodeling. *Cytokine Growth Factor Rev* **15**, 457-75 (2004).
3. Ioannidis, J.P. et al. Meta-analysis of genome-wide scans provides evidence for sex- and site-specific regulation of bone mass. *J Bone Miner Res* **22**, 173-83 (2007).

4. Ioannidis, J.P. Why most published research findings are false. *PLoS Med* **2**, e124 (2005).

5. Ioannidis, J.P. Genetic associations: false or true? *Trends Mol Med* **9**, 135-8 (2003).

6. Ioannidis, J.P. et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *Jama* **292**, 2105-14 (2004).

7. Langdahl, B.L. et al. Large-scale analysis of association between polymorphisms in the transforming growth factor beta 1 gene (TGFB1) and osteoporosis: the GENOMOS study. *Bone* **42**, 969-81 (2008).

8. Ralston, S.H. et al. Large-scale evidence for the effect of the COLIA1 Sp1 polymorphism on osteoporosis outcomes: the GENOMOS study. *PLoS Med* **3**, e90 (2006).

9. Uitterlinden, A.G. et al. The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. *Ann Intern Med* **145**, 255-64 (2006).

10. van Meurs, J.B. et al. Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. *Jama* **299**, 1277-90 (2008).

11. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).

12. Styrkarsdottir, U. et al. New sequence variants associated with bone mineral density. *Nat Genet* **41**, 15-7 (2009).

13. Styrkarsdottir, U. et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* **358**, 2355-65 (2008).

14. Kanis, J.A., Delmas, P., Burckhardt, P., Cooper, C. & Torgerson, D. Guidelines for diagnosis and management of osteoporosis. The European Foundation for Osteoporosis and Bone Disease. *Osteoporos Int* **7**, 390-406 (1997).

15. Blake, G.M., Knapp, K.M., Spector, T.D. & Fogelman, I. Predicting the risk of fracture at any site in the skeleton: are all bone mineral density measurement sites equally effective? *Calcif Tissue Int* **78**, 9-17 (2006).

16. Peacock, M., Turner, C.H., Econs, M.J. & Foroud, T. Genetics of osteoporosis. *Endocr Rev* **23**, 303-26 (2002).

17. Weedon, M.N. et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575-83 (2008).

18. Gudbjartsson, D.F. et al. Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**, 609-15 (2008).

19. Visscher, P.M. Sizing up human height variation. *Nat Genet* **40**, 489-90 (2008).

20. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

21. Abecasis, G.R. & Willer, C. Metal: Meta Analysis Helper. 1.0 14/4/2008 edn (2008).

22. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).

23. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).

24. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).

25. Banziger, C. et al. Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125**, 509-22 (2006).

26. Matsuda, A. et al. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene* **22**, 3307-18 (2003).

27. Glass, D.A., 2nd et al. Canonical Wnt signaling in differentiated osteoblasts controls osteoclast differentiation. *Dev Cell* **8**, 751-64 (2005).

28. Potthoff, M.J. et al. Histone deacetylase degradation and MEF2 activation promote the formation of slow-twitch myofibers. *J Clin Invest* **117**, 2459-67 (2007).

29. Cho, Y.S. et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* **41**, 527-34 (2009).

30. Smits, P. et al. The Transcription Factors L-Sox5 and Sox6 Are Essential for Cartilage Formation. **1**, 277-290 (2001).

31. Zhou, G. et al. Dominance of SOX9 function over RUNX2 during skeletogenesis. *Proc Natl Acad Sci U S A* **103**, 19004-9 (2006).

32. Reiner, O. et al. The evolving doublecortin (DCX) superfamily. *BMC Genomics* **7**, 188 (2006).

33. Kim, S.H. et al. The forkhead transcription factor Foxc2 stimulates osteoblast differentiation. *Biochem Biophys Res Commun* (2009).

34. Nifuji, A., Miura, N., Kato, N., Kellermann, O. & Noda, M. Bone morphogenetic protein regulation of forkhead/winged helix transcription factor Foxc2 (Mfh1) in a murine mesodermal cell line C1 and in skeletal precursor cells. *J Bone Miner Res* **16**, 1765-71 (2001).

35. Winnier, G.E., Hargett, L. & Hogan, B.L. The winged helix transcription factor MFH1 is required for proliferation and patterning of paraxial mesoderm in the mouse embryo. *Genes Dev* **11**, 926-40 (1997).

36. Stankiewicz, P. et al. Genomic and genic deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of FOXF1 cause alveolar capillary dysplasia and other malformations. *Am J Hum Genet* **84**, 780-91 (2009).

37. Tang, Y. et al. Disruption of transforming growth factor-beta signaling in ELF beta-spectrin-deficient mice. *Science* **299**, 574-7 (2003).

38. Alford, A.I. & Hankenson, K.D. Matricellular proteins: Extracellular modulators of bone development, remodeling, and regeneration. *Bone* **38**, 749-57 (2006).

39. Gowen, L.C. et al. Targeted disruption of the osteoblast/osteocyte factor 45 gene (OF45) results in increased bone formation and bone mass. *J Biol Chem* **278**, 1998-2007 (2003).

40. Malaval, L. et al. Bone sialoprotein plays a functional role in bone formation and osteoclastogenesis. *J Exp Med* **205**, 1145-53 (2008).

41. Yoshitake, H., Rittling, S.R., Denhardt, D.T. & Noda, M. Osteopontin-deficient mice are resistant to ovariectomy-induced bone resorption. *Proc Natl Acad Sci U S A* **96**, 8156-60 (1999).

42. Meyers, V.E., Zayzafoon, M., Douglas, J.T. & McDonald, J.M. RhoA and cytoskeletal disruption mediate reduced osteoblastogenesis and enhanced adipogenesis of human mesenchymal stem cells in modeled microgravity. *J Bone Miner Res* **20**, 1858-66 (2005).

43. Wang, L., Yang, L., Debidda, M., Witte, D. & Zheng, Y. Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. *Proc Natl Acad Sci U S A* **104**, 1248-53 (2007).

44. McKinsey, T.A., Zhang, C.L., Lu, J. & Olson, E.N. Signal-dependent nuclear export of a

histone deacetylase regulates muscle differentiation. *Nature* **408**, 106-11 (2000).

45. Schroeder, T.M. & Westendorf, J.J. Histone deacetylase inhibitors promote osteoblast maturation. *J Bone Miner Res* **20**, 2254-63 (2005).
46. Kang, J.S., Alliston, T., Delston, R. & Derynck, R. Repression of Runx2 function by TGF-beta through recruitment of class II histone deacetylases by Smad3. *Embo J* **24**, 2543-55 (2005).
47. Ioannidis, J.P., Thomas, G. & Daly, M.J. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* **10**, 318-29 (2009).
48. Online Mendelian Inheritance in Man, OMIM (TM). (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD),).
49. MCKussic, V.A. *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders.* , (Johns Hopkins University Press, Baltimore, 1998).
50. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
51. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* **7**, 403-22. (1991).
52. Aulchenko, Y.S. et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
53. Arden, N.K., Baker, J., Hogg, C., Baan, K. & Spector, T.D. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. *J Bone Miner Res* **11**, 530-4 (1996).
54. Dawber, T.R., Kannel, W.B. & Lyell, L.P. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* **107**, 539-56 (1963).
55. Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J. & Castelli, W.P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* **110**, 281-90 (1979).
56. Splansky, G.L. et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* **165**, 1328-35 (2007).
57. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
58. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
59. Li, Y. & Abecasis, G.R. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*, 2290 (2006).
60. 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* **39**, 906-13 (2007).
61. Kutyavin, I.V. et al. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res* **34**, e128 (2006).
62. Little, J. et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Eur J Epidemiol* **24**, 37-55 (2009).
63. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
64. Therneau, T. Kinship: mixed effectss Cox models, sparce matrices, and modelling

data from large pedigrees. R package version 1.1.0-19 edn (2008).

65. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).
66. R Developmental Core Team. A language and environment for statistical computing. (R Foundation for Statistical Computing, 2007).
67. de Bakker, P.I. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
68. StataCorp. Stata Statistical Software. Release 10 edn (StataCorp LP, College Station, TX, 2007).
69. Grundberg, E. et al. Systematic assessment of the human osteoblast transcriptome in resting and induced primary cells. *Physiol Genomics* **33**, 301-11 (2008).
70. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
71. Rivadeneira, F. et al. Estrogen receptor beta (ESR2) polymorphisms in interaction with estrogen receptor alpha (ESR1) and insulin-like growth factor I (IGF1) variants influence the risk of fracture in postmenopausal women. *J Bone Miner Res* **21**, 1443-56 (2006).
72. Schuit, S.C. et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone* **34**, 195-202 (2004).
73. McCloskey, E.V. et al. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporos Int* **3**, 138-47 (1993).

CHAPTER 4.2

Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture

Nat Genet , advance online publication, 15 April 2012; doi:10.1038/ng.2249.

ABSTRACT

Bone mineral density (BMD) is the most widely used predictor of fracture risk. We performed the largest meta-analysis to date on lumbar spine and femoral neck BMD, including 17 genome-wide association studies and 32,961 individuals of European and East Asian ancestry. We tested the top-associated BMD markers for replication in 50,933 independent subjects and for risk of low-trauma fracture in 31,016 fracture (cases) and 102,444 controls. We identified 56 loci (32 novel) associated with BMD at genome-wide significant level ($P<5x10^{-8}$). Several of these factors cluster within the RANK-RANKL-OPG, mesenchymal-stem-cell differentiation, endochondral ossification and the Wnt signalling pathways. However, we also discovered loci containing genes not known to play a role in bone biology. Fourteen BMD loci were also associated with fracture risk ($P<5x10^{-4}$, Bonferroni corrected), of which six reached $P<5x10^{-8}$ including: 18p11.21 (*FAM210A*), 7q21.3 (*SLC25A13*), 11q13.2 (*LRP5)*, 4q22.1 (*MEPE*), 2p16.2 (*SPTBN1*) and 10q21.1 (*DKK1*). These findings shed light on the genetic architecture and pathophysiological mechanisms underlying BMD variation and fracture susceptibility.

INTRODUCTION

Osteoporosis is a disease characterized by low bone mass and microarchitectural deterioration of bone tissue leading to increased risk of fracture. The disease accounts for approximately 1.5 million new fracture cases each year representing a huge economic burden on health care systems, with annual costs estimated to be $17 billion in the USA alone and expected to rise 50% by the year 2025.[1] Osteoporosis is defined clinically through the measurement of bone mineral density (BMD), which remains the single best predictor of fracture.[2,3]

Twin and family studies have shown that 50%-85% of the variance in BMD is genetically determined.[4] Osteoporotic fractures are also heritable by mechanisms that are partly independent of BMD.[5] Over the past 5 years, genome-wide association studies (GWAS) have revolutionized the understanding of the genetic architecture of common, complex diseases.[6] This strategy is providing key insights into the mechanisms of disease with prospects of designing effective strategies for risk assessment and development of new interventions.[7]

Previous GWAS have identified to-date 24 loci which influence BMD variation.[8-14] While several variants in these BMD loci have also been nominally-associated with fracture risk[15,16], none have shown robust association at genome-wide significant levels ($P<5x10^{-8}$). We report here the results of the largest effort to date searching for BMD loci in >80,000 subjects and testing them for association with fracture in >130,000 cases and controls. In addition, we employed bioinformatics tools and gene expression analyses to place the identified variants in the context of pathways relevant to bone biology.

RESULTS

This study was performed across three main stages (Fig. 1): 1) discovery of BMD loci, 2) follow-up replication and 3) association of the BMD loci with fracture.

Discovery of BMD Loci (Stage 1)

We first performed a meta-analysis of GWAS for BMD of the femoral neck (FN-BMD; n=32,961) and lumbar spine (LS-BMD; n=31,800) including ~2.5 million autosomal genotyped or imputed SNPs from 17 studies from populations across North America, Europe, East Asia and Australia, with a variety of epidemiological designs and patient characteristics (see Online Methods). We also performed meta-

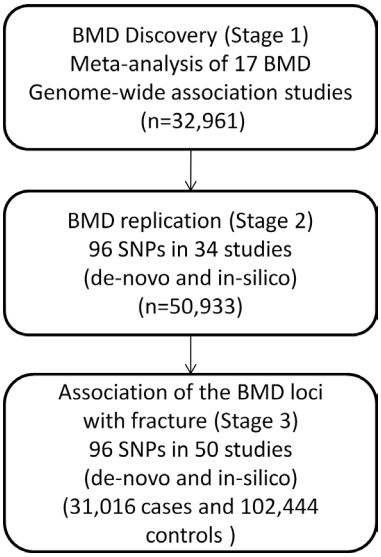


Figure 1. Description of Study design. Stage 1: Meta-analysis of 17 Genome-wide association studies for BMD. **Stage 2**: 96 top independent SNPs (82 autosomal SNPs with $P<5x10^{-6}$, 5 X-chromosome SNPs and 9 SNPs from conditional analysis) were followed-up for *de-novo* and *in-silico* replication of the BMD association in 34 studies. **Stage 3**: the same 96 SNPs were tested for association with fracture in 50 studies with *de-novo* and *in-silico* data.

analysis in men and women separately to identify sex-specific associations. The quantile-quantile (Q-Q) plots of the discovery meta-analysis displayed strong (and not early) deviation of the observed statistics from the null distribution for both BMD traits (Supplementary Fig. 1). After double Genomic Control (GC) correction of the overall (λ_{FN_POOLED}=1.112; λ_{LS_POOLED}=1.127) and sex-stratified analyses (λ_{FN_WOMEN}=1.091; λ_{FN_MEN}=1.059; λ_{LS_WOMEN}=1.086; λ_{LS_MEN}=1.061), SNPs in 34 loci surpassed GWS level while a total of 82 loci were associated at $P<5x10^{-6}$ (Supplementary Fig. 2-3). Thirty eight loci were associated with FN-BMD, 25 with LS-BMD and 19 with both. The overlap reflects the correlation between the femoral neck and lumbar spine measurements (Pearson correlation = 0.53). Of these 82 loci, 59, 18 and 5 were prioritized from the analysis in the sex-combined, women and men sample sets, respectively (Supplementary Table 1). The meta-analysis was extended to include the evaluation of 76,253 X-chromosome imputed markers across 14 of the discovery GWAS including 31,801 participants (see Online Methods). Five X-chromosome loci were associated at $P<5x10^{-5}$ four of which were derived from the sex-combined analysis and one from the analysis in men only (Supplementary Table 1). We further performed genome-wide conditional analyses in all sex-combined stage 1 studies.

Each study repeated the GWAS analysis but additionally adjusted for 82 SNPs representing the autosomal loci associated at $P<5x10^{-6}$ (see Online Methods). We then meta-analyzed these studies in the same way as for the primary GWA study meta-analysis. Nine loci showed at least two independent association signals arising from this conditional analysis (Supplementary Fig. 4 and Supplementary Table 2) suggesting that allelic heterogeneity is underlying BMD variation. We also assessed all possible pair-wise interactions of the 82 SNPs, but none were significant after adjusting for the number of tests (Supplementary Fig. 5 and Supplementary Table 3). A total of 96 independent SNPs (82 autosomal SNPs with $P<5x10^{-6}$ + 9 SNPs from conditional analysis and 5 X-chromosome SNPS) from 87 genomic loci were selected for further replication (Fig. 1).

Follow-up replication (Stage 2)

We de-novo genotyped these 96 SNPs and tested them for association with BMD in up to 50,933 additional participants from 34 studies (see Online Methods).The meta-analysis of the 96 SNPs in the discovery and replication studies (n=83,894) yielded 64 replicating SNPs from 56 associated loci. Of these loci, 32 were novel (Table 1 and Supplementary Table 4A) and 24 were reported previously[8-14] (Supplementary Table 4B). Thirty two SNPs did not reach genome-wide significance after replication (Supplementary Table 4C) and included 10 markers remaining associated at a suggestive level. Of all analyzed SNPs only one (rs9533090 mapping to 13q14.11 near *RANKL*) displayed high degree of heterogeneity of effects ($I^2>50\%$) across studies, despite being the marker with highest significance ($P=4.82x10^{-68}$) in the fixed-effect meta-analysis (Supplementary Table 4B). After applying random effects meta-analysis, this marker was still genome-wide significant ($P=3.98x10^{-13}$).

Two of the novel loci were discovered in the sex-stratified meta-analysis: 8q13.3 in women and Xp22.31 in men; however, only the association at Xp22.31 showed significant evidence for sex-specificity as reflected by significant heterogeneity of effects across sex strata ($P_{het}=1.62x10^{-8}$). Yet, we acknowledge that the association at 8q13.3 in women may be driven by a lower number of men in the discovery and replication datasets (Table 1 and Supplementary Table 5). Furthermore, evidence for BMD site-specificity ($P_{het}<5x10^{-4}$) was observed in a fraction of the loci including 6 of the 32 novel and 4 of the 24 known loci (Table 1 and Supplementary Fig. 6). Among the newly identified loci, 2q14 (*INSIG2*), 12p11.22 (*PTHLH*) and 16q12.1 (*CYLD*) displayed site-specificity with FN-BMD while 8q13.3 (*LACTB2*), 10p11.23 (*MPP7*) and 10q22.3 (*KCNMA1*) displayed site-specificity with LS-BMD.

Table 1. Estimated effects of novel genome-wide significant SNPs on bone mineral density of femoral neck (FN-) BMD and lumbar spine (LS-) BMD across stages.

			Functional evidence							
SNP	Locus	Closest Gene / Candidate	eQTL	Knock-out mouse	OMIM	Tags funct.	GRAIL	Pathway	A[a]	Freq[b]
rs479336	1q24.3	*DNM3*							T	0.74
rs7584262	2p21	*PKDCC*		•					T	0.23
rs17040773	2q13	*ANAPC1*	•						A	0.76
rs1878526	2q14.2	*INSIG2*							A	0.22
rs1026364	3q13.2	*KIAA2018*				•			T	0.37
rs344081	3q25.31	*LEKR1*							T	0.87
rs3755955	4p16.3	*IDUA*		•		•			A	0.16
rs11755164	6p21.1	*SUPT3H/RUNX2*		•			•	•	T	0.40
rs9466056	6p22.3	*CDKAL1/SOX4*							A	0.38
rs3801387	7q31.31	*WNT16*				•	•	•	A	0.74
rs13245690[g]	7q31.31	*C7orf58*							A	0.62
rs7812088	7q36.1	*ABCF2*							A	0.13
rs7017914[e]	8q13.3	*XKR9/LACTB2*	•			•			A	0.49
rs7851693	9q34.11	*FUBP3*							C	0.64
rs3905706	10p11.23	*MPP7*							T	0.22
rs1373004	10q21.1	*MBL2/DKK1*		•				•	T	0.13
rs7071206	10q22.3	*KCNMA1*							T	0.78
rs7084921	10q24.2	*CPN1*							T	0.39
rs10835187	11p14.1	*LIN7C*	•						T	0.55
rs7953528	12p11.22	*KLHDC5/PTHLH*		•	•		•	•	A	0.18
rs2887571	12p13.33	*ERC1/WNT5B*					•	•	A	0.76
rs12821008	12q13.12	*DHH*						•	T	0.39
rs1053051	12q23.3	*C12orf23*	•						T	0.52
rs1286083	14q32.12	*RPS6KA5*							T	0.81
rs4985155	16p13.11	*NTAN1*	•			•			A	0.67
rs9921222	16p13.3	*AXIN1*		•			•	•	T	0.48
rs13336428	16p13.3	*C16orf38/CLCN7*		•	•				A	0.43
rs1566045	16q12.1	*SALL1/CYLD*		•					T	0.80
rs1564981[g]	16q12.1	*CYLD*		•					A	0.47
rs4790881	17p13.3	*SMG6*							A	0.69
rs7217932	17q24.3	*SOX9*		•	•		•	•	A	0.46
rs4796995	18p11.21	*FAM210A*							A	0.63
rs10416218	19q13.11	*GPATCH1*				•			T	0.73
rs5934507[f]	Xp22.31	*FAM9B/KAL1*							A	0.74

Boldface indicates $P < 5 \times 10^{-8}$ or Site-specificity $P<5x10^{-4}$. Black dots in the six functional evidence columns indicate that: first: SNP is an expression quantitative trait loci (eQTL); second: there is a knockout mouse with skeletal phenotypes (MGI database 2011); third: candidate gene has been involved in a monogenic syndrome with skeletal phenotypes (OMIM database 2011); fourth: most significant SNP tags a SNP predicted to have impact on function of the candidate gene; fifth: gene is best candidate in GRAIL analysis; and sixth: candidate gene is part of a bone active pathway. [a]The effect estimates (Beta) are expressed as standardized values per copy of the SNP allele (A) from fixed effects meta-analysis. [b]Freq is the allele frequency of A. [c]Effect estimates were calculated in the Stage 2 samples. [d]Site-specificity null hypothesis, i.e., $\beta_{LS} = \beta FN$. [e]rs7017914 was discovered in the meta-analysis of women-only, the effects and p-value for this marker are for the meta-analysis of women samples. [f]rs5934507 was discovered in the meta-analysis of men-only, the effects and p-value for this marker are for the meta-analysis of men samples. [g]The snps rs13245690 and rs1564981 were independently associated to their main signals in conditional analysis.

FN-BMD				LSBMD				
	STAGE 1 (32,961)	STAGE 2 (50,933)	STAGE 1 + STAGE 2 (83,894)		STAGE 1 (31,800)	STAGE 2 (45,708)	STAGE 1 + STAGE 2 (77,508)	
Beta[c]	P	P	P	Beta[c]	P	P	P	P_{het} Site[d]
-0.04	1.1×10^{-7}	**1.3×10^{-8}**	**8.5×10^{-15}**	-0.03	0.01	5.0×10^{-4}	2.1×10^{-5}	0.05
0.03	1.4×10^{-7}	3.4×10^{-4}	**1.3×10^{-9}**	0.01	0.13	0.28	0.07	0.01
0.03	4.3×10^{-6}	6.1×10^{-5}	**1.5×10^{-9}**	0.01	0.61	0.21	0.19	5.2×10^{-3}
0.00	0.70	0.97	0.79	0.04	7.3×10^{-6}	3.4×10^{-6}	**1.2×10^{-10}**	**8.6×10^{-5}**
0.03	2.0×10^{-6}	2.5×10^{-5}	**4.1×10^{-10}**	0.02	0.04	7.3×10^{-3}	7.6×10^{-4}	0.11
0.03	1.1×10^{-4}	2.5×10^{-3}	2.2×10^{-6}	0.06	2.8×10^{-5}	**3.5×10^{-8}**	**4.5×10^{-12}**	0.12
-0.05	3.9×10^{-7}	**6.1×10^{-9}**	**1.5×10^{-14}**	-0.05	1.4×10^{-7}	**5.5×10^{-9}**	**5.2×10^{-15}**	0.80
-0.01	0.23	0.12	0.05	-0.03	3.5×10^{-7}	9.2×10^{-6}	**5.6×10^{-11}**	2.1×10^{-3}
-0.03	**1.8×10^{-8}**	1.6×10^{-6}	**2.7×10^{-13}**	-0.03	6.5×10^{-5}	1.1×10^{-4}	**3.6×10^{-8}**	0.34
-0.08	**4.2×10^{-14}**	**2.0×10^{-27}**	**5.0×10^{-40}**	-0.10	**1.4×10^{-16}**	**1.5×10^{-36}**	**3.2×10^{-51}**	0.09
0.00	8.6×10^{-5}	0.69	8.2×10^{-4}	0.03	**1.1×10^{-9}**	1.3×10^{-3}	**6.0×10^{-11}**	0.05
0.04	1.2×10^{-6}	4.4×10^{-4}	**7.3×10^{-9}**	0.04	2.9×10^{-5}	1.1×10^{-3}	2.2×10^{-7}	0.86
0.02	**4.7×10^{-8}**	7.1×10^{-3}	**1.9×10^{-8}**	-0.01	0.35	0.41	0.98	**9.1×10^{-5}**
0.05	**3.1×10^{-8}**	**1.4×10^{-15}**	**3.4×10^{-22}**	0.04	0.06	6.7×10^{-8}	6.1×10^{-8}	0.02
-0.02	0.63	1.7×10^{-3}	0.03	0.05	**2.9×10^{-9}**	**6.7×10^{-9}**	**2.4×10^{-16}**	**5.8×10^{-11}**
-0.04	1.4×10^{-5}	1.5×10^{-4}	**1.5×10^{-8}**	-0.05	5.4×10^{-8}	2.2×10^{-6}	**1.6×10^{-12}**	0.28
0.01	0.29	0.26	0.81	-0.05	**1.5×10^{-12}**	**6.2×10^{-9}**	**5.0×10^{-19}**	**5.9×10^{-9}**
0.03	1.4×10^{-4}	1.6×10^{-6}	**9.0×10^{-10}**	0.03	0.01	1.9×10^{-5}	9.2×10^{-7}	0.58
-0.01	0.17	0.08	0.03	-0.02	3.0×10^{-5}	2.4×10^{-4}	**4.9×10^{-8}**	0.03
0.04	5.8×10^{-8}	2.4×10^{-6}	**1.9×10^{-12}**	-0.02	0.94	0.05	0.13	**2.3×10^{-7}**
-0.03	1.1×10^{-4}	1.6×10^{-5}	**6.5×10^{-9}**	-0.04	2.2×10^{-7}	2.9×10^{-6}	**5.6×10^{-12}**	0.37
0.03	1.9×10^{-4}	5.2×10^{-4}	3.3×10^{-7}	0.05	1.5×10^{-7}	**1.9×10^{-9}**	**1.2×10^{-15}**	0.06
-0.03	1.4×10^{-5}	1.8×10^{-5}	**9.6×10^{-10}**	-0.02	2.5×10^{-6}	2.4×10^{-3}	7.9×10^{-8}	0.76
-0.05	**2.9×10^{-8}**	**9.3×10^{-9}**	**2.0×10^{-15}**	-0.04	**1.7×10^{-11}**	7.1×10^{-6}	**1.8×10^{-14}**	0.92
-0.03	3.5×10^{-4}	1.4×10^{-7}	**1.7×10^{-10}**	-0.03	8.7×10^{-7}	1.8×10^{-4}	**2.2×10^{-9}**	0.98
-0.03	2.5×10^{-7}	2.4×10^{-6}	**5.2×10^{-12}**	-0.04	**2.2×10^{-8}**	**8.3×10^{-10}**	**1.0×10^{-16}**	0.26
-0.04	2.9×10^{-7}	**1.1×10^{-10}**	**1.5×10^{-16}**	-0.04	5.9×10^{-5}	**5.8×10^{-10}**	**1.7×10^{-13}**	0.80
-0.06	**5.0×10^{-12}**	**3.0×10^{-12}**	**1.9×10^{-22}**	-0.01	7.8×10^{-3}	0.55	0.04	**7.5×10^{-6}**
-0.02	1.1×10^{-3}	0.01	4.4×10^{-5}	-0.03	**6.2×10^{-8}**	5.4×10^{-4}	**2.0×10^{-10}**	0.50
0.05	**1.7×10^{-8}**	**1.2×10^{-11}**	**9.8×10^{-19}**	0.04	6.0×10^{-4}	1.7×10^{-6}	**3.4×10^{-9}**	0.13
0.03	**3.7×10^{-8}**	2.7×10^{-5}	**1.9×10^{-11}**	0.01	0.31	0.15	0.08	3.8×10^{-3}
0.02	3.2×10^{-6}	1.1×10^{-3}	**4.9×10^{-8}**	0.01	5.2×10^{-4}	0.11	6.7×10^{-4}	0.29
-0.02	5.7×10^{-6}	7.1×10^{-4}	5.5×10^{-8}	-0.03	**9.2×10^{-9}**	1.2×10^{-4}	**6.6×10^{-11}**	0.38
-0.08	0.01	8.3×10^{-4}	1.6×10^{-4}	-0.09	5.7×10^{-6}	3.2×10^{-4}	**1.2×10^{-8}**	0.17

After replication, the conditional analysis provided significant evidence of association ($P<5x10^{-8}$) in 8 of the 9 loci containing secondary signals (Supplementary Fig. 4 and Supplementary Table 2). Three loci included variants localized less than 40 Kb from the initial main signal suggesting allelic heterogeneity and included the 1p31.3 (represented by rs17482952 near *WLS*), 6q25.1 (rs7751941 near *ESR1*) and the 16q12.1 (rs1564981 near *CYLD*) loci. The secondary signal in 16q12.1 (rs1564981) showed a strong association with LS-BMD, while the main signal in this locus (rs1566045) was only associated with FN-BMD. The other five secondary signals were represented by variants localized at more than 180kb away from the initial main signal and contained different candidate genes including the 1p36.12 (rs7521902 near *WNT4*), 7p14.1 (rs10226308 near *SFRP4*), 7q31.31 (rs13245690 near *C7orf58*), 12q13.13 (rs736825 near *HOXC6*) and the 17q21.31 (rs4792909 near *SOST*) loci. The secondary signal mapping to the 13q14.11 locus (rs7326472) did not achieve genome-wide significance after replication.

Association of the BMD loci with fracture (Stage 3)

We tested the 96 markers for association with fracture in 31,016 cases and 102,444 controls from 50 studies with fracture information. This collection included: 5,411 cases and 21,909 controls tested in the BMD GWAS discovery samples, 9,187 cases and 45,057 controls tested by in-silico replication and 16,418 cases and 35,478 controls tested by de-novo genotyping (Figure 1 and Online Methods). In this fracture meta-analysis fourteen loci were significantly associated with any type of fracture at a Bonferroni level ($P=5x10^{-4}$), of which five included novel BMD loci. None of the markers displayed large estimates of heterogeneity (Table 2, Supplementary Table 6 and Supplementary Fig. 7). Markers at six of these loci reached $P<5x10^{-8}$ including 18p11.21 (*FAM210A; also known as C18orf19*), 7q21.3 (*SLC25A13*), 11q13.2 (*LRP5*), 4q22.1 (*MEPE*), 2p16.2 (*SPTBN1*) and 10q21.1 (*DKK1*). The proportion of the overall fracture risk explained by BMD ranged between 0.09 and 0.40 across markers (Supplementary Table 7) and was estimated in a subset of Stage 2 samples (including n=8,594 cases and 23,218 controls) by modeling the BMD SNP effect on fracture risk with and without the inclusion of BMD as covariate. In general, the effect of these SNPs on BMD was larger than on fracture risk (Fig. 2A) except for the most significantly associated fracture locus 18p11.21 (Fig. 2B). SNPs in genes of the RANK-RANKL-OPG pathway (*TNFRSF11A-TNFSF11-TNFRSF11B*) despite being the strongest-associated BMD loci were not significantly associated with fracture. All 31 BMD loci with nominal association with fracture risk ($P<0.05$)

showed consistent direction (decreasing BMD allele increased risk of fracture). When we performed subgroup analyses using "cleaner" phenotype definitions such as limiting to clinically-validated fractures and stratifying by anatomical site (i.e. "non-vertebral" fractures and "vertebral" fractures), we did not gain any additional signals (Supplementary Table 8). At a nominally significant level (P<0.05) only three loci were associated with *vertebral fracture* and all 14 BMD loci were associated with *non-vertebral fracture*, but these difference in effects between fracture sites were not significant. Therefore, the power of our study did not benefit from improving phenotype definition at expense of (a lower) sample size.

Allele Risk Modelling for Osteoporosis and Fracture

The combined effect of all significant autosomal SNPs on BMD, osteoporosis and any type of fracture was modelled in the PERF study (n=2,836), a prospective study in postmenopausal Danish women aged 55-86 years.[17] This study comprises an independent validation setting since it was excluded from the overall meta-analysis for this purpose (see Supplementary Note for details). Risk alleles in the score (i.e., BMD-decreasing alleles) were weighted by their individual effect on BMD and grouped in 5 bins (Supplementary Table 9). The difference in mean FN-BMD between individuals in the highest bin of the risk score (9% of the population; n=244) and those in the middle bin (34% of the population; n=978) was -0.33 SDs (Fig. 3A). This analysis was based on 63 SNPs and explained 5.8% (95%CI [4.0-7.6]) of the total genetic variance in FN-BMD.

The ability of this genetic score to predict the risk for osteoporosis (defined as T-score<-2.5) and for fracture was modelled in the PERF study using the middle bin as reference (OR=1). Women in the highest bin had 1.56 (95%CI [1.12-2.18]) increased odds for osteoporosis (Fig. 3B), while women in the lowest bin were protected for osteoporosis (OR=0.38 (95%CI [0.23-0.63])). A model based on the 16 BMD SNPs associated with fracture risk showed that women in the highest bin had 1.60 (95%CI [1.15-2.24]) increased odds for fracture, while women in the lowest bin had a decreased risk for fracture (OR=0.54 (95%CI [0.36-0.83])) (Fig. 3C). Despite serving as a robust proof of principle of the relation between the BMD-decreasing alleles and the risk of osteoporosis and fracture, prediction ability was modest. The Receiver operating characteristics (ROC) curve analysis showed a significant but relatively small discrimination ability of the genetic score alone with an area under the curve (AUC) of 0.59 (95%CI [0.56- 0.62]) for osteoporosis (Supplementary Fig. 8). Adding this score to a model with age and weight alone (AUC 0.75 (95%CI

Table 2 Association of identified BMD-associated loci with risk for any type of low-trauma fracture

			Functional evidence							
SNP	**Locus**	**Closest Gene/ Candidate**	eQTL	Knock-out mouse	OMIM	Tags funct.	GRAIL	Pathway	Risk A.	Freq[b]
Loci significantly associated with fracture risk at P < 5 x 10^{-8}										
rs4233949	2p16.2	*SPTBN1*				•			G	0.63
rs6532023	4q22.1	*MEPE/ SPP1*	•	•			•	•	G	0.67
rs4727338	7q21.3	*SLC25A13*							G	0.32
rs1373004	10q21.1	***MBL2/ DKK1***		•				•	T	0.13
rs3736228	11q13.2	*LRP5*		•	•	•	•	•	T	0.15
rs4796995	18p11.21	***FAM210A***							G	0.39
Other significant loci associated with fracture risk at P < 5 x 10^{-4} (Bonferroni)										
rs6426749	1p36.12	*ZBTB40*							G	0.83
rs7521902	1p36.12[a]	*WNT4*	•				•	•	A	0.27
rs430727	3p22.1	*CTNNB1*		•			•	•	T	0.47
rs6959212	7p14.1	*STARD3NL*							T	0.33
rs3801387	7q31.31	***WNT16***				•	•	•	A	0.74
rs7851693	9q34.11	***FUBP3***							G	0.37
rs163879	11p14.1	*DCDC5*							T	0.66
rs1286083	14q32.12	***RPS6KA5***							T	0.81
rs4792909	17q21.31[a]	*SOST*		•	•			•	G	0.62
rs227584	17q21.31	*C17orf53*	•						A	0.67

Odds Ratios (OR) estimated per risk allele copy for any low-trauma fracture among cases as compared with controls. Q_{het} *is the Cochran's Q statistic and* I^2 *is the measure of heterogeneity.* [a]*rs7521902 and rs4792909 are secondary independent signals.* [b]*Freq is the frequency of the risk allele. Boldface indicates gene names from novel loci and/or P<5x10*$^{-8}$*. Black dots in the six functional evidence columns indicate that: first: SNP is an expression quantitative trait loci (eQTL); second: there is a knockout mouse with skeletal phenotypes (MGI database 2011); third: candidate gene has been involved in a monogenic syndrome with skeletal phenotypes (OMIM database 2011); fourth: most significant SNP tags a SNP predicted to have impact on function of the candidate gene; fifth: gene is best candidate in GRAIL analysis; and sixth: candidate gene is part of a bone active pathway.*

Meta-analysis without studies included in BMD discovery		Combined meta-analysis results			
25,605 cases, 80,535 controls		31,016 cases, 102,444 controls			
OR (95% CI)	P	OR (95%CI)	P	Q_{het} P	I^2
1.07 (1.04-1.09)	1.4×10^{-7}	1.06 (1.04-1.08)	**2.6×10^{-8}**	0.36	6
1.06 (1.04-1.09)	8.8×10^{-7}	1.06 (1.04-1.09)	**1.7×10^{-8}**	1.00	0
1.08 (1.05-1.10)	1.0×10^{-8}	1.08 (1.05-1.10)	**5.9×10^{-11}**	0.03	31
1.09 (1.06-1.13)	7.2×10^{-7}	1.10 (1.06-1.13)	**9.0×10^{-9}**	0.64	0
1.09 (1.05-1.12)	2.1×10^{-6}	1.09 (1.06-1.13)	**1.4×10^{-8}**	0.78	0
1.06 (1.04-1.09)	6.4×10^{-7}	1.08 (1.06-1.10)	**8.8×10^{-13}**	0.12	20
1.06 (1.03-1.09)	2.4×10^{-4}	1.07 (1.04-1.10)	3.6×10^{-6}	0.07	24
1.10 (1.06-1.14)	3.5×10^{-6}	1.09 (1.06-1.13)	1.4×10^{-7}	0.87	0
1.05 (1.03-1.08)	2.4×10^{-5}	1.06 (1.03-1.08)	2.9×10^{-7}	0.93	0
1.04 (1.02-1.07)	1.0×10^{-3}	1.05 (1.02-1.07)	7.2×10^{-5}	0.43	2
1.08 (1.05-1.11)	4.9×10^{-9}	1.06 (1.04-1.08)	2.7×10^{-7}	0.69	0
1.04 (1.01-1.06)	1.9×10^{-3}	1.05 (1.02-1.07)	3.5×10^{-5}	0.65	0
1.06 (1.03-1.09)	6.4×10^{-6}	1.05 (1.03-1.07)	3.3×10^{-5}	0.05	28
1.05 (1.02-1.08)	9.8×10^{-4}	1.05 (1.03-1.08)	7.2×10^{-5}	0.01	34
1.07 (1.04-1.11)	4.0×10^{-5}	1.07 (1.04-1.10)	6.9×10^{-6}	0.31	10
1.05 (1.02-1.08)	2.2×10^{-4}	1.05 (1.03-1.07)	4.1×10^{-5}	0.49	0

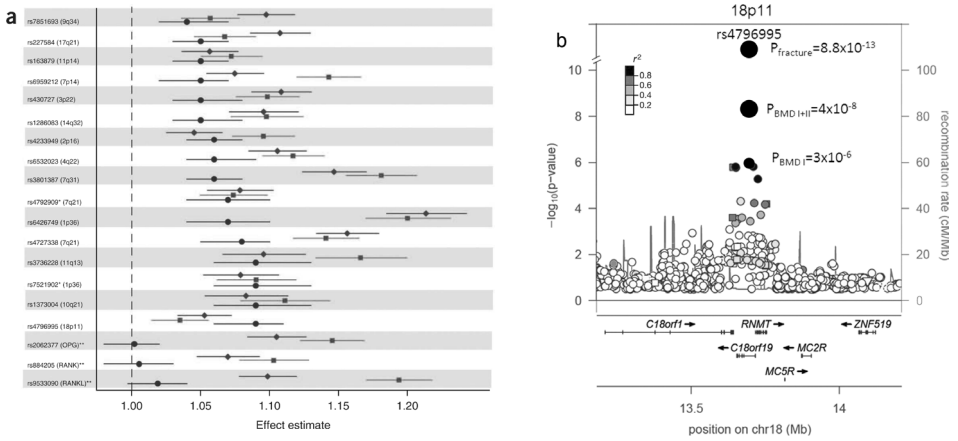


Figure 2. Association of BMD loci with fracture risk. a) Phenotype-wide effect for the BMD loci associated with fracture and those part of the OPG/RANK/RANKL pathway. Genetic effect estimates (± 95%CI) are shown for fracture (circles), lumbar-spine BMD (rectangles) and femoral-neck BMD (diamonds) for the 14 loci associated with fracture risk. Horizontal lines represent 95% confidence limits. Effect estimates are shown after transformation of the standardized mean difference(SMD) in the BMD effect to odds ratio equivalents[34] (e.g. a 0.02 SMD in the BMD effect corresponds to an OR of 1.04). Secondary signals for markers rs227584 and rs6426749 are marked with an asterisk and the signals mapping to the OPG (rs2062377), RANK (rs884205), and RANKL (rs9533090) genes are marked with double asterix. b) Regional association plot for the 18p11.21 locus displaying the P-values for the top SNP associated with fracture (rs4796995) together with P-values of the BMD discovery (Stage 1) and combined with the BMD replication (Stage 1 + 2). SNPs are plotted by position in a 500Kb window of chromosome 18 against association with FN-BMD (–log10 P). Estimated recombination rates (from HapMap) are plotted to reflect the local LD structure. SNPs surrounding most significant SNP are coded with different gray-levelts to reflect LD between markers (pairwise r^2). Genes, exons and transcription direction are derived from the UCSC genome browser.

[0.73-0.77])) did not substantially increase discrimination (AUC 0.76 (95%CI [0.74 -0.78])). A similar pattern was observed for fracture discrimination with an AUC of 0.57 (95%CI [0.55-0.59]) in a model with the score alone and of 0.62 (95%CI [0.60-0.64]) in a model with age, weight and height. A model considering all 63 SNPs did not change the AUC for fracture risk prediction (0.57 (95%CI [0.54-0.59])).

Functional annotations and pathway analyses

1000 Genomes Project (1KGP). For the purpose of fine-mapping and identifying additional SNPs with putative functional implication using linkage disequilibrium (LD), a subset of nine discovery studies (FN-BMD n=21,699; LS-BMD n=20,835) used 1KGP data (Release June/2010) to re-impute genotypes contained in the 55

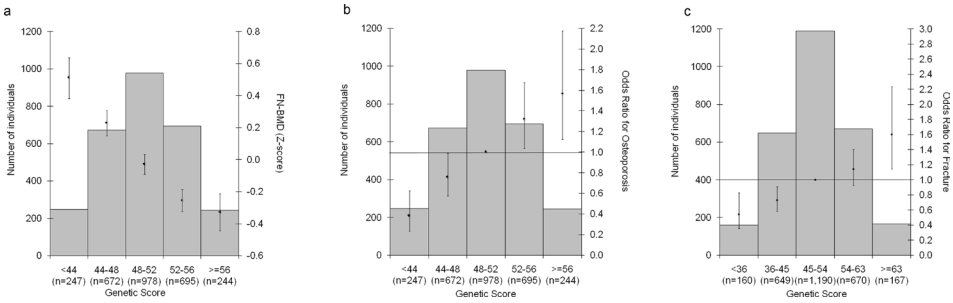


Figure 3. Combined effect of BMD-decreasing alleles and fracture risk-increasing risk alleles modelled in the population-based Prospective Epidemiological Risk Factor (PERF) study (n=2,836 women) on A) Baseline FN-BMD standardized residuals (Z-scores), B) Risk for Osteoporosis and C) Risk for Any type of fracture. The genetic score of each individual for A) and B) was based on the 63 SNPs displaying genome-wide significant association with BMD (55 main and 8 secondary signals), and for C) was based on the 16 BMD SNPs associated with fracture. Both genetic scores are weighted for relative effect sizes estimated without the PERF study. Weighted allele counts summed for each individual were divided by the mean effect size making them equivalent to the percent of alleles carried by each individual and binned into 5 categories. Histograms describe counts of individuals in each genetic score category (left axis scale). Diamonds (right axis scale) represent A) mean FN-BMD standardized levels, risk estimates in the form of odds ratio for B) Osteoporosis (defined as NHANES T-score<-2.5) and for C) Any type of Fracture using the middle category as reference (OR=1). Vertical lines represent 95% confidence limits.

autosomal BMD loci (see Supplementary note for details). In 13 of the 55 BMD loci (X-chromosome SNP not included) we identified markers in a surrounding 1 Mb region that were imputed from 1KGP, and that were more significant than the original HapMap signal (Supplementary Tables 10 and 11) highlighting the benefit of using a denser reference panel of markers. All *HapMap* markers in LD with variants with functional annotation and displaying higher significance in the 1KGP meta-analysis are shown on Supplementary Table 12. In 14 of the 56 discovered BMD loci a marker from the *HapMap* imputation was highly correlated (r^2 >0.8) with at least one putative functional variant annotated in the 1KG reference. Three of the 14 BMD loci associated with fracture contained putative functional variants tagged by the top SNPs of the BMD meta-analysis. These included the known rs3736228 functional marker in *LRP5* (encoding p.Ala1330Val) ,[16,18] the intronic marker rs3779381 within a promoter and/or regulatory region of *WNT16*, and one intronic marker (rs4305309) within a promoter and/or regulatory region of *SPTBN1*.

eQTL analyses. Expression profiles at the GWS BMD loci were analyzed within four datasets (see Supplementary Note). In trans-iliac bone biopsies, expression of

five genes correlated with LS-BMD and/or FN-BMD of the donors with P < 0.001 including *PSME4* (2p16.2), *DKK1* (10q21.1), *C17orf91* (17p13.3), *SOST* (17q21.31_1) and *DUSP3* (17q21.31_1) (Supplementary Table 13). Among them *DKK1* (10q21.1) was the most significantly correlated with FN-BMD ($P=1.3x10^{-5}$) and LS-BMD ($P=3.2x10^{-4}$). Variants in all these BMD loci (with exception of 17p13.3) were also associated with fractures. The SNP-eQTL analyses were performed across diverse tissues examining the correlation between marker alleles and transcript levels at the associated BMD loci. Fourteen of the BMD-associated SNPs correlated with the expression of one or more of the nearby genes with $P < 5\times10^{-5}$ and were either the strongest cis-variants, or good surrogates thereof, for those genes (Supplementary Tables 14 and 15). The most significant BMD-SNP eQTL was observed for rs10835187[T] with reduced expression of the *LIN7C* gene at the 11p14.1 locus ($P = 2.8\times10^{-39}$ in adipose tissue). Of particular interest were BMD-SNP cis-variants at three loci that were also associated with fracture including: 1p36.12, 4q22.1 and 17q21.31. At 1p36.12, rs6426749[G] correlated with reduced *WNT4* expression in fibroblast, osteoblast and adipose tissue; at 4q22.1 rs6532023[G] correlated with reduced SPP1 (osteopontin) expression in adipose tissue and at 17q21.31 rs227584[A] correlated with increased C17orf65 expression in monocytes, adipose tissue, whole blood and lymphoblasts.

GRAIL analysis. We applied the GRAIL text-mining algorithm[19] to investigate connections between genes in the 55 autosomal BMD loci. This analysis revealed significant (GRAIL SNP P<0.01) connections between genes in 18 of the 55 input loci (Fig. 4 and Supplementary Table 16). The strongest connections were seen for members of three key biologic pathways: RANK-RANKL-OPG pathway (*TNFRSF11A-TNFSF11-TNFRSF11B*); mesenchymal stem cell differentiation (*RUNX2, SP7, SOX9*); and Wnt-signalling (*LRP5, CTNNB1, SFRP4, WNT3, WNT4, WNT5B, WNT16, AXIN1*) with the ten most frequently connecting terms being: 'bone', 'catenin', 'signaling', 'differentiation', 'rank', 'osteoblast', 'diacylglycerol', 'kappab', 'development', and 'osteoclast'. To assess the significance of this "biological" gene connection enrichment we applied GRAIL to 2000 random matched sets of 55 SNPs (See Supplementary note for details) and we did not observe any set with 15 or more loci with significant enriched connectivity (Supplementary Fig. 9) providing strong statistical evidence of the significant clustering of our BMD loci (P < 0.0005).

DISCUSSION

In this, the largest GWAS for osteoporosis traits to date, we identified 32 novel genomic loci bringing to 56 the number of loci robustly associated with BMD variation. Furthermore, we report for the first time that six of these BMD loci are

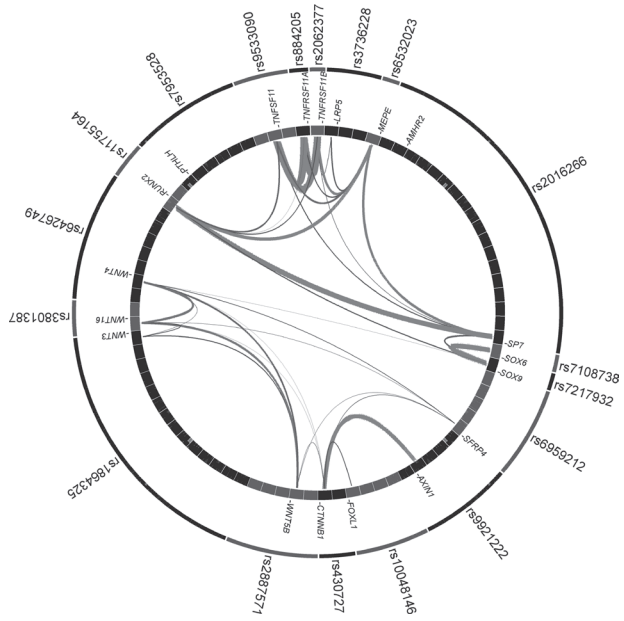


Figure 4. Graphic representation of GRAIL connections between SNPs and corresponding genes for the 18 SNPs as determined with GRAIL P<0.01. The top 10 keywords linking the genes were: 'bone', 'catenin', 'signaling', 'differentiation', 'rank', 'osteoblast', 'diacylglycerol', 'kappab', 'development', and 'osteoclast'. Thicker lines imply stronger literature-based connectivity. Gray and black boxes depict loci boundaries represented per top-associated marker (outer circle) and per gene in the region (inner circle).

associated with low-trauma fractures at $P<5x10^{-8}$. In terms of other complex traits, our results indicate hundreds of variants with small effects may be contributing to the genetic architecture of BMD and fracture risk.[20] Our hypothesis-free assessment of common variants of the genome provides novel insights into biology, implicating several factors clustering in bone-active pathways.

Our results highlight the highly polygenic nature underlying BMD variation and the critical role of several biological pathways influencing osteoporosis and fracture susceptibility (Supplementary Fig. 10). On top of the Wnt factors known to be associated with BMD (*CTNNB1*, *SOST*, *LRP4*, *LRP5*, *WLS*, *WNT4*, *MEF2C*) several of the newly discovered loci also implicate additional Wnt signalling factors (including *WNT5B*, *WNT16*, *DKK1*, *PTHLH, SFRP4 and AXIN1*). Another clearly delineated pathway is that involved in mesenchymal cell differentiation, including the newly identified *RUNX2, SOX4* and *SOX9* BMD loci along with the previously known *SP7.*

Another bone-relevant pathway includes that of "Endochondral Ossification" which involves essential processes during the fetal development of the mammalian skeleton and which implicated several of our identified BMD loci including: *SPP1, MEF2C, RUNX2, SOX6, PTHLH, SP7* and *SOX9*. In addition, the biological relevance of our associations is accentuated by the identification of genes underlying rare monogenetic forms of osteoporosis and/or high bone mass such as *SOST, CLCN7*, *LRP5* (refs. 21-23) (Supplementary Table 17) which also contain common variants involved in normal BMD variation at the population level.[11,14,16] This is supportive of a genetic architecture where both common and rare genetic variation may reside in the same locus.[24] Other genes have not been reported to be associated with monogenic forms for osteoporosis but have clear involvement in bone development in animal models. For example, SNPs in the 16q12.1 BMD locus map near *CYLD*. Human mutations in this gene have been described to cause familial cylindromatosis a condition without phenotypic skeletal manifestations. However, it has been shown that *Cyld* knock-out mice have significant bone loss leading to a severe osteoporosis phenotype[25] and also that *CYLD* regulates osteoclastogenesis.[26] Moreover, evidence from the GWAS and eQTL analyses also suggests some loci contain more than one common variant with independent effects on BMD and fracture risk. On the other hand, when no correlation is observed between gene expression and a particular SNP, it is difficult to draw conclusions. A correlation might be missed if the expression of the transcript was not measured in a relevant tissue or if the expression of a particular splice-variant was not measured.[27]

BMD and fracture genetic effects correlate to some extent, but some important fracture risk variants may have minimal impact on BMD and vice versa. This is the case for the 18p11.21 signal (Fig. 2B) mapping to a gene coding for a protein of unknown function, which despite a modest effect on BMD (0.02% variance explained) displayed the most significant association with fracture risk (OR=1.08, 95%CI[1.06-1.10], $P=8.8x10^{-13}$). This is in contrast to variants with known stronger effects on BMD which were not significantly associated with fracture risk. For example, variants in the RANK-RANKL-OPG pathway, known to play a critical role in osteoclastogenesis, had clear associations with BMD but not fracture risk (Fig. 2A). Even though loci discovery was based on the BMD phenotype, these findings reflect the heterogeneous and complex nature of the mechanistic pathways leading to fracture. Therefore, given our study design, we cannot rule out the possibility that yet unidentified genetic loci are influencing risk of fracture independently of BMD. Future well-powered GWAS meta-analyses on fracture risk will address this question while corroborating the associations with fracture that we report for some of the

BMD loci (particularly those not associated with fracture at $P<5x10^{-8}$).

Our study also provides indication that there is sex- and site-specificity underlying BMD variation. One of the GWAS signals (Xp22.31) was only significant in the sex-stratified analysis in men and displayed significant sex heterogeneity ($P_{het}=1.62x10^{-8}$). This is expected considering the sexual dimorphism of bone.[28,29] In fact, in a recent GWAS, the rs5934507 SNP mapping to Xp22.31, which is associated with BMD in the current study, has been previously associated with male serum testosterone levels.[30] Thus, it is likely that rs5934507 affects serum testosterone, which in turn regulates BMD. In line with the different types of bone composition at the different skeletal sites (predominantly trabecular at the lumbar spine while predominantly cortical at the femoral neck) we observed some indication of site specificity in 10 of the 56 BMD loci, suggesting differential genetic influences on BMD determination across skeletal sites. As has been previously shown[31], we did not find in our results major differences in effect sizes between individuals of European and East Asian ancestry (Supplementary Fig. 7). However, this may be due to reduced power given the smaller number of individuals of East Asian ancestry. We tested a genetic risk score to identify individuals at risk of osteoporosis and fracture and showed that cumulatively, the identified variants generate a gradient of risk. These gradients reach ORs of 1.56 for osteoporosis and 1.60 for fractures when comparing participants with the highest risk scores with those reflecting the mean score. Yet, at present there is limited clinical utility in using this score as evidenced by the non-significant contribution to case discrimination after considering clinical risk factors with strong effects on osteoporosis and fracture risk (like age and weight). This is not unexpected given the small fraction of genetic risk for either BMD or fracture that has been identified thus far.

Our study has limitations. The identified SNPs are probably not the causal variants; it is more likely that these markers are in LD with the underlying causal variants. Additional analyses on potential functional SNPs identified in this study will be required to determine if they are causal to these relationships with BMD. Moreover, the causal genes underlying the GWAS signals may be different from the candidate genes we describe, considering that our understanding of their role in bone biology is limited. Further exploration of these loci with more detailed sequencing, gene expression, and translational studies will be required. Such studies can also disentangle the diverse types of complex relationships we currently cannot distinguish in the BMD loci with secondary signals, i.e., if these are the result of true allelic heterogeneity or if they are driven by a second gene in the same region.[32] Similarly, despite our large sample size, power limitations still play a role for

detecting additional associations with smaller effect sizes and/or arising from rarer variants. Finally, given the different levels of data availability and the difficulties for standardization across studies, we did not evaluate the effect of additional risk factors for osteoporosis, such as menopausal status and smoking, which can influence the genetic associations with BMD. Nonetheless, despite these limitations we have identified many novel and previously unsuspected associations with BMD variation and fracture risk.

Finally, the relatively weak effects of the variants discovered by GWAS do not undermine the biological relevance of the genes identified, as exemplified by the identification of genetic signals at the location of genes coding for proteins currently targeted by novel osteoporosis treatments (Supplementary Fig. 10). The novel genes identified in our study may represent new candidates to target for osteoporosis drug discovery. Most established treatments for osteoporosis currently focus on curtailing bone resorption (eg. bisphosphonates, RANKL inhibitors) while only few anabolic treatments are currently approved for the treatment of osteoporosis (i.e. recombinant truncated or altered PTH). Other anabolic compounds under Phase II development include PTHrP fragments and Wnt-signaling enhancers such as anti-Sclerostin antibodies.[33] Several of the variants robustly associated with BMD map in or close to genes of proteins involved in these pharmacologic pathways, namely osteoprotegerin (*TNFRSF11B*), RANK (*TNFRSF11A*), RANKL (*TNFSF11*), PTHrP (*PTHLH*), Low-density lipoprotein receptor-related protein 5 (*LRP5*), Sclerostin (*SOST*), and Dickkopf-1 (*DKK1*).

In conclusion, these findings highlight the highly polygenic and complex nature underlying BMD variation, shedding light on the pathophysiological mechanisms underlying fracture susceptibility and may contribute to the identification of future drug targets for the treatment of osteoporosis.

METHODS

Study design. This study is part of the GEnetic Factors for OSteoporosis consortium (GEFOS), a coalition of teams of investigators dedicated to identify the genetic determinants of osteoporosis. The discovery samples comprised 17 GWA studies (n=32,961) from populations across North America, Europe, East Asia and Australia, with a variety of epidemiological designs (Supplementary Table 18A) and patient characteristics (Supplementary Table 18B); a subset of which had fracture information (Supplementary Table 18C). Subjects from 34 additional studies with BMD data (n=50,933) were used for replication while association with fracture was

tested across 50 studies with fracture information, most of them also used for the BMD analysis (n=31,016 cases and 102,444 controls) (Figure 1 and Supplementary Tables 19A-C and 20A-C). All studies were approved by their institutional ethics review committees and all participants provided written informed consent.

BMD measurements and fracture definition. BMD of the lumbar spine (LS-BMD) and femoral neck (FN-BMD) was measured in all cohorts using dual-energy X-ray absorptiometry following standard protocols (Supplementary Tables 18B, 19B and 20B). Three clinically-distinct fracture definitions were used: 1) *Any type,* consisting of low-trauma fractures at any skeletal site (except fingers, toes and skull) occurring after age 18 years assessed by X-ray, radiographic report, clinical record, clinical interview and/or questionnaire; 2) *Validated non-vertebral*, consisting of fractures occurring after age 50 years with diagnosis confirmed by hospital records and/or radiographs; and 3) *Radiographic vertebral fractures*, from lateral morphometry scored on X-rays. The first is most-inclusive, while the latter two are more stringent fracture definitions commonly used in randomized trials.[35,36] Controls were defined as individuals without a history of fracture using for each fracture type the same age limit categories of the cases.

Stage 1 Genome-wide association analysis. *Genotyping and Imputation*: GWAS genotyping was done by each study following standard protocols followed by imputation to ~2.5 million SNPs from HapMap[37] Phase II release 22 using Genome Build 36. Quality control was performed independently for each study. To facilitate meta-analysis, each group performed genotype imputation with BIMBAM[38], IMPUTE[39], or MACH[40] software using genotypes from the HapMap Phase II release 22 (CEU or CHB/JPT as appropriate). HapMap release 21 was used as reference for SNPs residing on the X chromosome and IMPUTE software was used for imputation. Overall imputation quality scores for each SNP were obtained from IMPUTE (proper_info) and MACH (rsq_hat) statistics. Details on the genotyping platform used, genotype quality control procedures and software for imputation employed for each study are presented in the Supplementary Tables 18D and 19D.

Association analysis with BMD: each study performed genome-wide association analysis for FN-BMD and LS-BMD using sex-specific, age- weight- and principal components-adjusted standardized residuals analyzed under an additive (per allele) genetic model. Analyses of autosomal and chromosome X markers were done separately. Analysis of imputed genotype data accounted for uncertainty in each genotype prediction by using either the dosage information from MACH or the genotype probabilities from IMPUTE and BIM-BAM. Studies used MACH2QTL[40] directly or via GRIMP[41] (which uses genotype dosage value as a predictor in a

linear regression framework), SNPTEST[39], Merlin[42], BIM-BAM or the linear mixed effects model of the Kinship and ProbABEL[43] (Supplementary Tables 18D and 19D). For analysis of the X-chromosome either SNPTEST or R package was used in each participating study. We coded "effect allele homozygous genotype" as "2" and "other allele homozygous genotype" as "0" in the genotyped SNPs in men on the X chromosome. The imputed genotypes were coded as continuous variables from 0 to 2 to take into account imputation uncertainty. The genomic control method[44] was used to correct the standard error (SE) by the square root of the genomic inflation factor (lambda): $SE_{corrected} = SE \times \sqrt{lambda}$.

Meta-analysis of the GWA studies: before performing meta-analysis on the genome-wide association data, SNPs with poor imputation quality scores (rsq_hat < 0.3 in MACH, proper_info < 0.4 in IMPUTE or the ratio of observed to expected dosage variance < 0.3 in BIMBAM) and markers with a minor allele frequency < 1% were excluded for each study. All individual GWAS were genomic control corrected before meta-analysis.[44] Individual study-specific genomic control values ranged from 0.98 to 1.08. (Supplementary Table 18D). A total of 2,483,766 autosomal SNPs were meta-analyzed across: 17, 16 and 13 studies for FN-BMD (Pooled, women-only, and men-only analyses, respectively) and 16, 13 and 12 studies for LS-BMD (Pooled, women-only, and men-only analyses, respectively). A total of 76,253 X-linked SNPs were meta-analyzed across: 14, 13 and 10 studies for LS and FN-BMD (Pooled, women-only, and men-only analyses, respectively). In our discovery analysis, we chose to implement a fixed effects models approach as it is generally preferable for the purposes of initial discovery, where the aim is to screen and identify as many of the true variants as possible.[45,46] SNPs present in less than three studies were removed from the meta-analysis yielding ~ 2.2 million SNPs in the final results. Genomic inflation factors (λ) were 1.11, 1.09, 1.06 for FN-BMD BMD (Pooled, women-only, and men-only analyses, respectively) and 1.13, 1.09, 1.06 for LS-BMD (Pooled, women-only, and men-only analyses, respectively). A second GC correction was applied to the overall meta-analysis results, although such second correction is considered overly conservative.[47] Significance for BMD association was set at $P<5x10^{-8}$ while a Bonferroni correction was used for the association with fracture.[48]

Selection of SNPs for follow up: we took forward the most significant 96 SNPs for replication. Based on power estimations, after adding 30,000 samples in stage 2 these variants had a priori Power >=85% to reach $P= 5 \times 10^{-8}$ in the meta-analysis. Loci were considered independent when separated by at least 1 Mb down and upstream of the top GWAS signal. The 96 variants included the 82 index SNPs representing each of the 82 loci reaching $P<5x10^{-6}$ in Stage 1, 9 SNPs that lie within the same 2Mb

windows as the 82 and were independent from the main signal (secondary signals), and the top-five most associated SNPs of the X-chromosome (with $P<5x10^{-5}$).

Association analyses with fracture risk: effect estimates (odds ratio) for association of allele dosage of the top hits with fracture risk were obtained from logistic regression models adjusted for age, age^2 weight, sex, height and four principal components. The proportion of the fracture risk explained by FN-BMD was calculated from the regression coefficients as $(\beta_{unadjusted} - \beta_{BMDadjusted})/ \beta_{unadjusted}$ in a subset of replication samples for which both FN-BMD and complete fracture information was available.

Stage 2 follow up. *Samples and genotyping*: fracture association results were also obtained for the 82 most significant SNPs from 54,244 individuals of European ancestry from 7 GWAS (in-silico genotyping) that had not been included in the stage 1 analyses (Supplementary Tables 19A, 19B and 19C). Subjects from 34 studies of the GENOMOS consortium with BMD and/or fracture information were studied for replication (Supplementary Tables 3A, 3B and 3C). De-novo replication genotyping was done in the UK (Kbiosciences), Iceland (deCODE Genetics), Australia (University of Queensland Diamantina Institute) and the USA (WHI GeCHIP) using KASPar, Centaurus, OpenArray and iSelect assays respectively (Supplementary Note). Minimum genotyping quality control criteria were defined as: Sample call rate > 80%, SNP call rate > 90%, HWE $P > 1x10^{-4}$, MAF > 1%.

Association analyses and meta-analysis: We tested the association between the 96 SNPs and BMD and fracture risk in each *in-silico* and *de-novo* "Stage 2" study separately as described for the "Stage 1" studies. We subsequently meta-analyzed effects and standard errors from the "Stage 2" studies, followed by a meta-analysis of the summary statistics of both "Stage 1" and "Stage 2" using the inverse-variance method in METAL. At this replication stage, where more than 30 studies were synthesized, we chose to first assess the underlying heterogeneity considering both the Cochran's Q statistic and the I^2 metric. If the heterogeneity was not significant fixed effects models were applied. If the Cochrane Q P-value<0.0005 and the I^2 was > 50% we used the more conservative random effects models.

Additional analyses. Further analyses were performed for the SNPs carried forward for replication. Each of these analyses is described in detail in the **Supplementary Note**.

In brief, we performed: 1) a conditional genome-wide association analysis to examine whether any of the 82 BMD loci harbored additional independent signals; 2) tested gene-by-gene pair-wise interactions between these BMD loci; 3) assessed within the independent setting of the PERF study (for details on study design see Supplementary Tables 20A, 20B & 20C) the predictive ability derived from

the cumulative effect of the 63 genome-wide significant autosomal BMD SNPs in relation to BMD levels and osteoporosis risk; and that of the 16 BMD SNPs also associated with fracture risk in relation to fracture risk; 4) identified SNPs having $r^2 \geq 0.80$ with the lead SNP that were potentially functional (nonsense, non-conservative non-synonymous, synonymous, exonic splicing, transcription factor binding sites, etc) using regional imputation with the 1000 Genomes data (June 2010 release); 5) tested the relationship between gene expression profiles from a) trans-iliacal bone biopsies and BMD in 84 unrelated postmenopausal women[49] and b) also examined cis- associations between each of the 55 significant BMD SNPs and expression of nearby genes in different tissues including lymphoblastoid cell lines[50-52], primary human fibroblasts and osteoblasts[53], adipose tissue[54], whole blood[54] and circulating monocytes[55]; and finally 6) evaluated the connectivity and relationships between identified loci using the literature-based annotation with Gene Relationships across Implicated Loci (GRAIL[19]) statistical strategy.

ACKNOWLEDGMENTS

We thank all study participants for making this work possible. This research and the Genetic Factors for Osteoporosis (GEFOS) consortium have been funded by the European Commission (HEALTH-F2-2008-201865-GEFOS). We acknowledge funding from the following organizations: the US National Institutes of Health (NIH; R01 AG18728, R01 HL088119, R01AR046838, U01 HL084756, P30 DK072488, T32 AG000262, F32 AR059469, P01 AG-18397, R01 AG041517, M01 RR-00750 and N01-AG-12100), the NIA Intramural Research Program (AG-023629, AG-15928, AG-20098 and AG-027058), Hjartavernd (the Icelandic Heart Association), the Althingi (the Icelandic Parliament), the Australian National Health and Medical Research Council (511132), the Australian Cancer Research Foundation and the Rebecca Cooper Foundation, the Australian National Health and Medical Research Council Career Development Award (569807 to E.L.D.), an MRC New Investigator Award (MRC G0800582 to D.M.E.), the Health Research Council of New Zealand, Sanofi-Aventis, Eli Lilly, Pfizer, Proctor & Gamble Pharmaceuticals, Roche, the Medical Benefits Fund (MBF) Living Well Foundation, the Ernst Heine Family Foundation, Arthritis Research UK (17539 and 15389), The Victorian Health Promotion Foundation, Geelong Region Medical Research Foundation, Australia (628582). Action Research UK, the European Commission (QLRT-2001-02629), the UK Food Standards Agency, BioPersMed (COMET K-project 825329), the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT), the Austrian Federal Ministry of Economics and Labour (BMWA), the Austrian Federal Ministry of Economy, Family and Youth (BMWFJ), the Styrian Business Promotion Agency (SFG), the Red de Envejecimiento y Fragilidad (RETICEF), Instituto Carlos III, the Spanish Ministry of Education and Science (SAF2010-15707), the Government of Catalonia (2009SGR971 and 2009SGR818), Instituto de Salud Carlos III–Fondo de Investigaciones Sanitarias (PI 06/0034 and PI08/0183), Healthway Health Promotion Foundation of Western Australia, Australasian Menopause Society and the Australian National Health and MRC Project (254627, 303169 and 572604), the Finnish Ministry of Education, Merck Frosst Canada, Eli Lilly Canada, Novartis Pharmaceuticals, Procter and Gamble Pharmaceuticals Canada, Servier Canada, Amgen Canada, The Dairy Farmers of Canada, The Arthritis Society, the US National Heart, Lung, and Blood Institute (NHLBI; N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, HL080295, HL075366, HL087652, HL105756 NINDS, HL 043851 and HL69757, CA 047988, and the Framingham Heart Study (N01-

HC-25195) and its contract with Affymetrix, Inc, for genotyping services (N02-HL-6-4278)). Untied Educational Grants were provided by Amgen, Eli Lilly International, GE-Lunar, Merck Australia, Sanofi-Aventis Australia and Servier. Additional support was provided by the US National Center for Research Resources (M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping Core), the US National Institute of Diabetes and Digestive and Kidney Diseases (DK063491 to the Southern California Diabetes Endocrinology Research Center), deCODE Genetics, The UK National Institute for Medical Research (NIMR) Biomedical Research Centre, the Cancer Research Campaign, the Stroke Association, the British Heart Foundation, the UK Department of Health, the Europe Against Cancer Programme Commission of the European Union, the Ministry of Agriculture, Fisheries and Food, EU Biomed 1 (BMHICT920182, CIPDCT925012, ERBC1PDCT 940229 and ERBC1PDCT930105), the UK MRC (G9321536 and G9800062), the Wellcome Trust collaborative Research Initiative 1995, MAFF AN0523, EU Framework Programme 5 (FP5; QLK6-CT-2002-02629), the Food Standards Agency (N05046), the Netherlands Organization for Scientific Research (NWO), Erasmus University Medical Center, the Centre for Medical Systems Biology (CMSB1 and CMSB2) of the Netherlands Genomics Initiative (NGI), the F.I.R.M.O. Fondazione Raffaella Becagli, the National Institute for Arthritis, Musculoskeletal and Skin Diseases, the National Institute on Aging (R01 AR/AG 41398, N01AG62101, N01AG62103, N01AG62106, 1R01AG032098 and R01 AR 050066), the Canadian Institutes for Health Research (86748), Federal Program of the Ministry of Education and Science of the Russian Federation Scientific and Pedagogical Staff of Innovative Russia in 2009–2013 (P-601), the Federal Program Research and Development of Prior Directions of Scientific-Technological Complex of Russia in 2007–2012 (16.512.11.2032), the Swedish Research Council (K2010-54X-09894-19-3, 2006-3832, K2010-52X-20229-05-3) and K20006-72X-20155013 the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, the Novo Nordisk foundation. University of Athens, Greece (Kapodistrias 2009), the UK NIHR Musculoskeletal BRU Oxford; the UK NIHR Nutrition BRU Southampton, The Center for Inherited Disease Research (CIDR), National Institutes of Health (HHSN268200782096C), the Hong Kong Research Grant Council (HKU 768610M), The Bone Health Fund of the HKU Foundation, The KC Wong Education Foundation, Small Project Funding (201007176237), Matching Grant, Committee of Research and Conference Grants (CRCG) Grant the Osteoporosis and Endocrine Research Fund, the Genomics Strategic Research Theme of The University of Hong Kong, Chinese University of Hong Kong, the Korea Health 21 Research & Development Project, the Korean Ministry of Health & Welfare, Republic of Korea (A010252), the Korea Healthcare Technology Research & Development Project, the Ministry for Health, Welfare and Family Affairs (A110536), The Netherlands Ministry of Health, Welfare and Sports Directorate of Long-Term Care, the World Anti-Doping Agency, the Danish Ministry of Culture, the Institute of Clinical Research of the University of Southern Denmark, the Chief Scientists Office of the Scottish Government (CZB/4/276), a Royal Society University Research Fellowship (to J.F.W.), the European Union Framework Program 6 EUROSPAN project (LSHG-CT-2006-018947), the European Union's Seventh Framework Programme (FP7/2007-2013; HEALTH-F2-2009-223004 PHASE), the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011 and 911-03-012), the Research Institute for Diseases in the Elderly (RIDE2; 014-93-015), the Netherlands Genomics Initiative/Netherlands Consortium for Healthy Aging (050-060-810), the German Bundesministerium fuer Forschung und Technology (01 AK 803 A-H and 01 IG 07015 G), the NIHR Biomedical Research Centre (grant to Guys' and St. Thomas' Hospitals and King's College London), the Chronic Disease Research Foundation, the Canadian Institutes of Health Research, the Canadian Foundation for Innovation, the Fonds de la Recherche en Santé Québec, The Lady Davis Institute, the Jewish General Hospital, the Ministère du Développement Economique, de l'Innovation et de l'Exportation du Quebec, the Swedish Sports Research Council (87/06), the Swedish Society of Medicine, the Kempe Foundation (JCK-1021), the Medical Faculty of Umeå University (ALFVLL:968:22-2005, ALFVL:-937-2006, ALFVLL:223:11-2007 and ALFVLL:78151-2009), the County Council of Västerbotten (Spjutspetsanslag; VLL:159:33-2007); the US National Cancer Institute, the Donald W. Reynolds Foundation, the Fondation Leducq, the Academy of Finland (126925, 121584, 124282, 129378 (Salve), 117787 (Gendi) and

41071 (Skidi)), the Social Insurance Institution of Finland, Kuopio, Tampere and Turku University Hospital Medical Funds (9M048 for TeLeht), the Juho Vainio Foundation, the Paavo Nurmi Foundation, the Finnish Foundation of Cardiovascular Research, the Finnish Cultural Foundation, the Tampere Tuberculosis Foundation, the Emil Aaltonen Foundation (K08AR055688 to T.L.). A detailed list of acknowledgments by study is given in the Supplementary Note.

AUTHOR CONTRIBUTIONS

Study-specific design and management were performed by U.S., M.A., L.M., J.P., S.B., M.L.B., B.M.B., C. Christiansen, C. Cooper, G.D., I.F., M.F., D.G., J.G.-M., M. Kähönen, M. Karlsson, J.-M.K., P.K., B.L.L., W.D.L., P.L., Ö.L., R.S.L., J.M., D.M., J.M.O., U.P.-K., J.A.R., P.M.R., F. Rousseau, P.E.S., N.L.S.T., R.U., W.V.H., J.V., M.T.Z., K.M.G., T.P., D.I.C., S.R.C., R.E., J.A.E., V.G., A.H., R.D.J., G.J., J.W.J.,K.-T.K., T.L., M. Lorentzon, E.M., B.D.M., G.C.N., M.P., H.A.P.P., R.L.P., O.R., I.R.R.,P.N.S., P.C.S., A.R.S., F.A.T., C.M.v.D., N.J.W., L.A.C., M.J.E., T.B.H., A.W.C.K., B.M.P., J. Reeve, T.D.S., E.A.S., M.C.Z., U.T., C.O., J.B.R., M.A.B., K. Stefansson, A.G.U., S.H.R., J.P.A.I., D.P.K. and F. Rivadeneira. Study-specific genotyping was performed by K.E., U.S., E.L.D., L.O., L.V., S.-M.X., A.K.A., D.J.D., S.G., R.K., C.K., A.Z.L., J.R.L., S.M., S.M.-B., S.S., S.T., O.T., S.C., E.K., J.M., B.O.-P., Y.S.A., E.G., L.H., H.J., T. Kwan, R. Luben, C.M.-G., S.T.P., S. Reppe, J.I.R., J.B.J.v.M., D.V., K.M.G., D.I.C., G.R.C., P.D., R.D.J., T.L., Y.L., M. Lorentzon, R.L.P., N.J.W., L.A.C., C.O., M.A.B., A.G.U. and F. Rivadeneira. Study-specific phenotyping was performed by U.S., E.L.D., O.M.E.A., A.M., S.-M.X., N. Alonso, S.K.K., S.G.W., A.K.A., T.A., J.R.C., Z.D., N.G.-G., S.G., G.H., L.B.H., K.A.J., G.K., G.S.K., C.K., T. Koromila, M. Kruk, M. Laaksonen, A.Z.L., S.H.L., P.C.L., L.M., X.N., J.P., L.M.R., K. Siggeirsdottir, O.S., N.M.v.S., J.W., K.Z., M.L.B., C. Christiansen, M.F., M. Kähönen, M. Karlsson, J.-M.K., Ö.L., J.M., D.M., B.O.-P., J.M.O., U.P.-K., D.M.R., J.A.R., P.M.R., F. Rousseau, W.V.H., J.V., M.C.-B., E.G., T.I., R. Luben, S. Reppe, G.S., J.B.J.v.M., D.V., F.M.K.W., K.M.G., J.A.C., D.I.C., E.M.D., R.E., J.A.E., V.G., A.H., R.D.J., G.J., Y.L., M. Lorentzon, E.M., G.C.N., B.A.O., M.P., H.A.P.P., R.L.P., O.R., I.R.R., J. Robbins, P.N.S., C.M.v.D., M.J.E., J. Reeve, E.A.S., M.C.Z., C.O., M.A.B., A.G.U., D.P.K. and F. Rivadeneira. Study-specific data analysis were performed by K.E., U.S., E.E., Y.-H.H., E.L.D., E.E.N., L.O., O.M.E.A., N. Amin, J.P.K., D.L.K., G.L., C.L., R.L.M., A.M., L.V., D.W., S.-M.X., L.M.Y.-A., H.-F.Z., J.E., C.M.K., S.K.K., P.J.L., G.T., J.F.W., V.A., A.K.A., T.A., J.R.C., G.H., L.J.H., C.K., T. Koromila, A.Z.L., S.M.-B., T.V.N., M.S.P., J.P., L.M.R., A.V.S., O.S., S.T., S.C., J.M., B.O.-P., U.P.-K., R. Li, R. Luben, S. Reppe, J.I.R., A.R.W., Y.Z., S. Raychaudhuri, D.I.C., J.A.E., R.D.J., T.L., K.N., O.R., D.M.E., D.K., J.B.R., M.A.B., J.P.A.I., D.P.K. and F. Rivadeneira. Analysis plan design was performed by K.E., E.E., U.S., D.K., D.P.K., J.P.A.I. and F. Rivadeneira. K.E., E.E., Y.-H.H. and E.E.N. carried out meta-analyses. K.E., E.E. and A.R.W. determined gene-by-gene interaction. Risk modeling and analysis of secondary signals were

performed by K.E. and F. Rivadeneira. Expression QTLs were analyzed by U.S., G.T., E.G., S. Reppe, K.M.G. and T.P. Y.-H.H. performed functional SNP prediction. GRAIL was carried out by K.E., E.L.D., D.W. and S. Raychaudhuri. Standardization of phenotype and genotype replication data sets was performed by K.E., U.S., E.E., E.L.D., L.O., G.T., L.H. and C.M.-G. Interpretation of results was carried out by K.E., U.S., E.E., Y.-H.H., E.L.D., E.E.N., L.O., O.M.E.A., N. Amin, D.L.K., C.-T.L., R.L.M., A.M., L.V., D.W., S.-M.X., L.M.Y.-A., J.E., C.M.K., S.K.K., A.W.C.K., J. Reeve, M.C.Z., C.O., D.K., J.B.R., M.A.B., A.G.U., S.H.R., J.P.A.I., D.P.K. and F. Rivadeneira. The manuscript draft was prepared by K.E., U.S., E.E., Y.-H.H., E.L.D., E.E.N., L.O., O.M.E.A., A.M., C.O., D.K., J.B.R., M.A.B., A.G.U., S.H.R., J.P.A.I., D.P.K. and F. Rivadeneira. The steering committee for GEFOS includes U.S., E.E., U.T., A.G.U., S.H.R., J.P.A.I. and F. Rivadeneira.

COMPETING FINANCIAL INTERESTS

The coauthors affiliated with deCODE Genetics in Reykjavík Iceland withhold stock options in that company.

URLS

GEFOS Consortium, http://www.gefos.org/; GENOMOS Consortium, http://www.genomos.eu/; HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; 1000 Genomes Project, http://www.1000genomes.org/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; METAL, http://www.sph.umich.edu/csg/abecasis/Metal/.

Note: Supplementary information is available on the Nature Genetics website**.**

REFERENCES

1. Burge, R. *et al.* Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res* **22**, 465-75 (2007).
2. Johnell, O. *et al.* Predictive value of BMD for hip and other fractures. *J Bone Miner Res* **20**, 1185-94 (2005).
3. Kanis, J.A. *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. *Osteoporos Int* **18**, 1033-46 (2007).
4. Peacock, M., Turner, C.H., Econs, M.J. & Foroud, T. Genetics of osteoporosis. *Endocr Rev* **23**, 303-26 (2002).
5. Ralston, S.H. & Uitterlinden, A.G. Genetics of osteoporosis. *Endocr Rev* **31**, 629-62 (2010).
6. Hardy, J. & Singleton, A. Genomewide association studies and human disease. *N*

Engl J Med **360**, 1759-68 (2009).

7. Manolio, T.A. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* **363**, 166-76 (2010).
8. Richards, J.B. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
9. Styrkarsdottir, U. *et al.* Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* **358**, 2355-65 (2008).
10. Rivadeneira, F. *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* **41**, 1199-206 (2009).
11. Styrkarsdottir, U. *et al.* New sequence variants associated with bone mineral density. *Nat Genet* **41**, 15-7 (2009).
12. Hsu, Y.H. *et al.* An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. *PLoS Genet* **6**, e1000977 (2010).
13. Kung, A.W. *et al.* Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am J Hum Genet* **86**, 229-39 (2010).
14. Duncan, E.L. *et al.* Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet* **7**, e1001372 (2011).
15. Richards, J.B. *et al.* Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern Med* **151**, 528-37 (2009).
16. van Meurs, J.B. *et al.* Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. *Jama* **299**, 1277-90 (2008).
17. Bagger, Y.Z. *et al.* Links between cardiovascular disease and osteoporosis in postmenopausal women: serum lipids or atherosclerosis per se? *Osteoporos Int* **18**, 505-12 (2007).
18. Kiel, D.P. *et al.* Genetic variation at the low-density lipoprotein receptor-related protein 5 (LRP5) locus modulates Wnt signaling and the relationship of physical activity with bone mineral density in men. *Bone* **40**, 587-96 (2007).
19. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).
20. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8 (2010).
21. Brunkow, M.E. *et al.* Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. *Am J Hum Genet* **68**, 577-89 (2001).
22. Guo, Y.F. *et al.* Polymorphisms of the low-density lipoprotein receptor-related protein 5 (LRP5) gene are associated with obesity phenotypes in a large family-based association study. *J Med Genet* **43**, 798-803 (2006).
23. Kornak, U. *et al.* Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* **104**, 205-15 (2001).
24. Panagiotou, O.A., Evangelou, E. & Ioannidis, J.P. Genome-wide significant associations for variants with minor allele frequency of 5% or less--an overview: A HuGE review. *Am J Epidemiol* **172**, 869-89 (2010).
25. Jin, W. *et al.* Deubiquitinating enzyme CYLD negatively regulates RANK signaling and osteoclastogenesis in mice. *J Clin Invest* **118**, 1858-66 (2008).

26. Sundaram, K., Shanmugarajan, S., Rao, D.S. & Reddy, S.V. Mutant p62P392L stimulation of osteoclast differentiation in Paget's disease of bone. *Endocrinology* **152**, 4180-9 (2011).

27. Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res* **19**, 1942-52 (2009).

28. Duan, Y., Beck, T.J., Wang, X.F. & Seeman, E. Structural and biomechanical basis of sexual dimorphism in femoral neck fragility has its origins in growth and aging. *J Bone Miner Res* **18**, 1766-74 (2003).

29. Karasik, D. & Ferrari, S.L. Contribution of gender-specific genetic factors to osteoporosis risk. *Ann Hum Genet* **72**, 696-714 (2008).

30. Ohlsson, C. *et al.* Genetic determinants of serum testosterone concentrations in men. *PLoS Genet* **7**, e1002313 (2011).

31. Styrkarsdottir, U. *et al.* European bone mineral density loci are also associated with BMD in East-Asian populations. *PLoS One* **5**, e13217 (2010).

32. Wood, A.R. *et al.* Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Human molecular genetics* **20**, 4082-92 (2011).

33. Mosekilde, L., Torring, O. & Rejnmark, L. Emerging anabolic treatments in osteoporosis. *Curr Drug Saf* **6**, 62-74 (2011).

34. Hasselblad, V. & Hedges, L.V. Meta-analysis of screening and diagnostic tests. *Psychol Bull* **117**, 167-78 (1995).

35. Harris, S.T. *et al.* Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. Vertebral Efficacy With Risedronate Therapy (VERT) Study Group. *JAMA* **282**, 1344-52 (1999).

36. Jackson, R.D. *et al.* Calcium plus vitamin D supplementation and the risk of fractures. *N Engl J Med* **354**, 669-83 (2006).

37. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).

38. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet* **3**, e114 (2007).

39. 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* **39**, 906-13 (2007).

40. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).

41. Estrada, K. *et al.* GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* **25**, 2750-2 (2009).

42. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).

43. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).

44. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155-66 (2001).

45. Ioannidis, J.P., Thomas, G. & Daly, M.J. Validating, augmenting and refining genome-wide association signals. *Nature reviews. Genetics* **10**, 318-29 (2009).

46. Pereira, T.V., Patsopoulos, N.A., Salanti, G. & Ioannidis, J.P. Discovery properties of genome-wide association signals from cumulatively combined data sets. *Am J Epidemiol* **170**, 1197-206 (2009).

47. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).

48. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).

49. Reppe, S. *et al.* Eight genes are highly associated with BMD variation in postmenopausal Caucasian women. *Bone* **46**, 604-12 (2010).

50. Ge, B. *et al.* Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* **41**, 1216-22 (2009).

51. Montgomery, S.B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-7 (2010).

52. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat Genet* **39**, 1217-24 (2007).

53. Grundberg, E. *et al.* Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet* **7**, e1001279 (2011).

54. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423-8 (2008).

55. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).

CHAPTER 4.3

Assessment of Gene-by-Sex Interaction Effect on Bone Mineral Density

Submitted

CHAPTER 4.4

Genome-wide CNV association study identifies a rare deletion on 6p25.1 associated with fractures

Submitted

CHAPTER 5

GENERAL DISCUSSION

In this thesis, genetic variants were tested to characterize the genetic architecture of osteoporosis-related traits like BMD, height and fracture risk. The main objective of this thesis was to identify novel genetic risk factors for osteoporosis-related traits using the hypothesis-free Genome-wide Association Study (GWAS) approach. This chapter places the main findings of this thesis under a unifying perspective together with a discussion on methodological considerations. Finally, recommendations for future research on the genetics of osteoporosis are presented.

MAIN FINDINGS

Analytical Methods for GWAS

At the time when this thesis was started (2008), the first meta-analysis of GWAS data from large cohort studies was setting up the grounds for an explosion of genetic discoveries[1] . After four years, in 2012, the NHGRI GWAS Catalog www.genome.gov/gwastudies. (Accessed in 04/01/2012) contains 6,897 genetic variations identified via 1,130 published GWAS papers. Considering that before 2006 only a handful of common genetic variants were successfully replicated in candidate gene approaches, these numbers illustrate the advantage as in the highly-powered hypothesis-free GWAS approach. Most of the recent papers incorporate the use of imputed data using a common reference panel such as the International HapMap which contains ~ 2.5 million SNPs in the CEU panel. In **Chapter 2.1** we described a web- and grid-based solution GRIMP which allows the parallel analysis of multiple phenotype from different studies. This user-friendly interface provides access to distributed computing resources for primarily biomedical researchers with or without experience in the field, but who still need to run analyses with extreme computational demands. This solution has enabled the genetic discoveries presented in this thesis and also of many others which are not presented. GRIMP also ensures that regardless of the number of markers, phenotypes or covariates, the association analysis will be performed in less than 3 hrs, thus setting the grounds for future imputed datasets of even larger size.

Genetics of human stature

Human stature is a classic phenotype that has inspired researchers since several decades.[2,3] It is not only a precise and easily measured trait but it is also

highly heritable (h^2~80%)[4]. Therefore, one could say that the study of the genetic architecture of human height can serve as a model for other quantitative human traits. From the osteoporosis-research point of view, height is also a very important bone trait as it is associated with increased risk of fracture, independently of BMD. When this thesis was started, 44 genetic variants had been identified arising from five GWAS of height.[5-8]

In **Chapter 3.1** we aimed to identify common genetic variants associated with height in a GWAS run on a geographically restricted setup including 10,074 individuals of Northwestern European ancestry with replication of the most significant findings in 6,912 individuals. In this effort, we identified a novel genetic variant (rs6717918) on chromosome 2q37.1 associated with height at $P=3.4x10^{-9}$. This variant is located near the natriuretic peptide precursor type C (*NPPC*) gene. In addition, we also found a SNP (rs10472828) located on 5p14 near the natriuretic peptide receptor 3 (*NPR3*) gene, encoding a receptor of the *NPPC* ligand, associated with body height. Finding genetic variants mapping near both the ligand and the receptor of a given gene associated with a complex trait points strongly to the identification of the underlying biology. Our finding of involvement of the NPPC/NPR3 pathway is corroborated by the observation that *NPR3* knock-out mice were previously found to have significant skeletal overgrowth.[9] After our paper was published, it was shown that the product of *NPPC* (CNP) can rescue impaired growth and early death of chondrodysplastic CNP knockout mice.[10,11] Finally, a study on rodent pups with GH deficiency showed evidence supporting the use of CNP peptide concentrations as biomarkers of linear growth.[12] Taken together our results and those in mice models, it has been proposed that CNP-based therapies may be developed in the future for treating short stature syndromes.[13] Considering that the common genetic variant we identified confers a height change of 0.44 cm per allele, these results highlight that the identification of genetic loci with common gene variants might be of more value for the development of future treatments and/or biomarkers rather than for risk prediction applications.

In **Chapter 3.2** we increased the discovery sample size by one order of magnitude compared to the previously described study. This was only possible with the collaboration of the **G**enetic **I**nvestigation of **AN**thropometric **T**raits (**GIANT**) consortium which by that time included a Stage I discovery set of 46 GWAS (133,653 individuals) plus 50,074 individuals of European ancestry in the Stage II replication. In this chapter we showed that hundreds of variants in at least 180 loci (the ones reaching $P < 5x10^{-8}$) are associated with height. While the 180 loci explain on average 10% of the phenotypic variation of height, including markers associated at lower significant levels ($0.05>P>5x10^{-8}$) can increase the variance explained to 13.3%

(range 9.7–16.8%). In this chapter, we also showed that the 180 identified loci are not randomly distributed in the genome, have an overrepresentation of factors involved in monogenetic syndromes and show a significant level of connectivity based on known biological function described in published literature (GRAIL P=0.016). Another important observation from this research is the broad presence of secondary signals in at least 10% of the loci which may be driven by allelic heterogeneity. Methodological considerations of the secondary signals analysis are described further below. The results shown in this chapter are not only of relevance for the study of the genetic architecture of adult height but also for the study of common complex diseases in general.

Genetics of Osteoporosis

In chapters 4.1, 4.2 and 4.3 we report results from three meta-analyses of GWAS on BMD. **Chapter 4.1** used the setup of the Genetic Factors for Osteoporosis (GEFOS) consortium including 5 studies with GWAS data on Femoral Neck (FN-) and Lumbar Spine (LS-) BMD. In this, the first attempt to capture common variation underlying BMD using imputed and genotyped data we increased the number of BMD loci from 7 to 20. These 13 novel loci included: 1p31.3 (*WLS*), 2p21 (*SPTBN1*), 3p22 (*CTNNB1*), 4q21.1 (*MEPE*), 5q14 (*MEF2C*), 7p14 (*STARD3NL*), 7q21.3 (*FLJ42280*), 11p11.2 (*LRP4*, *ARHGAP1*, *F2*), 11p14.1 (*DCDC5*), 11p15 (*SOX6*), 16q24 (*FOXL1*), 17q21 (*HDAC5*) and 17q12 (*CRHR1*). In this chapter we started identifying further members of the Wnt pathway in addition to the well known *LRP5* gene and including: *CTNNB1*, *WLS* (previously known as *GPR177*) and *LRP4*. The variance explained by these loci was 2.9% for LS-BMD and 1.9% for FN-BMD. In this study, we tested the association of the genetic risk score with fracture risk in a limited subset of 4,865 individuals of the Rotterdam Study I. Allele scores for LS- and FN-BMD were significantly associated with vertebral and non-vertebral fractures, respectively. Evidence for site-specificity was found for 3 loci: *MEF2C* and *SOX6* for association with FN-BMD, and *RANKL* with LS-BMD. Such skeletal site-specificity is expected given that the genetic correlation between the two BMD measurements is considerably less than 1.

To overcome the low powered setting to evaluate risk of fracture, in **Chapter 4.2** we expanded the GEFOS consortium to 17 GWAS studies of BMD and in addition included 34 replicating studies of the GENOMOS consortium with DNA and no GWAS information. In the Stage I + Stage II meta-analysis 83,894 samples were combined for FN-BMD and 77,508 for LS-BMD. Finally, all independent variants with a $P < 5x10^{-6}$ (including known and novel) were tested in 31,016 any type of fracture

cases and 102,444 controls. As a result we identified 56 loci (32 novel) associated at genome-wide significant with either FN- or LS-BMD. The first male-specific BMD variant was found mapping in the Xp22.31 chromosome. Following the trend of the previous chapter, we identified six additional loci displaying site-specificify: 2q14 (*INSIG2*), 12p11.22 (*PTHLH*) and 16q12.1 (*CYLD*) for LS-BMD and 8q13.3 (*LACTB2*), 10p11.23 (*MPP7*) and 10q22.3 (*KCNMA1*) for FN-BMD. More importantly, given the higher powered setting, we were able to identify fourteen BMD loci also associated with fracture risk at $P<5x10^{-4}$, of which six reached $P<5x10^{-8}$ including: 18p11.21 (*FAM210A*), 7q21.3 (*SLC25A13*), 11q13.2 (*LRP5)*, 4q22.1 (*MEPE*), 2p16.2 (*SPTBN1*) and 10q21.1 (*DKK1*). We also evaluated the significance of connectivity between candidate genes underlying the BMD associations in the same way as in chapter 3.2 for Height and found an even more significant connectivity between the identified loci ($P < 0.0005$ for BMD, $P=0.016$ for Height). This time the relevance of the WNT pathway for BMD determination was highlighted by the identification of 6 additional member of the WNT pathway: *WNT5B*, *WNT16*, *DKK1*, *PTHLH, SFRP4 and AXIN1.* These results show that the BMD is not only a good predictor of fracture but in fact, a great phenotype characterizing bone biology. In this effort we also applied the same procedure as of chapter 3.2 to identify secondary signals and found that 8 of the BMD loci (14% of the total) showed at least one additional independent signal. Novel methods for the identification of additional independent signals may increase the proportion of loci with secondary signals (See discussion below on *Methodological considerations*). The identification of these 56 loci did not lead to a significant improvement in prediction of either osteoporosis nor fracture risk on top of the known (non-genetic) risk factors, suggesting that many more need to be identified. In line with the observation of potential therapeutic application for other genes discovered by GWAS, we also noted that several of the current osteoporosis drug targets are also identified by our GWAS, such as *PTHLH*, *DKK1*, *RANKL*, *OPG*, and *SOST*. In this chapter we also investigated the effect of regional imputation using a denser reference panel (1000Genomes June06 Pilot version) further discussion on these denser panels can be found in the *Methodological considerations* section below.

The contribution of sex-specific genetic factors for osteoporosis has been widely postulated.[14] While on chapter 4.2 we found indication of sex-specificity in only one locus at the X chromosome, no autosomal variant has been identified to be either sex-specific or more formally, with a gene by sex interaction. Therefore, in **Chapter 4.3** we examined genome-wide gene-by-sex autosomal interactions for LS-BMD and FN-BMD including 25,000 individuals from seven GWAS of BMD. Twelve SNPs

with an interaction P < $1x10^{-5}$ were followed-up in 24,000 additional samples. In this study we were not able to identify any autosomal locus reaching the strict P < $5x10^{-8}$ threshold for gene-by-sex interaction, which means that if any autosomal loci has a significant interaction with sex, it must have a very small effect (i.e., less than 0.08% of the variation of these traits). We discuss the power issues in similar scenarios such as gene x environment GWAS in the *Methodological considerations* section below.

All previous chapters focused on the study of common variation, in **Chapter 4.4** we aimed to evaluate the contribution of another type of genetic variants, the Copy Number Variations (CNVs) which are a type of structural variants of the genome in which large (>1Kb) segments of the genome are either lost or duplicated. Considering that common CNVs are well tagged by common SNPs,[15] we sought to identify low-frequency CNVs in 5,178 individuals from the Rotterdam Study I cohort as a discovery set and then assessed the association with osteoporotic fractures. A rare 210Kb deletion in 6p25 was identified in the RSI study and further replicated in other array-based studies (p=0.02). We attempted replication by sequencing the breakpoints and genotyping the deletion in ~10,000 cases and 16,542 controls from 15 studies across Europe and Australia, however, using this method only 6 from the 9 studies identified the deletion compromising the power of the replication phase. There is some indication that the prevalence of the deletion is geographically. Despite large sample sizes no single deletion was identified in some countries (i.e. ~38,000 subjects in the DeCode study from Iceland, 4000 subjects in UFO study from Sweden) while successful identification of the deletion was observed in samples from Spain, Netherlands and England. This study highlights the importance of regional replication of such low-frequency variants in studies of genetic epidemiology. Technical aspects of reasons for not replication are discussed on the paper.

The studies described in this thesis cover only a part of the recent research of the genetics of osteoporosis-related traits, mostly describing the state of the art of the genetics of BMD and height. We and others have published more than 20 different GWAS for osteoporosis-related traits in the last 5 years. Published results have been summarized in recent reviews.[16-18]

METHODOLOGICAL CONSIDERATIONS

In chapter 3.2 we showed that the largest study to date of common genetic variants associated with adult height identified 180 loci which together can explain 10% of the variance of this classic phenotype. If we consider that the heritability of this

trait is 80%, we have explained 10%/80% = 12.5% of the narrow sense heritability of height. By using a different methodology, the group of Peter Visscher has suggested that common SNPs explain a larger proportion (45%) of the heritability of height.[19] This method does not identify the exact loci underlying the association, but rather estimates a genomic-average by using all the SNPs in the microarray. In this section we discuss, among other issues, the research areas under development that may help for the identification of genetic variants and/or increase the proportion of variance explained for traits similar as the ones discussed in this thesis.

Improvements in discovery sample size

Increasing sample size is clearly a key factor for the identification of novel genetic factors of complex diseases. This is evident from results in this thesis from restricted geographic studies in only 10,000 individuals in the discovery (chapter 3.1) to a study of major scale in 130,000 individuals in chapter 3.2, allowing to increase dramatically the number of identified novel loci. The same applies to the study of BMD. In Figure 1, we have summarized the number of loci discovered for BMD, Height and BMI. One can see that the genetic architecture of BMD and Height seems to be similar while BMI shows a different pattern. There are four main reasons that might explain the success on the discovery of GWAS loci for these traits as compared to BMI:

1) Both BMD and Height are more heritable than BMI
2) BMD and Height are more precise and stable measurements
3) These phenotypes are easily standardized across studies.
4) Possible differences in the genetic architecture (common vs rare variants)

We have demonstrated in chapter 3.2 that just increasing the sample size to a hypothetical number of half million samples with GWAS on height would increase the number of height loci to ~ 700, but these would only explain ~15% of the variance. However, that analysis does not take into account variants of weaker effects than those detected there. Using the same method[20], we predicted using the results from chapter 4.1 that around 44 loci explaining 7% of the BMD variance would be identified in a meta-analysis including 100,000 samples (Fig. 2).[21] In fact, these predictions were validated in chapter 4.2, however we noted some limitations of this method. First, only variants with at least 1% of power of detection can be included, otherwise over-inflation occurs. Consequently, this method only predicts the discovery of variants for which the individual effect sizes are equal or larger than

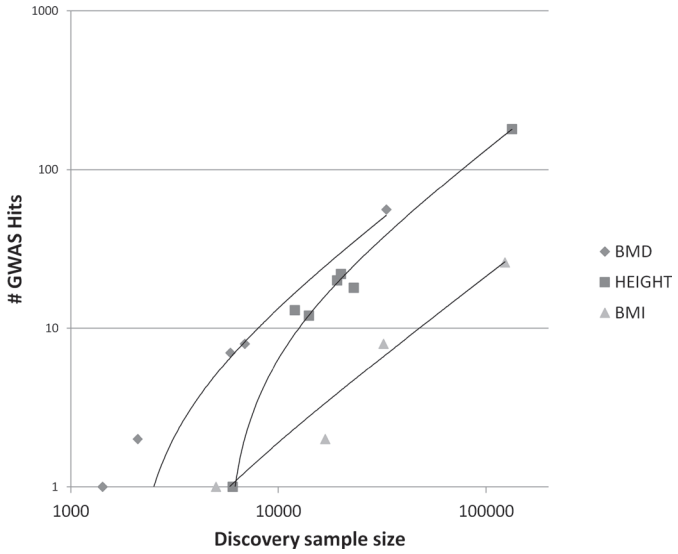


Figure 1. Number of loci identified as function of GWAS discovery sample size. Examples for studies of BMD, Height and BMI

ones used for the prediction. Given the complexity of even larger meta-analyses, other options need to be considered in addition to increasing the sample size.

Additional variants in the same locus

As discussed previously, in chapters 3.2 and 4.2 we showed that about 10 to 14% of the known Height and BMD loci contains at least an additional variant independently associated with the trait. We say at least one because the method we used to assess independence adjusted the phenotypes against the effect of the already identified genetic variants. In the second univariate association, multiple SNPs can still be independently associated with the trait but only the most significant is usually selected as the "secondary signal". While additionally independent variants may exist in the same locus, these could not be fully identified without reiterating the conditional analysis. An alternative would be to use multivariate analysis but this can be extremely computationally intensive at a genome-wide scale. A novel method to identify additional variants in the same locus has been recently proposed.[22] This method requires only individual level data of one of the largest GWAS included in the meta-analysis is required to approximate the conditional analysis from summary-level statistics from a GWAS meta-analysis. Using this new method, 36

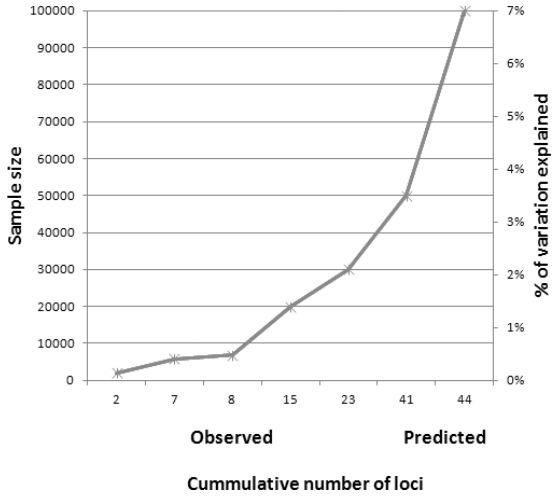


Figure 2. Prediction of the number of loci identified depending on the GWAS discovery sample size

height loci show 49 additional SNPs associated with height, which doubles the variance explained by these 36 loci. Another point raised by that study[22] is that in some genomic regions, markers can be still in strong LD while being at a distance of more than 1 Mb away. At least 20 regions have been reported with this long range LD patterns.[23] Other important aspect of doing multivariate association analysis is that when pairs of associated SNPs have some level of correlation, the marginal SNP effects can be either under or overestimated.[24] Also, it has been shown that some of those secondary signals may be explained by a 3rd not yet genotyped marker.[24] Further, the study of secondary signals might be especially important in the context of potential interactions between a common regulatory variant and rare coding variants as shown for eQTLs[25] but also recently described for some complex disorders such as Hirschsprung disease[26]. This type of genetic interaction between variants is so far the only one showing conclusive results as the search of other gene by gene interactions has been so far inconclusive.

Interaction of gene variants with other risk factors

The large majority of GWAS studies have focused on the study of main effects of genetic variants across pooled samples. The term “Gene x Environment” interaction refers to the situation when a environmental risk factor acts as an effect modifier of a genetic risk factor. Sex is probably the strongest genetic phenotype (at least in mammals), so in general terms it should not be considered an “environmental” risk factor. Sex is also one of the strongest determinants of osteoporosis (56% of women

above 50 have osteopenia or osteoporosis vs 18% in men) and many other human conditions such as human stature (males are on average 20 cm taller than females). Despite these strong differences across sexes, and the fact that we have performed formal tests for interaction effects in autosomal chromosomes (chapter 4.3), we have only identified one X-linked variant (chapter 4.2) showing significant difference in effects between females and males. The same genetic variant on chromosome Xp22.31 has been associated with testosterone levels in males. One could then assume that genetic variants affecting sex-hormone levels such as testosterone for males and estradiol for females may be key regulators of such strong sex differences. However, the BMD genetic variants found at 6q21 near *ESR1* (estrogen receptor alpha) were equally associated in both females and males. The presence/absence of sex-specific loci may be also trait dependent. In the 2nd round of Giant meta-analysis where three main traits were analyzed: Height[27], BMI[28] and waist-hip-ratio[29], only the latter found (seven) loci with significant differences in effect sizes between females and males. This should reflect underlying biology since we can rule out power issues as the waist-hip-ratio was the smallest meta-analysis of the three. Interestingly, one of those seven loci is a BMD locus, (*RSPO3).* Variants in this locus were also strongly associated with FN-BMD in females, but not in males, however the test for sex heterogeneity for FN-BMD was only marginally significant (P=0.04). The development of better methods and the systematic evaluation of sex chromosomes is required to identify further gene interactions with other risk factors.

Combining meta-analysis of related traits

Pleiotropic effects, that is a gene or gene variant having an effect on two or more traits, is a common feature of human traits.[30] By looking at the genome-wide significant hits for height and BMD reported in chapters 3.2 and 4.2 we identified five loci where the same variant is associated with both BMD and height (rs479336 near *DNM3*, rs1366594 near *MEF2C*, rs6959212 near *STARD3NL* and rs7851693 near *FUBP3*, rs7084921 near *CPN1*). Interestingly, three of those variants are also associated with fracture risk at P $<5x10^{-3}$ (Table 1). One can exploit this pleiotropic properties to identify additional risk variants by combining two GWAS of disease-related traits.[31] Doing such bivariate or multivariate analysis normally requires individual level data. A novel method has been recently proposed to use the GWAS meta-analysis results of two traits to combine them while accounting for any possible underlying correlation amongst them.[32] Such methods would identify even more pleiotropic loci beyond those already identified at genome-wide significance level.

One limitation of these methods is the assumption that the same variant or a highly correlated one must be highly associated with both traits. There are two loci where two variants show an independent effect on BMD and Height (6q25.1 near *ESR1*, 6p21.1 near *RUNX2*). In such scenarios of both allelic heterogeneity and pleiotropy, different methods for identifying pleiotropic loci need to be developed.

Improvements in density and sample size of reference panels

All the GWAS meta-analyses presented in this thesis were based in imputation using HapMap II CEU as reference panel. This dataset contains ~ 2.5 Million common snps genotyped in 60 individuals. Thus, not very surprisingly, most of the genetic variants found here and in other GWAS are common. A debate has started on the relevance of common variants and has "divided" the field of human geneticists into those considering that rare variants would have a greater impact on phenotype variation[33], in contrast to those considering the common disease, common variants approach[34]. The problem is that rare variants have not been thoroughly assessed by HapMap and where available are generally speaking difficult to impute. It has been shown that to improve the imputation quality for low-frequency variants (MAF < 5%), the best approach is to use a large and diverse reference set[35]. Despite the HapMap III covered a diverse set of populations in a considerable sample size (1000 samples), the coverage of that reference panel was lower than the HapMap II (1.5 M vs 2.5M) and still focused largely on common variation. To overcome this gap, the 1000 Genomes project started in 2008, and has effectively shown that by means of low-coverage whole-genome sequencing one can generate a better reference panel for imputation. The first pilot releases of the 1000G data (March 2009 and June 2010) were constructed with the whole-genome sequence of the same 60 CEU individuals genotyped in HapMap II. Therefore, there were no major improvements in either the imputation quality or the coverage of rare variants (with 120 haplotypes there is very limited power to find variants with a MAF <2.5%) . In chapter 4.2 we evaluated the improvement in test statistics by regional imputation using the 1000G June 2010 release in the 56 BMD loci. In 13 of these loci we identified at least one variant with a p-value of an order of magnitude or more significant than the original variant identified in the HapMap II discovery meta-analysis. We also used this data to suggest putative functional variants within the BMD loci. Again not surprisingly given that the same 60 samples were used in the reference panel, there were no low-frequency variants detected in this analysis.

In February 2012 the version 2 of Phase I data of 1000 Genomes was released. This Phase I data consists on both low-pass whole-genome and high-pass exome-

Table 1. Loci associated with both BMD and Height at Genome-wide significant level

				FNBMD (n=83,894)		LSBMD (n=77,508)		Height (n=183,727)		Fracture Risk (n=130,000)	
Locus	SNP	Gene	A*	Beta	P-value	Beta	P-value	Beta	P-value	OR	P-value
1q24.3	rs479336	*DNM3*	T	-0.04	8.51E-15	-0.03	2.14E-05	-0.04	1.40E-23	1.01	0.36
5q14.3	rs1366594	*MEF2C*	C	-0.08	4.47E-61	-0.01	0.01	-0.03	2.00E-18	1.03	2.72E-03
7p14.1	rs6959212	*STARD3NL*	T	-0.04	1.18E-13	-0.07	3.76E-38	-0.02	1.60E-09	1.05	7.16E-05
9q34.11	rs7851693	*FUBP3*	G	-0.05	3.37E-22	-0.03	6.08E-08	-0.03	2.60E-17	1.05	3.54E-05
10q24.2	rs7084921	*CPN1*	C	-0.03	9.03E-10	-0.03	9.15E-07	0.03	1.60E-13	1.03	0.21

**Effect allele*

Table 2. Number of samples sequenced at low-coverage in the 1000G Phase I project.

Acronym	Population	Samples	Ethnic background
ASW	African-Americans (SW US)	61	AFR
LWK	Luhya individuals	97	AFR
YRI	Yoruba individuals	88	AFR
CLM	Colombian in Medellin, Colombia	60	AMR
MXL	Mexican individuals from LA	66	AMR
PUR	Puerto Rican in Puerto Rico	55	AMR
CHB	Han Chinese in Beijing	97	ASN
CHS	Han Chinese South	100	ASN
JPT	Japanese individuals	89	ASN
CEU	CEPH individuals from Utah	87	EUR
FIN	Finnish individuals from Finland	93	EUR
GBR	British individuals from England	89	EUR
IBS	Iberian populations in Spain	14	EUR
TSI	Toscan individuals	98	EUR
	Total	1,094	

sequence of 1,094 individuals (Table 2). Besides being the largest reference set in terms of samples, this reference panel contains a large number of SNPs (~39M) and other type of genetic variants such as Indels and large deletions (Table 3).

The methods to impute and test those indels and deletions into the GWAS data are still under development. A large proportion (~50%) of these variants are rare (MAF < 1%), actually many are singletons (only found in one sample), thus making them very difficult to be imputed. Using the 1000G Phase I alpha release, we imputed 11.2 million autosomal SNPs with a sufficient imputation quality (MACH_RSQ > 0.5) in the RSI cohort. Half of those 11.2 million SNP, had a MAF < 5% (data not shown).

A recently proposed approach employs next-generation sequencing of subsets of individuals (such as extreme cases and controls) and uses those novel variants to impute the full GWAS cohort.[36,37] Such approach has already been deployed to identify low-frequency variants in individuals of Iceland by the deCode study.[37-39]

Another option is to generate regional reference panels to improve the chances of finding low-frequency variants. The Genome of the Netherlands (GoNL, http://www.dutchgenomeproject.com/)[40] is currently in the final analysis phase of 250 trios (500 independent samples) with whole-genome sequence at 12x. Also, the Rotterdam Study is currently assembling a dataset of 3000 random samples with exome-sequencing at an average 30 x coverage. These resources will create a denser reference panel which may help identifying low-frequency variants in individuals of Dutch origin.

Clinical applications – Risk Prediction

One of the expectations from genetic studies of common diseases is to be able to improve risk prediction.[41] We have shown that in general there is little added value in terms of prediction so far. This is due to several factors, the most important being that the genetic variants we have identified explain less than 10% of the phenotypic variance of these osteoporosis-related traits. In chapter 4.2 we used the BMD association results to create a genetic risk score to categorize individuals in five groups of low, low-normal, normal, high-normal and high genetic risk score bins. There we saw that the group with a high genetic risk score had an ORs of 1.56 for osteoporosis and 1.6 for fractures when comparing them with ”normal” risk score involving 34% of the population. In the same way the group with a low genetic risk score had an ORs of 0.40 for osteoporosis and 0.5 for fracture risk, or protective ORs of 2.62 and 2 for osteoporosis and fracture risk, respectively. So it seems that the accumulation of BMD-increasing alleles in average has a higher “protective” effect

Table 3. Type and number of genetic variants included in the Phase I v2 of the 1000 Genomes Project.

Type of variant	No. variants
Autosomal SNPs	38,200,000
X-chromosome SNPs	1,451,000
InDels	3,800,000
Deletions	14,000

than the accumulation of BMD-decreasing alleles. It is also possible that at some point in the distribution the accumulation of protecting common alleles would have a higher effect than the accumulation of risk alleles. A similar behavior has been recently reported in a study on individuals of the extremes of the height distribution.[42] In that paper it is shown that while the polygenic model can explain the associations in the 1% tails of height of extremely short individuals, their observed data indicate that the polygenic model breaks near the 0.25% lowest percentile. Using simulated data, it has also been shown that for those extremely short individuals it would be more likely that 10 rare variants with a modest effect of 1 SD could explain this behavior than a combination of up to 180 additional common variants. Sibling analysis in the same paper suggested that either de-novo and/or recessive models would better explain the observed data. These results fit with another recent paper from the same group showing an excess of rare large deletions associated with short stature.[43]

Clinical applications – Identifying potential drug targets

Bone remodeling repair micro-damages and provides calcium from bone to preserve homeostasis. This dynamic process enables interventions that limit resorption (antiresorptive therapy) or augment formation (anabolic therapy). Treatment with antiresorptive agents (such as bisphosphonates and RANKL inhibitors) inhibits bone resorption by osteoclasts. Therefore, these treatments only prevent further loss of bone and they do not stimulate new bone formation.[44] As of today, the only FDA-approved anabolic therapy is recombinant parathyroid hormone (PTH). However, PTH is a protein that needs to be administered subcutaneously, it

has potential side effects (e.g., hypercalcemia and hypercalciuria) and the duration of treatment with this drug is limited to a maximum therapy duration of 18 months in Europe and 24 months in the US, because rodents administered high doses of PTH were shown to develop osteosarcomas.[44-46]

We have identified several new candidate genes related to either height and/or bone mineral density. Our results for BMD show that several of the current drug targets for osteoporosis are detected in our unbiased GWAS approach. Given the large number of known and newly discovered loci, we propose seven points which might help prioritize them for potential clinical use.

Identify causal variant. Identifying a significant associated SNP does not necessarily mean the marker itself is the functional allele that causes disease risk. It is more likely that the observed association at a marker might be the result of linkage disequilibrium (LD) of the associated variant with the causal allele. Dense genotyping of markers in an associated locus can refine the number of potential candidates. Other option is to use studies of different ethnic backgrounds (individuals from Caucasian or Asian ancestry have wide LD blocks, individuals from African ancestry have smaller LD blocks), to refine the association signal. High-throughput sequencing in sufficient sample sizes might help to investigate if rare variants are driving the association in a given locus.

Identify causal gene. Only 1% of the human genome codes for proteins. An association in this region might be more promising to identify the underlying gene driving the association. However, only about ~12% of the GWAS hits can be explained by coding markers.[47] Further, the GWAS approach does not pinpoint genes directly. For the majority of the intronic or intergenic GWAS hits other options such as eQTL analysis might help to pin point candidate genes. Two main types of eQTLs can be used for this purpose, the most common is the SNP vs gene expression eQTL mapping. But one can also use direct association of gene expression patterns vs the phenotype of interest. In chapter 4.2 we show that both techniques can provide complementary answers to candidacy.

Animal model with relevant phenotype. Sometimes the closest gene or a nearby has a knockout mice with a relevant phenotype to the one studied in the GWAS. That was the case of the height-associated *NPPC* signal explained in chapter 3.1 among others. The availability of such animal data with a biological involvement can give candidate genes a higher priority.

Known Mendelian forms of disease. Further candidacy of nearby genes can be enriched by the identification of known variants associated with Mendelian forms of the studied disease. As we pointed out in the introduction, several genes were identified in severe forms of bone diseases. We note that for most of them (*LRP5*,

COL1A1, *SOST*, *CLCN7*) common markers in or nearby those genes are also associated with normal BMD variation. For height, we found that the 180 loci contained 21 of 241 genes found to underlie human syndromes characterized by abnormal growth.

Tissue specificity of candidate gene expression. Many of the GWAS signals can point to a well known candidate gene, however if the expression of that gene is ubiquitous, then a potential therapy on that gene product may have many side effects. For osteoporosis treatment, genes that are mostly acting in bone-active cell types such as osteocytes, osteoblasts and osteoclasts are preferred for further follow-up research. Tissue specific expression profiles can be a useful way to approximate therapeutic applications.

Identifying biological mechanism. As the number of significant loci increases one can make us of computational methods such as GRAIL to identify main pathways involved in the phenotype of interest. The overrepresentation of genes in a given network can predict which are the most candidate genes and the mechanism underlying the association.[48] The identification of underlying pathways was a hallmark with Wnt and TGT-Beta arising as prominent for BMD and height, respectively.

Summarizing evidence to select candidates. By identifying which loci have an enrichment of the points previously described can help in prioritizing genes for further investigation in clinical settings. While an equal weight to each point can be given, one could argue that some points such as having tissue specificity and/or having a known Mendelian form of disease may have a higher weight than knowing the causal variant.

Table 4 contains a list of candidate genes for osteoporosis treatment based on the categories mentioned above. This list was obtained from the GWS hits for BMD described in chapter 4.2. High in the rank is a well known bone-related gene, *LRP5*. Both *SOST* and *DKK1* suppress WNT signaling by binding to different domains of *LRP5*. The expression of *DKK1* in multiple tissues might eventually limit the efficacy of *DKK1*-targeted therapies. On the other hand, *SOST* is primarily expressed in bone tissue (mainly osteocytes). Inhibiting sclerostin is an attractive mechanism to stimulate bone formation and is currently progressing preclinical testing.[49] Another gene identified as associated with BMD in chapter 4.1 (*MEF2C*) has been recently shown to be essential for the transcription of *SOST*.[50,51] However, *MEF2C* is expressed in many different tissue types, thus its use as drug target is probably limited by a higher propensity to see side effects as seen for *DKK1*.

The next two candidates on Table 4 are *TNFRSF11A* and *TNFRSF11B* which code for RANK and OPG, respectively. RANK is activated by RANKL. Denosumab inhibits this maturation of osteoclasts by binding to and inhibiting RANKL. This protects bone

from degradation and helps counter the progression of osteoporosis.[52] The drug therefore mimics the effects of OPG. Denosumab has been recently approved to treat osteoporosis in the Netherlands and in other countries. The last two candidates genes are *SP7* and *SOX9* which code for the transcription factors Osterix and Sox9 which are master transcription factors for osteogenesis and chrondrogenesis respectively. Both have in common their interaction with another transcription factor, *RUNX2*. The interplay between *SOX9*, *RUNX2* and *Osterix* regulates the fate of the mesenchymal cells into osteoblasts or osteocytes.[53] To our knowledge, these molecules are currently not being considered as potential targets for drug treatments despite their high biological relevance. While *RUNX2* is mostly bone-specific expressed, it has also been shown that it can suppress breast cancer proliferation.[54] These dual beneficial effects need to be taken carefully into account. A previously approved drug for the treatment of breast cancer (tamoxifen) was also found to be useful for the treatment for osteoporosis. However, tamoxifen was later linked to increased risk of endometrial cancer and therefore it is currently only approved for patients with high risk of breast cancer recurrence.[55] Raloxifen, which is also an estrogen receptor modulator is currently used for the treatment of osteoporosis.

We also note that other interesting loci might also be those with a significant fracture risk association even though not much of their biology is known. That is the case of the two more significant fracture loci (near the genes *FAM210A* and *SLC25A13*) identified in chapter 4.2 (Table 5). These discoveries open completely new research lines that warrant further investigation to fully characterize the biological mechanisms underlying the observed associations.

The scope of this thesis focused on providing additional genetic factors for adult stature and osteoporosis. So far we have only identified genetic variants (mostly SNPs) that are nearby potential candidate genes. Further research is required to demonstrate that the candidate gene is the one driving the association. Additional research would be required to find potential drug compounds. We envision that specialized groups will take the results driven by this thesis to propose novel drug compounds for osteoporosis. We recently learned that such is the case for the research of future therapies for mental disorders where pharmaceutical companies, such as Novartis, GlaxoSmithKline and AstraZeneca, are shutting down the traditional neuroscience research divisions globally because of the expensive costs and little success of their clinical trials.[56] Instead, these companies are partnering with experts in the field of genetics of psychiatry and cognitive disorders to create a new research division that will follow the leads of recent genetic studies driven by GWAS, sequencing and copy number analyses.[56]

Table 4. Candidacy of genes for osteoporosis for further research based on biological evidence

SNP associated with BMD	Candidate gene	Closest gene	eQTL	Knockout mouse with skeletal phenotype (MGI)	Monogenic syndrome with skeletal phenotype (OMIM)	Potential Funct. variant	GRAIL Priority	Pathway	Treatment Available?
rs3736228	LRP5	YES		1278315	259770 607634	Same	YES	Wnt	Monoclonal antibodies for DKK1 and SOST
rs884205	TNFRSF11A	YES	NM_003839	1314891	602080		YES	OPG/RANK/ RANKL	RANKL inhibitor
rs2062377	TNFRSF11B	YES		109587	602080		YES	OPG/RANK/ RANKL	RANKL inhibitor
rs2016266	SP7	YES		2153568	613849		YES	MSC differentiation	NO
rs7217932	SOX9	YES		98371	608160		YES	Endochondral Ossification	NO

Table 5. Candidacy of genes for osteoporosis for further research based on their association of variants with fracture risk

SNP associated with BMD and Fracture	Candidate gene	Closest gene	eQTL	Knockout mouse with skeletal phenotype (MGI)	Monogenic syndrome with skeletal phenotype (OMIM)	Potential Functional variant	GRAIL Priority	Pathway	Treatment Available?	Fracture Risk P-val
rs4796995	*FAM210A*	YES							NO	**8.8×10^{-13}**
rs4727338	*SLC25A13*	YES							NO	**5.9×10^{-11}**
rs1373004	*DKK1*			1329040				Wnt	Anti-DKK1	**9.0×10^{-9}**
rs3736228	*LRP5*	YES		1278315	259770 607634	rs3736228	YES	Wnt	Anti-DKK1 and anti-SOST	**1.4×10^{-8}**
rs6532023	*SPP1*		NM_000582	98389			YES	Endochondral Ossification	NO	**1.7×10^{-8}**
rs4233949	*SPTBN1*	YES				rs4305309			NO	**2.6×10^{-8}**

FUTURE DIRECTIONS

In this thesis we have shown that hundreds of variants affect adult height and it is the same case for many other complex traits such as bone mineral density. We have applied the novel hypothesis-free approach of GWAS to identify those loci harboring at least one variant affecting these osteoporosis-related traits. It is expected that many additional variants (both common and rare) in the same locus are independently associated and adding those will increase the percentage of variation explained. Future research to identify these additional variants is currently underway for height in the GIANT consortium using an extended set of studies genotyped with the MetaboChip[57], a microarray enriched with variants identified in different GWAS of metabolic traits. A similar approach may be envisioned to look for other osteoporosis-related traits which are currently under analysis within the GEFOS consortium.

Another possible next step is to increase the coverage of the genotyped data sets. Currently, the most cost-effective option is to use imputation on GWAS data using the sequencing-based reference sets such as the 1000 Genomes Project, or other regional reference panels such as the GoNL project. This will lead to meta-analyses of thousands of samples of > 40 million SNPs. These large samples sizes are required to identify low-frequency variants with subtle effects. However, this solution has the drawback that at some point (i.e. MAF <1%) imputation will not be sufficient to fully characterize those variants. To fully characterize those rare variants de-novo re-sequencing data will be required. The number of sequenced samples required to have enough power to detect significant associations will be probably only be achieved within two to three years if the sequence costs continue decreasing at the same rate as of today.

Finally, better characterization of the novel loci is required; we can speculate that this may be especially relevant for those loci for which we have shown a significant effect on fracture risk which have a higher prior to be important for future therapies for the treatment of osteoporosis and its sequela.

CONCLUDING REMARK

We have expanded our understanding of the genetic architecture of adult stature and osteoporosis. Many more factors still need to be discovered to improve the current prediction models for osteoporosis and fracture risk. Hopefully, the insights provided by our research will eventually be translated into clinical interventions.

REFERENCES

1. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat Genet* **41**, 399-406 (2009).
2. Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* **5**, 329-348 (1885).
3. Fisher, R.A. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918).
4. Visscher, P.M., Medland, S.E., Ferreira, M.A. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* **2**, e41 (2006).
5. Weedon, M.N., Lettre, G., Freathy, R.M. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* **39**, 1245-1250 (2007).
6. Weedon, M.N., Lango, H., Lindgren, C.M. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575-83 (2008).
7. Sanna, S., Jackson, A.U., Nagaraja, R. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* **40**, 198-203 (2008).
8. Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G. *et al.* Many sequence variants affecting diversity of adult human height. *Nat Genet* **40**, 609-15 (2008).
9. Matsukawa, N., Grzesik, W.J., Takahashi, N. *et al.* The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc Natl Acad Sci U S A* **96**, 7403-8 (1999).
10. Yasoda, A., Kitamura, H., Fujii, T. *et al.* Systemic administration of C-type natriuretic peptide as a novel therapeutic strategy for skeletal dysplasias. *Endocrinology* **150**, 3138-44 (2009).
11. Fujii, T., Komatsu, Y., Yasoda, A. *et al.* Circulating C-type natriuretic peptide (CNP) rescues chondrodysplastic CNP knockout mice from their impaired skeletal growth and early death. *Endocrinology* **151**, 4381-8 (2010).
12. Prickett, T.C., Bothwell, J.C., Yandle, T.G., Richards, A.M. & Espiner, E.A. Pharmacodynamic responses of plasma and tissue C-type natriuretic peptide to GH: correlation with linear growth in GH-deficient rats. *J Endocrinol* **212**, 217-25 (2012).
13. Potter, L.R. Regulation and therapeutic targeting of peptide-activated receptor guanylyl cyclases. *Pharmacol Ther* **130**, 71-82 (2011).
14. Karasik, D. & Ferrari, S.L. Contribution of gender-specific genetic factors to osteoporosis risk. *Ann Hum Genet* **72**, 696-714 (2008).
15. Craddock, N., Hurles, M.E., Cardin, N. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713-20 (2010).
16. Zheng, H.F., Spector, T.D. & Richards, J.B. Insights into the genetics of osteoporosis from recent genome-wide association studies. *Expert Rev Mol Med* **13**, e28 (2011).
17. Ralston, S.H. & Uitterlinden, A.G. Genetics of osteoporosis. *Endocr Rev* **31**, 629-62 (2010).
18. Duncan, E.L. & Brown, M.A. Clinical review 2: Genetic determinants of bone density and fracture risk--state of the art and future directions. *J Clin Endocrinol Metab* **95**, 2576-87 (2010).
19. Yang, J., Benyamin, B., McEvoy, B.P. *et al.* Common SNPs explain a large proportion

of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).

20. Park, J.H., Wacholder, S., Gail, M.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42**, 570-5 (2010).
21. Estrada, K. ECTS Meeting abstract. "Discerning the complex genetic architecture of human bone mineral density variation using genome-wide association studies. (2010).
22. Yang, J., Ferreira, T., Morris, A.P. *et al.* Conditional and joint multiple SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* (in press).
23. Price, A.L., Weale, M.E., Patterson, N. *et al.* Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* **83**, 132-5; author reply 135-9 (2008).
24. Wood, A.R., Hernandez, D.G., Nalls, M.A. *et al.* Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Human molecular genetics* **20**, 4082-92 (2011).
25. Lappalainen, T., Montgomery, S.B., Nica, A.C. & Dermitzakis, E.T. Epistatic selection between coding and regulatory variation in human evolution and disease. *Am J Hum Genet* **89**, 459-63 (2011).
26. Emison, E.S., Garcia-Barcelo, M., Grice, E.A. *et al.* Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *Am J Hum Genet* **87**, 60-74 (2010).
27. Lango Allen, H., Estrada, K., Lettre, G. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8 (2010).
28. Speliotes, E.K., Willer, C.J., Berndt, S.I. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
29. Heid, I.M., Jackson, A.U., Randall, J.C. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-60 (2010).
30. Sivakumaran, S., Agakov, F., Theodoratou, E. *et al.* Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet* **89**, 607-18 (2011).
31. Ferreira, M.A. & Purcell, S.M. A multivariate test of association. *Bioinformatics* **25**, 132-3 (2009).
32. Eijgelsheim, M., Estrada, K., Newton-Cheh, C. *et al.* Meta-analysis of genome-wide association studies of correlated traits. In Genetic determinants of Heart Rhythm and Conduction Disorders. (Erasmus University, Rotterdam, 2011).
33. Goldstein, D.B. Common genetic variation and human traits. *N Engl J Med* **360**, 1696-8 (2009).
34. Hirschhorn, J.N. Genomewide association studies--illuminating biologic pathways. *N Engl J Med* **360**, 1699-701 (2009).
35. Jostins, L., Morley, K.I. & Barrett, J.C. Imputation of low-frequency variants using the HapMap3 benefits from large, diverse reference sets. *Eur J Hum Genet* **19**, 662-6 (2011).
36. Zeggini, E. Next-generation association studies for complex traits. *Nat Genet* **43**, 287-8 (2011).
37. Holm, H., Gudbjartsson, D.F., Sulem, P. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat Genet* **43**, 316-20 (2011).
38. Sulem, P., Gudbjartsson, D.F., Walters, G.B. *et al.* Identification of low-frequency

variants associated with gout and serum uric acid levels. *Nat Genet* **43**, 1127-30 (2011).

39. Rafnar, T., Gudbjartsson, D.F., Sulem, P. *et al.* Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet* **43**, 1104-7 (2011).
40. Wijmenga, C. Netherlands 750 Genome Project: Family-Based Whole Genome Sequencing of 250 Parents and Child in *BGI Workshop Luncheon at ICHG / ASHG* (Montreal, CA., 2011).
41. Jostins, L. & Barrett, J.C. Genetic risk prediction in complex disease. *Human molecular genetics* **20**, R182-8 (2011).
42. Stacey, S.N., Sulem, P., Jonasdottir, A. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* **43**, 1098-103 (2011).
43. Chan, Y., Holmen, O.L., Dauber, A. *et al.* Common variants show predicted polygenic effects on height in the tails of the distribution, except in extremely short individuals. *PLoS Genet* **7**, e1002439 (2011).
44. Khosla, S., Westendorf, J.J. & Oursler, M.J. Building bone to reverse osteoporosis and repair fractures. *J Clin Invest* **118**, 421-8 (2008).
45. Vahle, J.L., Sato, M., Long, G.G. *et al.* Skeletal changes in rats given daily subcutaneous injections of recombinant human parathyroid hormone (1-34) for 2 years and relevance to human safety. *Toxicol Pathol* **30**, 312-21 (2002).
46. Deal, C. Future therapeutic targets in osteoporosis. *Curr Opin Rheumatol* **21**, 380-5 (2009).
47. Hindorff, L.A., Sethupathy, P., Junkins, H.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
48. Raychaudhuri, S. Mapping rare and common causal alleles for complex human diseases. *Cell* **147**, 57-69 (2011).
49. Hoeppner, L.H., Secreto, F.J. & Westendorf, J.J. Wnt signaling as a therapeutic target for bone diseases. *Expert Opin Ther Targets* **13**, 485-96 (2009).
50. Kramer, I., Baertschi, S., Halleux, C., Keller, H. & Kneissel, M. Mef2c deletion in osteocytes results in increased bone mass. *J Bone Miner Res* (2011).
51. Leupin, O., Kramer, I., Collette, N.M. *et al.* Control of the SOST bone enhancer by PTH using MEF2 transcription factors. *J Bone Miner Res* **22**, 1957-67 (2007).
52. McClung, M.R., Lewiecki, E.M., Cohen, S.B. *et al.* Denosumab in postmenopausal women with low bone mineral density. *N Engl J Med* **354**, 821-31 (2006).
53. Zhou, G., Zheng, Q., Engin, F. *et al.* Dominance of SOX9 function over RUNX2 during skeletogenesis. *Proc Natl Acad Sci U S A* **103**, 19004-9 (2006).
54. Chimge, N.O., Baniwal, S.K., Luo, J. *et al.* Opposing Effects of Runx2 and Estradiol on Breast Cancer Cell Proliferation: In Vitro Identification of Reciprocally Regulated Gene Signature Related to Clinical Letrozole Responsiveness. *Clin Cancer Res* (2012).
55. Grilli, S. Tamoxifen (TAM): the dispute goes on. *Ann Ist Super Sanita* **42**, 170-3 (2006).
56. Abbott, A. Novartis to shut brain research facility. *Nature* **480**, 161-2 (2011).
57. Metabochip. Cardio MetaboChip Consortium (CMC2) Design and SNP Content Review. *http://www.sph.umich.edu/csg/kang/MetaboChip/doc/Metabochip_summary_final_091809.pdf*, Accessed online on February, 2012 (2012).

CHAPTER 6
SUMMARY

Summary

Osteoporosis is a common complex disease mostly characterized by decreased bone mineral density (BMD). In this thesis we aimed to identify novel genetic risk factors of osteoporosis-related traits using the hypothesis-free Genome-wide Association Study (GWAS) approach.

In **Chapter 1** we begin with an introduction to the relevance of the study of osteoporosis and provide background information on previous studies on the genetics of osteoporosis-related traits.

In **Chapter 2** we present a novel web- and grid-based tool (GRIMP), which was a used for the GWAS described in this thesis. This software is of major importance for the enlargement of GWAS with the whole-genome sequence data available from the 1000 Genomes Project.

In **Chapter 3** we focus on studies of genetics of adult height. In **Chapter 3.1** we identified a novel genetic variant (rs6717918) on chromosome 2q37.1 associated with height at $P=3.4x10^{-9}$. Our results suggest that variation in the C-type natriuretic peptide signaling pathway, involving the *NPPC* and *NPR3* genes, plays an important role in determining human body height. Further, using 183,727 individuals with GWAS data, we show in **Chapter 3.2** that hundreds of genetic variants, in at least 180 loci, influence adult height, a highly heritable and classic polygenic trait

In **Chapter 4**, we focus on the study of genetics of osteoporosis. In **Chapter 4.1** we performed the first meta-analysis of imputed data of GWAS of femoral neck (FN-) or lumbar spine (LS-) BMD and identified 20 loci that reached genome-wide significance, of which thirteen map to regions not previously associated with BMD: 1p31.3 (*GPR177*), 2p21 (*SPTBN1*), 3p22 (*CTNNB1*), 4q21.1 (*MEPE*), 5q14 (*MEF2C*), 7p14 (*STARD3NL*), 7q21.3 (*FLJ42280*), 11p11.2 (*LRP4*, *ARHGAP1*, *F2*), 11p14.1 (*DCDC5*), 11p15 (*SOX6*), 16q24 (*FOXL1*), 17q21 (*HDAC5*) and 17q12 (*CRHR1*). In **Chapter 4.2** we increased the discovery sample size of a meta-analysis of FN-BMD and LS-BMD to more than 30,000 samples and included an additional set of more than 50,000 samples for replication. We identified 56 loci (32 novel) associated with genome-wide significance with either FN-BMD or LS-BMD. Furthermore, we found fourteen BMD loci also associated with fracture risk at $P<5x10^{-4}$, of which six reached $P<5x10^{-8}$ including: 18p11.21 (*FAM210A*), 7q21.3 (*SLC25A13*), 11q13.2 (*LRP5)*, 4q22.1 (*MEPE*), 2p16.2 (*SPTBN1*) and 10q21.1 (*DKK1*). In **Chapter 4.3** we attempted to identify gene-by-sex interactions for BMD in 25,000 individuals with GWAS data. We did not identify any autosomal loci reaching the strict $P < 5x10^{-8}$ for gene-by-sex interaction, although some loci remained suggestive after adding 24,000 samples in the replication phase. In **Chapter 4.4** we assessed the effect of Copy Number Variants (CNVs) on fracture risk using the Rotterdam Study I (RSI) as

discovery set. A rare (minor allele frequency [MAF]=0.1%) 210Kb deletion in 6p25 was associated with increased fracture risk in the RSI study and further replicated in other array-based studies (P=0.02). Geographically constrained prevalence of this deletion was suggested.

Finally, in **Chapter 5** the results are discussed in a more general manner, and findings are placed in a broader context. In addition, methodological considerations regarding the studies contained in this thesis are discussed. Future directions are also presented at the end of **Chapter 5**.

Summary

SAMENVATTING

Osteoporose is een vaak voorkomende, complexe aandoening. Deze ziekte wordt gekenmerkt door een verminderde bot mineraal dichtheid (BMD). In dit proefschrift hebben wij getracht om nieuwe genetische risicofactoren voor osteoporose-gerelateerde fenotypen te identificeren. Daarvoor maakten wij gebruik van de hypothese-vrije genoom-brede associatie studie (GWAS) onderzoeksopzet.

In **Hoofdstuk 1** beginnen wij met een inleiding over de relevantie van studies naar osteoporose en geven wij achtergrondinformatie over eerdere studies naar de genetica van op osteoporose-gerelateerde fenotypen.

In **Hoofdstuk 2** presenteren wij een nieuwe grid-gebaseerde web applicatie (GRIMP), die werd gebruikt voor de GWAS beschreven in dit proefschrift. Deze software is van groot belang voor de uitbreiding van GWAS met de whole-genome sequence gegevens van het 1000 Genomen Project.

In **Hoofdstuk 3** concentreren wij ons op studies van genetica van volwassen lichaamslengte. In **Hoofdstuk 3.1** ontdekten wij een nieuwe genetische variant (rs6717918) op chromosoom 2q37.1 geassocieerd met lichaamslengte met $P=3.4x10^{-9}$. Onze resultaten suggereren dat variatie in de C-type natriuretic peptide signaling pathway, waaronder de genen *NPPC* en *NPR3*, een belangrijke rol heeft in het bepalen van de volwassen lichaamslengte. Vervolgens analyseerden wij GWAS gegevens van 183.727 individuen en laten wij in **Hoofdstuk 3.2** zien dat honderden genetische varianten, in minstens 180 chromosomale loci, bepalend zijn voor de volwassen eindlengte. Lichaamslengte bij de mens is grotendeels erfelijk en volgt een klassieke, polygenetische overerving.

In **Hoofdstuk 4** richten wij ons op de bestudering van genetica van osteoporose. In **Hoofdstuk 4.1** voerden wij de eerste meta-analyse uit met geïmputeerde data van GWAS voor BMD in de femurnek (FN-) en lumbale wervelkolom (LS-) en identificeerden hierbij 20 loci die genoom-brede significantie bereikten, waarvan dertien gelokaliseerd zijn in gebieden die nog niet die eerder met BMD waren geassocieerd: 1p31.3 (*GPR177*), 2p21 (*SPTBN1*), 3p22 (*CTNNB1*), 4q21.1 (*MEPE*), 5q14 (*MEF2C*), 7p14 (*STARD3NL*), 7q21.3 (*FLJ42280*), 11p11.2 (*LRP4*, *ARHGAP1*, *F2*), 11p14.1 (*DCDC5*), 11p15 (*SOX6*), 16q24 (*FOXL1*), 17q21 (*HDAC5*) en 17q12 (*CRHR1*). In **Hoofdstuk 4.2** breidden wij de grootte van de meta-analyse voor BMD uit met meer dan 30.000 deelnemers en voegden daar nog eens meer dan 50.000 deelnemers aan toe voor replicatie. Wij identificeerden 56 loci (waarvan 32 tot nog toe onbekend waren) met een genoom-brede significantie voor FN-BMD of LS-BMD. Voorts ontdekten wij dat veertien BMD loci ook geassocieerd waren met

fractuur risico met $P<5x10^{-4}$, waarvan zes met $P<5x10^{-8}$, met inbegrip van: 18p11.21 (*FAM210A*), 7q21.3 (*SLC25A13*), 11q13.2 (*LRP5)*, 4q22.1 (*MEPE*), 2p16.2 (*SPTBN1*) en 10q21.1 (*DKK1*). In **Hoofdstuk 4.3** onderzochten wij gen-om-sexe interacties voor BMD in 25.000 deelnemers met GWAS data. Wij vonden echter geen autosomale loci die een $P < 5x10^{-8}$ voor gen-om-sexe interactie bereikten, alhoewel sommige loci wel suggestief bleven nadat gegevens van 24.000 deelnemers werd toegevoegd in de replicatie fase. In **Hoofdstuk 4.4** bestudeerden wij het effect van kopie nummer variaties (CNVs) op fractuur risico binnen de Rotterdam Studie I (RSI) als zijde de ontdekking fase. Een zeldzame ('minor' allel frequentie [MAF]=0,1%) 210Kb deletie in 6p25 was geassocieerd met een verhoogd fractuur risico in de RSI studie en dit werd gerepliceerd in andere array-gebaseerde studies (P=0.02). De uitkomst van de studie suggereert een geografisch beperkte prevalentie van deze deletie.

Ten slotte, bediscussiëren wij in **Hoofdstuk 5** de resultaten in de meer algemene zin en plaatsen wij ze in een bredere context. Tevens worden methodologische overwegingen voor de studies in dit proefschrift besproken. Ook presenteren wij aanbevelingen voor toekomstig onderzoek aan het eind van **Hoofdstuk 5**.

CHAPTER 7
APPENDICES

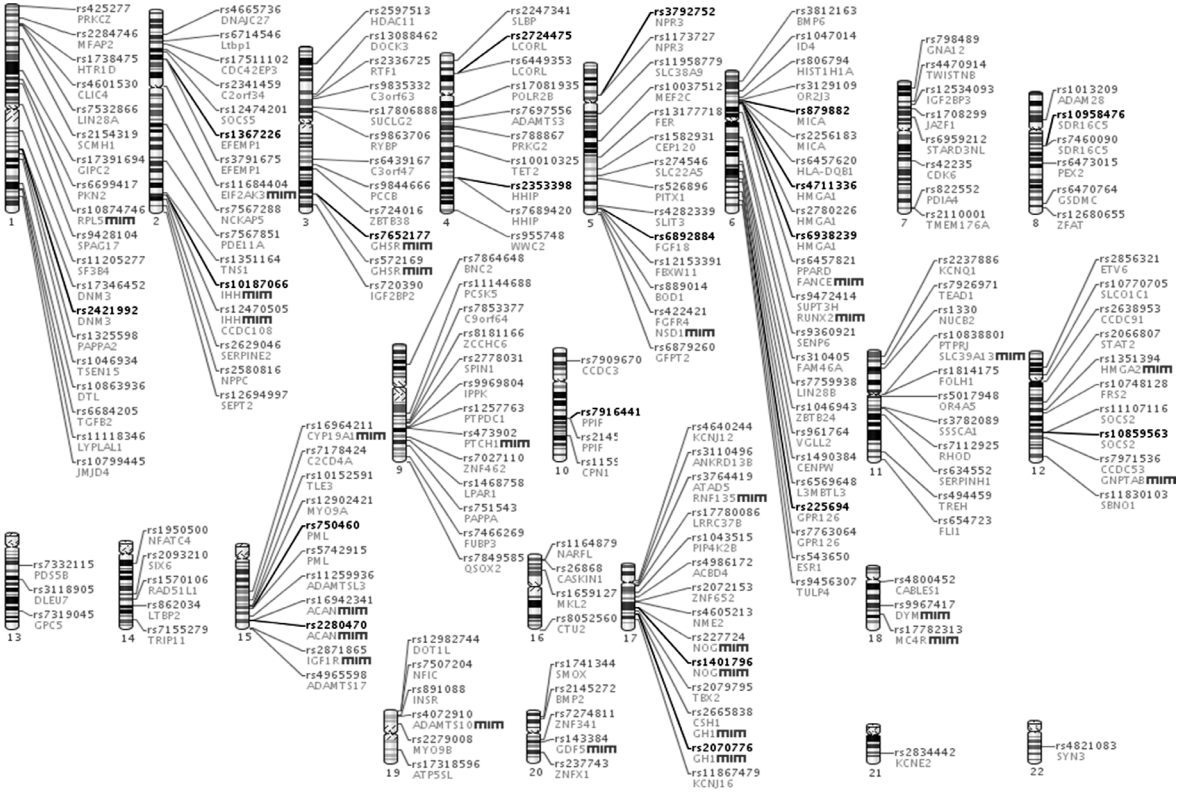
APENDIX 1. SUPPLEMENTARY INFORMATION

The chapters in Table 1 contain supplementary information and can be found in their respective journal websites.

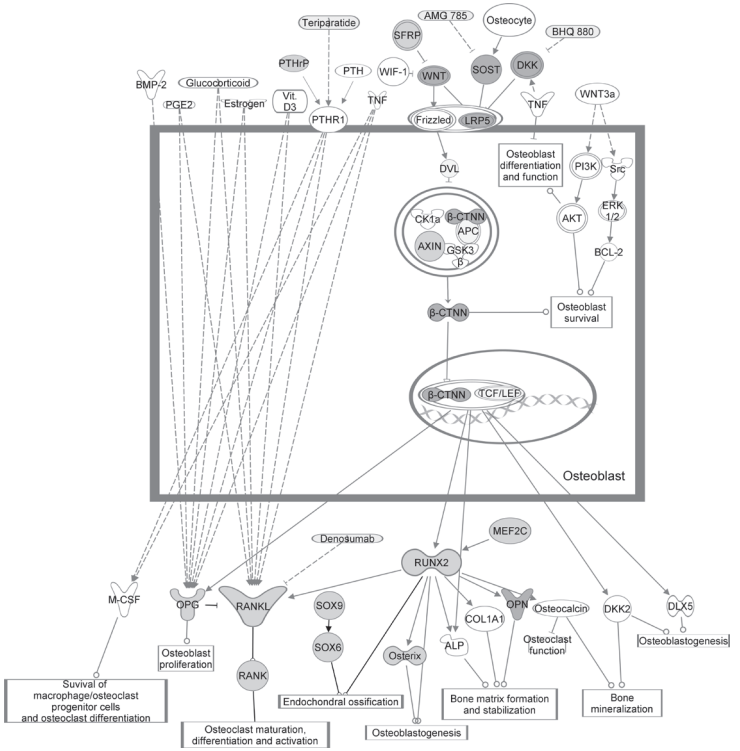
Table 1. URL for supplementary materials

Chapter	URL
3.1	http://hmg.oxfordjournals.org/content/suppl/2009/07/01/ddp296.DC1/ddp296supp.pdf
3.2	http://www.nature.com/nature/journal/v467/n7317/extref/nature09410-s1.pdf
4.1	http://www.nature.com/ng/journal/v41/n11/extref/ng.446-S1.pdf
4.2	http://www.nature.com/doifinder/10.1038/ng.2249

Selected supplementary materials from these published papers, and full supplementary information for unpublished manuscripts are presented in this chapter.

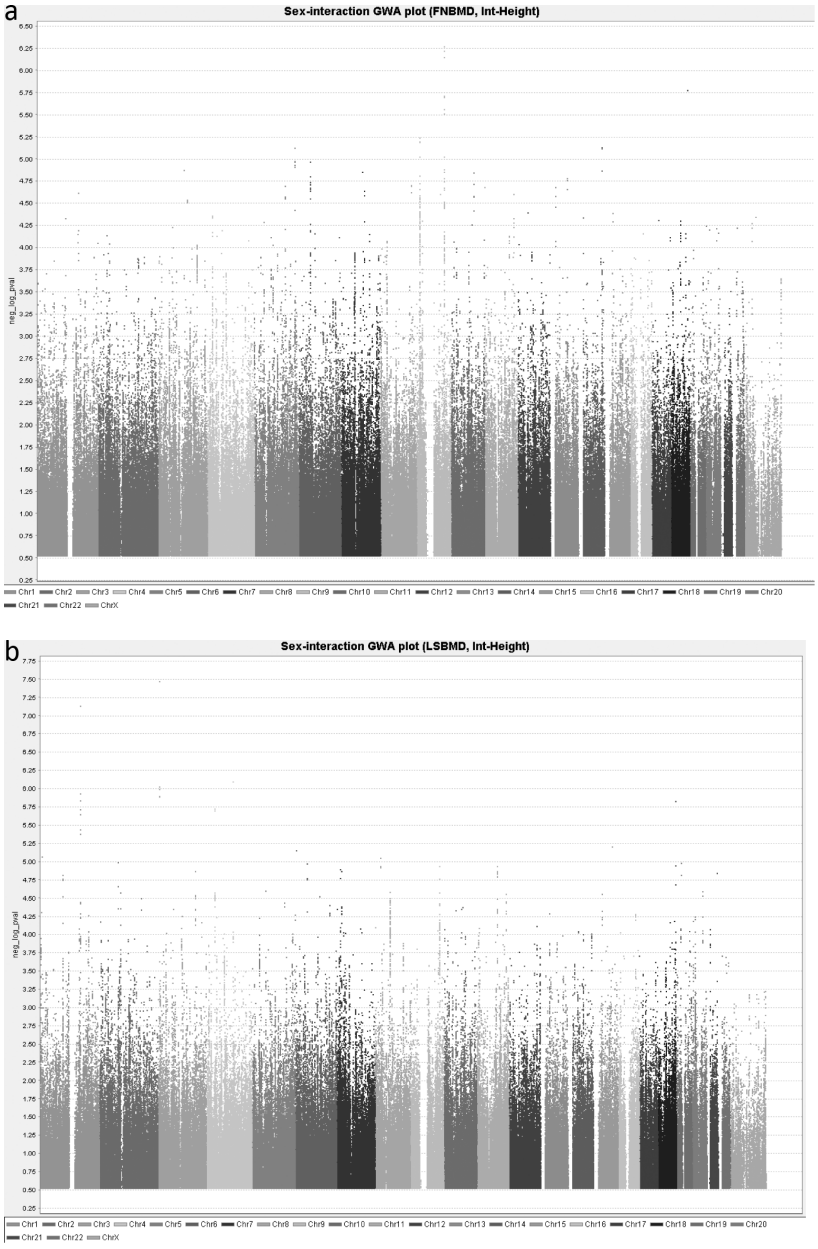


Chapter 3.2 - Supplementary Figure 2. 199 loci associated with adult height variation. Karyogram displaying the genome location of the 180 height SNPs identified from the primary meta-analysis (gray) and the 19 secondary signals (bold) discovered in the conditional analysis to be associated with height. The closest genes to the SNPs (gray) are followed by a MIM label if the gene underlies a skeletal growth-related Mendelian disorder described in OMIM. The plot was created using Affyrmation.

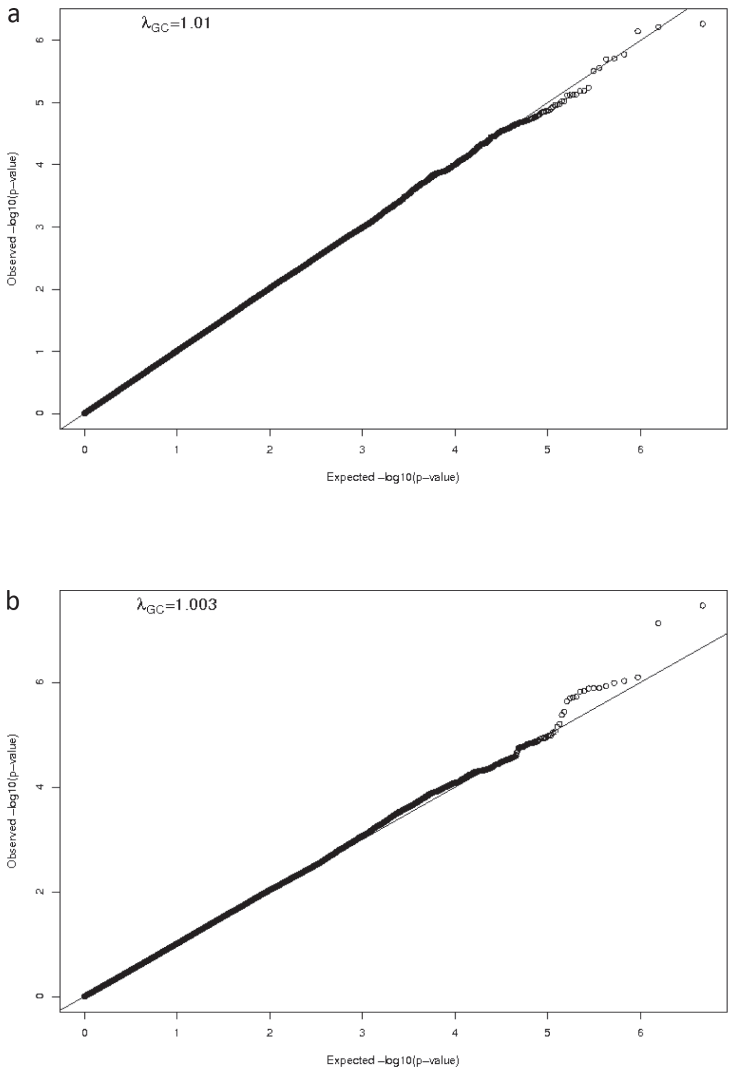


Chapter 4.2 - Supplementary Figure 10. Graphic representation of the different pathways critical to bone biology including osteoblastic differentiation; Wnt-mediated osteoblastic activation and function; RANK-mediated osteoclastic activation and endochondral ossification. The figure is derived from an image generated by Ingenuity Pathway Analysis (IPA) software version 2011 (Ingenuity Systems). The osteoblast is the main bone-forming cell. Genes identified by SNPs associated with BMD at genome-wide significant levels are in gray, genes identified by markers associated with fracture risk are in dark gray. Current pharmacological compounds used or under development for the treatment of osteoporosis are also depicted (Teriparatide, AMG 785, BHQ 880 and Denosumab).

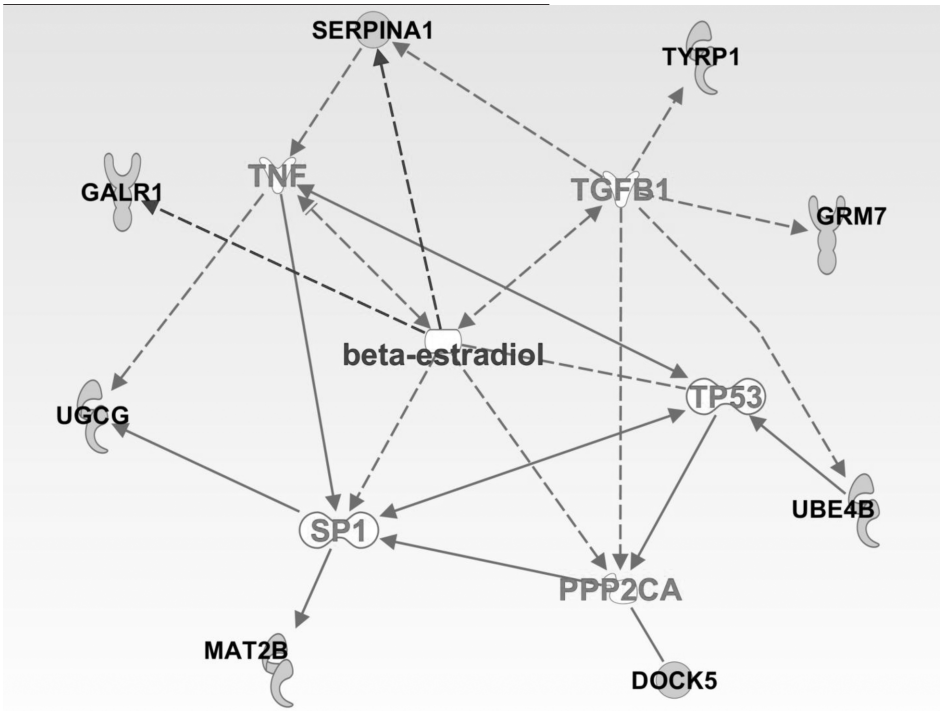
Chapter 4.3 Assessment of Gene by Sex Interaction Effect on Bone Mineral Density



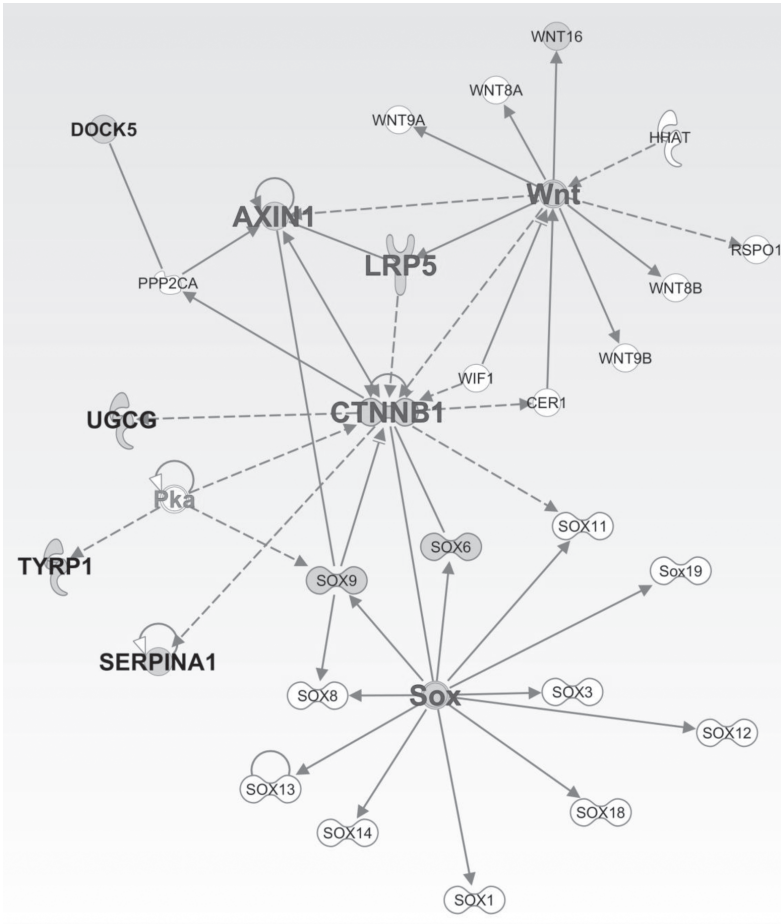
Supplementary Figure S1. Genome-wide association plots for a) FNBMD, b) LSBMD



Supplementary Figure S2. Quantile-quantile plots for a) FNBMD, b) LSBMD



Supplementary Figure S3. Functional interaction network focusing on β-estradiol



Supplementary Figure S4. Functional interaction network focusing on Wnt signaling

METHODS

Candidate Gene Analysis: SNP-by-Sex Interaction

We also tested eighty-two previously reported BMD-associated SNPs identified through a meta-analysis of 83,894 subjects[1] for SNP-by-sex interactions. This hypothesis-driven analysis also afforded us the opportunity to focus on the most promising SNPs from the BMD GEFOS II meta-analysis while reducing the penalty due to multiple testing[1]. In addition, this sex by SNP interaction analysis allows exploring whether the reason for 26 of those SNPs not replicating in the GEFOS II BMD meta-analysis can be attributed to sex-specificity.

Functional Interaction Network

Studies have suggested that genes under the same linkage analysis peak, and genes with similar expression profiles are related within hierarchical genetic networks or pathways[2]. To further understand the potential functions of the genes annotated to the most significant SNPs in our analyses, we focused on their functional interactions with sex hormones (specifically β-estradiol and testosterone), and Wnt signaling pathway members, because these pathways are critical in skeletal development, which is when skeletal features begin to differ between the sexes. We constructed functional interaction networks using Ingenuity Pathways Analysis (IPA) (Ingenuity, Redwood City, CA). The Ingenuity database is a knowledge repository of networks and biological relationships that have been systematically encoded into ontology based on more than 200,000 original peer-reviewed articles on mammalian biology. Details are available in the IPA Web site (www.ingenuity.com). Direct molecular interactions and regulatory events among genes/proteins have been manually extracted and curated from full text peer-reviewed articles of experimental findings in human, mice and rats to create the IPA database.

The details of the algorithm used to construct networks has been described elsewhere[3]. In brief, after annotating the SNPs from each of our top SNP-sex interaction GWAS associations to a gene, we overlaid results onto a global molecular network from the Ingenuity database to form a sub-network. With the assumption of biological commonalities among top associated genes for a specific disease, sub-networks are then combined into small-networks that maximize their specific

connectivity, which is a reflection of their interconnectedness with each other relative to all molecules they are connected to in the Ingenuity database. Additional molecules from the Ingenuity database are used to connect small-networks by merging them into a larger one. The finalized functional interaction networks are limited up to 35 molecules (genes/proteins) each to keep them to a usable size. A graphical network of the functional relationships between gene/gene products was extracted. Genes/gene products are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). Nodes are displayed using various shapes that represent the functional class of the gene products. Edges are displayed with various labels that describe the nature of the relationship between the nodes. Statistical significance was calculated as a p-value of the probability of the genes forming a specific network in comparison to all possible networks, with the right-tailed Fisher's Exact Test based on the hypergeometric distribution.

RESULTS

Candidate Gene Analysis of SNP-by-Sex for Previously Reported SNPs

Table 3 displays results for previously reported BMD-associated SNPs that demonstrated a significant interaction signal. One SNP, rs344081 (in *FLJ16641*), showed a significant sex interaction for FNBMD, however it should be noted that none of these SNPs from discovery were successfully replicated. The results for the complete list of BMD-related loci are presented in Supplementary Table S3.

Functional Interaction Networks with Sex Hormones & Wnts

We constructed functional interaction networks using 10 of the final 12 gene-sex interaction loci (see main paper table 1) with functional information along with sex hormone relevant proteins and genes. Due to lack of biological or functional annotation, *RELL1* and *C4orf32* were excluded from analyses. A functional interaction network that linked to β-estradiol was constructed (**Supplementary Figure S3**), and we found that 8 out of 10 gene-sex interaction genes were physically (*GALR1* and *SERPINA1*) interacting with or indirectly linked to (*DOCK5, GRM7, MAT2B, TYRP1,*

UBE4B and *UGCG* genes) β-estradiol ($p=10^{-25}$). The indirect links between *DOCK5, GRM7, MAT2B, TYRP1, UBE4B* and *UGCG* genes and β-estradiol are through their direct interaction with PPP2CA, TGFB1, SP1, TP53 and TNF molecules (gene, mRNA or protein). For example, UBE4B has been found to negatively regulate the level of p53 and to inhibit p53-dependent transactivation and apoptosis[4]. Based on chromatin immunoprecipitation assays, studies found p53 to be recruited to the Estrogen Receptor (ER)-α promoter along with other transcription factors, such as CARM1, CBP, c-Jun, and Sp1 and that this complex was formed in a p53-dependent manner, which suggests that p53 regulates ER expression through transcriptional control of the ER promoter[5]. β-estradiol increases expression of human SERPINA1 mRNA[6]. Estradiol and progesterone also have been shown to decrease expression of rat Galr1 mRNA[5].

We also constructed functional interaction networks with the same 10 gene-sex interaction loci (excl. *RELL1* and *C4orf32* as above) along with Wnt signaling pathway. As is shown in the functional interaction network (**Supplementary Figure S4**), *UGCG, TYRP1* and *SERPINA1* were found to be functionally interacting with Wnt signaling pathways ($p=10^{-11}$) and *UGCG* and *SERPINA1* to be regulated by protein-protein complex that consists of human β-catenin[7].

REFERENCES

1. Estrada, K. Styrkarsdottir, U., Evangelou, E., *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* advance online publication, 15 April 2012; doi:10.1038/ng.2249.
2. Pulendran, B., Querec, T.D., Akondy, R.S. *et al.* Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nature Immunology* **10**, 116-125 (2009).
3. Calvano, S.E., Xiao, W.Z., Richards, D.R. *et al.* A network-based analysis of systemic inflammation in humans (vol 437, pg 1032, 2005). *Nature* **438**, 696-696 (2005).
4. Leng, R.P. & Wu, H. UBE4B, a ubiquitin chain assembly factor, is required for MDM2-mediated p53 polyubiquitination and degradation. *Cell Cycle* **10**, 1912-1915 (2011).
5. Fuchs-Young, R., Shirley, S.H., Rundhaug, J.E. *et al.* Transcriptional Regulation of Estrogen Receptor-alpha by p53 in Human Breast Cancer Cells. *Cancer Research* **69**, 3405-3414 (2009).
6. Kanis, J.A., Stevenson, M., McCloskey, E.V., Davis, S. & Lloyd-Jones, M. Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis. *Health Technology Assessment* **11**, 1-+ (2007).
7. Cho, K.R., Schwartz, D.R., Wu, R. *et al.* Novel candidate targets of beta-catenin/T-cell factor signaling identified by gene expression profiling of ovarian endometrioid adenocarcinomas. *Cancer Research* **63**, 2913-2922 (2003).

APPENDIX 2: SUPPLEMENTARY TABLES

Chapter 4.3 Assessment of Gene by Sex Interaction Effect on Bone Mineral Density

Supplementary Table S1. Description of Participating Cohorts

Study	Gender	N	Age (yrs) mean (SD)	Height (cm) mean (SD)	Weight (kg) mean (SD)	FNBMD (g/cm2) mean (SD)	LSBMD (g/cm2) mean (SD)
Discovery Cohorts							
AMISH	Males	443	51.3 (15.5)	171.00 (6.40)	77.90 (12.20)	0.85 (0.13)	0.97 (0.13)
	Females	475	52.2 (14.6)	160.00 (5.90)	72.20 (14.50)	0.82 (0.14)	0.94 (0.15)
	Combined	918	51.7 (15.1)	165.00 (8.60)	74.90 (13.70)	0.84 (0.14)	0.96 (0.14)
CHS	Males	340	77.0 (4.8)	172.70 (6.50)	78.40 (12.20)	0.76 (0.14)	1.12 (0.24)
	Females	568	76.3 (4.2)	158.90 (6.40)	65.80 (13.70)	0.63 (0.11)	0.91 (0.24)
	Combined	908	76.6 (4.4)	164.10 (9.30)	70.50 (14.50)	0.68 (0.14)	0.99 (0.25)
DECODE	Males	1136	66.1(14.2)	176.50(6.70)	83.40(14.40)	0.80(0.10)	1.00(0.20)
	Females	6469	59.7(13.8)	164.40(6.20)	71.10(13.40)	0.70(0.10)	0.90(0.20)
	Combined	7605	60.7(13.9)	166.21(6.28)	72.94(13.55)	0.71(0.10)	0.91(0.20)
ERF	Males	908	48.8(14.5)	174.44(7.23)	83.00(14.23)	0.97(0.15)	1.17(0.17)
	Females	1191	47.7(14.4)	161.75(6.60)	69.18(13.76)	0.91(0.13)	1.12(0.16)
	Combined	2099	48.2(14.5)	167.23(9.32)	75.21(15.59)	0.93(0.14)	1.15(0.17)
FRAMINGHAM	Males	2561	54.9(15.3)	175.45(7.10)	86.54(14.58)	0.98(0.15)	1.31(0.20)
	Females	3240	55.8(15.7)	161.40(6.83)	69.57(15.09)	0.88(0.16)	1.17(0.20)
	Combined	5801	55.4(15.5)	167.61(9.85)	77.06(17.09)	0.93(0.16)	1.23(0.21)
HEALTHABC	Males	833	73.9 (2.9)	173.55 (6.40)	81.51 (12.45)	0.76(0.12)	1.07 (0.19)
	Females	734	73.6(2.8)	159.44 (5.88)	66.31 (12.15)	0.65(0.11)	0.91(0.17)
	Combined	1567	73.8(2.8)	166.94 (9.36)	74.39 (14.46)	0.71(0.13)	1.00(0.20)
RS I	Males	2110	68.1(8.2)	174.85(6.76)	78.58(10.74)	0.92(0.14)	1.17(0.20)
	Females	2794	70.3(9.6)	161.33(6.65)	69.59(11.29)	0.83(0.14)	1.04(0.18)
	Combined	4904	69.4(9.0)	167.14(6.70)	73.46(11.06)	0.87(0.14)	1.10(0.19)
RS II	Males	781	63.7(6.8)	176.05(6.47)	83.50(11.40)	0.97(0.13)	1.21(0.19)
	Females	898	63.8(7.4)	162.87(6.19)	72.80(12.50)	0.89(0.14)	1.11(0.19)
	Combined	1679	63.8(7.1)	169.00(6.32)	77.78(12.00)	0.93(0.14)	1.16(0.19)
Replication Cohorts							
RS III	Males	528	56.1(5.5)	178.80(6.70)	89.70(14.10)	1.00(0.10)	1.20(0.20)
	Females	683	56.1(5.5)	165.00(6.20)	75.20(14.30)	0.90(0.10)	1.20(0.20)
	Combined	1211	56.1(5.5)	171.02(6.42)	81.52(14.21)	0.94(0.10)	1.20(0.20)
SAFOS	Males	370	42.4(15.5)	170.68 (6.47)	85.71 (17.83)	0.90 (0.15)	1.05 (0.13)
	Females	611	44.2(15.3)	156.75 (6.38)	76.65 (18.03)	0.84 (0.14)	1.01 (0.14)
	Combined	981	43.5 (15.4)	162.00 (9.31)	80.07 (18.48)	0.87 (0.14)	1.02 (0.14)
TWINS UK 1	Females	2174	61.4 (12.2)	162.12 (6.20)	67.27 (12.23)	0.81(0.13)	1.00(0.15)
TWINS UK 23	Males	473	58.9(12.6)	175.26 (6.39)	81.85 (11.69)	0.87(0.14)	1.03(0.15)
	Females	3029	61.4(12.2)	162.12 (6.20)	67.27 (12.23)	0.80(0.14)	0.99(0.15)
	Combined	3502	61.4(12.2)	162.39 (6.47)	67.57 (12.39)	0.81(0.14)	0.99(0.15)
HKOS	Females	800	48.9(15.5)	155.11(6.71)	54.74(10.27)	0.70(0.17)	0.89(0.21)
GOOD	Males	938	18.9(0.56)	181.4(6.75)	73.84(11.89)	1.17(0.16)	1.21(0.15)
AROS	Males	171	54.3(15.7)	176.20(7.40)	78.10(12.40)	0.70(0.10)	0.90(0.20)
	Females	605	61.8(12.9)	161.70(6.70)	63.80(11.00)	0.70(0.10)	0.80(0.20)
	Combined	776	60.1(13.9)	165.00(9.00)	66.77(12.71)	0.67(0.14)	0.84(0.18)
BARCOS	Females	1443	65.5(9.1)	157.00(6.00)	64.95(10.48)	0.69(0.11)	0.86(0.15)
CABRIO-C	Males	529	63.9(8.5)	168.20(6.10)	81.50(11.00)	0.80(0.10)	1.00(0.20)

Supplementary Table S1. Description of Participating Cohorts

Study	Gender	N	Age (yrs) mean (SD)	Height (cm) mean (SD)	Weight (kg) mean (SD)	FNBMD (g/cm2) mean (SD)	LSBMD (g/cm2) mean (SD)
	Females	902	62.0(9.8)	155.90(6.00)	68.70(12.00)	0.70(0.10)	0.90(0.10)
	Combined	1431	62.7(9.4)	161.00(8.00)	73.58(13.16)	0.76(0.13)	0.96(0.15)
CAIFOS	Females	1082	80.2(2.7)	158.00(6.00)	67.64(12.10)	0.68(0.10)	0.95(0.18)
CAMOS	Males	715	65.4(16.6)	174.10(7.10)	81.70(13.40)	0.80(0.10)	1.00(0.20)
	Females	1593	67.3(14.9)	160.60(6.40)	69.40(13.70)	0.70(0.10)	1.00(0.20)
	Combined	2308	57.9(14.0)	165.00(9.00)	73.03(14.40)	0.76(0.13)	0.98(0.17)
DOPS	Females	1710	50.6(2.8)	165.00(6.00)	67.72(11.84)	0.80(0.11)	1.03(0.14)
EDOS	Males	354	62.4(13.6)	170.60(8.80)	77.30(17.20)	0.70(0.10)	0.90(0.20)
	Females	1615	66.2(12.4)	158.00(7.10)	65.10(13.70)	0.60(0.10)	0.80(0.20)
	Combined	1969	65.5(12.8)	160.00(9.00)	67.30(15.15)	0.64(0.13)	0.81(0.18)
FLOS	Males	159	53.9(14.7)	175.60(7.00)	80.80(13.60)	0.80(0.20)	1.00(0.10)
	Females	834	60.9(12.0)	160.00(6.70)	61.80(9.30)	0.70(0.20)	0.90(0.20)
	Combined	993	59.8(12.7)	163.00(9.00)	64.83(12.30)	0.72(0.17)	0.89(0.18)
GEOS	Females	2377	53.8(9.6)	159.00(6.00)	64.99(11.86)	0.88(0.14)	1.11(0.17)
GEVUR	Males	75	59.2(12.9)	170.80(7.50)	76.20(14.00)	0.90(0.20)	0.90(0.20)
	Females	397	62.2(8.2)	159.30(6.30)	70.60(13.20)	0.80(0.20)	0.90(0.20)
	Combined	472	61.8(9.0)	161.00(7.00)	71.25(13.37)	0.78(0.17)	0.89(0.18)
GROS	Males	41	70.2(12.8)	164.80(10.40)	71.50(12.30)	0.80(0.20)	0.90(0.20)
	Females	252	69.1(11.7)	161.80(7.30)	71.70(10.90)	0.80(0.10)	0.80(0.20)
	Combined	293	69.3(11.8)	162.00(8.00)	71.68(11.12)	0.80(0.15)	0.82(0.17)
MrOSS	Males	2893	75.4(3.2)	175.00(7.00)	80.70(12.07)	0.85(0.15)	1.15(0.22)
OAS	Males	589	68.1(4.2)	174.00(7.00)	83.70(12.53)	0.76(0.11)	1.05(0.17)
SLO-PREVAL	Males	121	67.9(6.5)	171.70(6.30)	81.60(12.60)	0.80(0.20)	1.00(0.20)
	Females	590	62.1(10.6)	160.40(6.30)	69.10(12.20)	0.70(0.10)	0.90(0.20)
	Combined	711	63.1(10.3)	162.00(8.00)	71.24(13.12)	0.71(0.13)	0.90(0.17)

Supplementary Table S2. Genotype, Imputation and Sample QC information for Participating Cohorts

	Genotyping		SNP Inclusion criteria				SNP Imputation			Samples			Association
Study	**Platform(s) / Chip(s)**	**Calling Algorithm**	**MAF**	**Call Rate**	**P-test HWE**	**Included SNPs**	**Method**	**MAF**	**Quality metric**	**Total # SNPs**	**Genotyped samples**	**Call rate* selection**	**Sample QC / Other exclusions**	**Software**
Discovery Cohorts														
AMISH	Affymetrix / 500K or 6.0	Birdseed	≥ 1%		95% >10-6	338,598	MACH	≥ 1%	MACH R2	2543013	1213	>95.0%	1. Missing BMD data. 2 Missing Covariate	MMAP (J.O'Connell)
CHS	Illumina 370CNV	BeadStudio	> 1%	≥97%	> 10^{-5}	306,655	BimBam	≥1%	Variance on the allele dosage >0.01	2,335,99	3291	>95%	1) presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack; 2) missing DNA; 3) non-Caucasian ethnicity; 4) gender mismatch; 5) discordance with prior genotyping.	R
DECODE	Illumina HH300 and 370CNV	BeadStudio	> 1%	> 96%	> 10^{-6}	281,410	IMPUTE	≥1%	MACH R2 ≥ 0.3	2,454,808	7605	≥91%	1) missing BMD measurement; 2) missing body weight and height.	SNPTEST
ERF	Illumina	Beadstudio	>0.5%	>98%	≥ 10^{-6}		MACH	≥1%			1602	>95%	1) gender mismatch; 2) ethnic outliers; 3) Missing phenotype data; 4) high IBS; 5) excess heterozygosity.	
	Affymetrix	BRLMM	>1%	>95%	≥ 10^{-6}	upto 487,573	MACH	≥1%	O/E Variance	2543887	2385	>95%	1) high IBS 2) high autosomal heterozygosity 3) ethnic outliers 4) missing trait 5) sex mismatch	ProbABEL
FRAMINGHAM	Affymetrix 500K Dual GeneChip + 50K gene-centered MOP set	BRLMM	≥ 1%	≥ 97%	≥ 10^{-6}	378163	MACH	≥ 1%	$(O/E)\sigma^2$ ratio ≥ 0.3	2540224	9274	≥ 97.0%	1. autosomal heterozygosity <0.33 or > 0.37 2. ethnic outliers (using Eigenstraat) 3. missing BMD or weight measurements	Kinship R-Package
HEALTHABC	Illumina Human1M-Duo BeadChip	Illumina BeadStudio	>1%	>97%	>10e-6	914,263	MACH v1.0.16	≥1%	O/E Variance	2,543,887	1663	≥ 97.0%	1) Missing DNA 2) 1st or 2nd degree relatives 3) missing body weight and height 4)ethnic outliers	ProbABEL
RS I	Illumina / HumanHap 550K V.3 ADHumanHap 550 V.3 DUO;	Beadstudio Genecall	≥ 1%	≥ 97.5%	> 10^{-6}	512,349	MACH	≥1%	MACH R2 ≥ 0.3	2,448,227	5,746	≥ 97.5%	1) missing DNA; 2) gender mismatch with typed X-linked markers; 3) excess autosomal heterozygosity > 0.336~FDR>0.1%; 4) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 5) ethnic outliers using IBS distances > 3SD from PLINK; 6) Missing body weight and height.	MACH2QTL via GRIMP
RS II	Illumina / HumanHap 550K V.3 ADHumanHap 550 V.3 DUO;	Beadstudio Genecall	≥ 1%	≥ 97.5%	> 10^{-6}	466,389	MACH	≥1%	MACH R2 ≥ 0.3	2,448,227	2,157	≥ 97.5%	1) missing DNA; 2) gender mismatch with typed X-linked markers; 3) excess autosomal heterozygosity > 0.336~FDR>0.1%; 4) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 5) ethnic outliers using IBS distances > 3SD from PLINK; 6) Missing body weight and height.	MACH2QTL via GRIMP

Supplementary Table S2. Genotype, Imputation and Sample QC information for Participating Cohorts

	Genotyping		SNP Inclusion criteria				SNP Imputation				Samples			Association
Study	**Platform(s) / Chip(s)**	**Calling Algorithm**	**MAF**	**Call Rate**	**P-test HWE**	**Included SNPs**	**Method**	**MAF**	**Quality metric**	**Total # SNPs**	**Genotyped samples**	**Call rate* selection**	**Sample QC / Other exclusions**	**Software**
Replication Cohorts														
In-silico replication														
RS III	Illumina / HumanHap610	Beadstudio Genecall	≥ 1%	≥ 97.5%	> 10^{-6}	514,073	MACH	≥1%	MACH R2 ≥ 0.3	2,448,227	1212	≥ 97.5%	1) missing DNA; 2) gender mismatch with typed X-linked markers; 3) excess autosomal heterozygosity > 0.336~FDR>0.1%; 4) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 5) ethnic outliers using IBS distances > 3SD from PLINK; 6) Missing body weight and height.	MACH2QTL via GRIMP
SAFOS	Illumina HumanHap 550	Beadstudio	≥ 1%	≥ 95%	> 10-6	531,800	MACH	≥ 1%	MACH R2 ≥ 0.3	1,387,467 *note HapMap3 reference data	1860	≥ 95%	1) Missing BMD measurements. 2) Missing covariate 2) Missing covariate data	MMAP (J.O'Connell)
TWINS UK	Illumina HumanHap 300 & 550. Illumina HumanCNV370 Duo	Beadstudio Genecall	≥ 1%	≥ 95%	> 10^{-6}	313,575	IMPUTE	≥1%	Prop_info >0.4	2,561,701	1511	≥ 95%	1) autosomal heterozygosity <0.33 or > 0.37; 2) ethnic outliers (using STRUCTURE);	GenABEL
TWINS UK-23	Illumina 610k	Beadstudio Genecall	≥ 1%	≥ 95%	> 10^{-6}	545,026	IMPUTE	≥1%	Prop_info >0.4	2,561,701	2801	≥ 95%	3) missing BMD or weight measurements.	GenABEL
HKOS	Human610-Quad Chip	Illumina BeadStudio	≥0.01	≥95%	≥10e-6	489,068	IMPUTE v2.1	≥0.01	proper_info>=0.3	2,426,092 for autosomes	800	95%	(1) genotyping call rate less than 95% ; (2) autosomal heterozygosity less than 28% or more than 30% ; (3) being related or identical to other individuals in the sample and (4) discordance of observed gender and estimated gender .	SNPTEST v2.2.0
GOOD	Illumina610	BeadStudio	≥ 1%	>98%	>10-6	521,160	MACH 1.0		MACH R2	2,543,887	938	98%	1) heterozygosity > 33%; 2) ethnic outliers 3) related individuals and duplicates	GRIMP
De novo genotyped														
GENOMOS	KASPar assay	SNPviewer2	≥ 1%	>90%	> 10-6	12	NA	NA	NA	NA	27715	80%	Discordance of observed gender and estimated gender .	Plink v1.07

Supplementary Table 3. Investigation of Sex-by-Gene Interaction for 56 BMD-related Loci

					FNBMD			LSBMD		
SNP	Locus	Closest Gene	Coded Allele/Non-coded Allele	Allele Freq	Beta	SE	P-value	Beta	SE	P-value
rs12407028	1p31.3	GPR177	t/c	0.6091	-0.0006	0.0024	8.00×10-1	0.0008	0.0036	8.24×10-1
rs6426749	1p36.12	ZBTB40	c/g	0.1736	0.001	0.0032	7.53×10-1	-0.0037	0.0047	4.32×10-1
rs479336	1q24.3	DNM3	t/g	0.7402	-0.001	0.0027	7.20×10-1	0.0043	0.0041	2.95×10-1
rs4233949	2p16.2	SPTBN1	c/g	0.3802	0.0009	0.0025	7.23×10-1	-0.0009	0.0037	8.07×10-1
rs7584262	2p21	LOC91461	t/c	0.2284	0.0059	0.0028	3.56×10-2	0.0012	0.0041	7.75×10-1
rs17040773	2q13	ANAPC1	a/c	0.7706	0.0024	0.003	4.18×10-1	0.0021	0.0044	6.30×10-1
rs1878526	2q14.2	INSIG2	a/g	0.2246	-0.0031	0.0029	3.00×10-1	-0.0098	0.0043	2.31×10-2
rs1346004	2q24.3	GALNT3	a/g	0.4866	-0.0023	0.0025	3.46×10-1	-0.0075	0.0035	3.48×10-2
rs430727	3p22.1	CTNNB1	t/c	0.4728	-0.0012	0.0024	6.05×10-1	-0.0061	0.0036	8.35×10-2
rs1026364	3q13.2	KIAA2018	t/g	0.3595	-0.003	0.0025	2.28×10-1	-0.0021	0.0036	5.60×10-1
rs344081	3q25.31	LEKR1	t/c	0.8748	0.0136	0.0036	1.31×10-4	0.0105	0.0051	3.92×10-2
rs3755955	4p16.3	IDUA	a/g	0.1601	-0.0069	0.004	8.17×10-2	-0.006	0.0059	3.14×10-1
rs6532023	4q22.1	MEPE	t/g	0.335	-0.0025	0.0025	3.01×10-1	-0.0021	0.0036	5.65×10-1
rs1366594	5q14.3	MEF2C	a/c	0.5261	0.0007	0.0023	7.56×10-1	-0.0012	0.0034	7.20×10-1
rs11755164	6p21.1	SUPT3H	t/c	0.4012	0.002	0.0025	4.27×10-1	-0.0006	0.0037	8.77×10-1
rs9466056	6p22.3	CDKAL1	a/g	0.38	-0.0051	0.0024	3.36×10-2	-0.01	0.0036	5.09×10-3
rs13204965	6q22.32	RSPO3	a/c	0.7597	-0.0061	0.0029	3.59×10-2	-0.0023	0.0043	5.82×10-1
rs4869742	6q25.1	C6orf97/ESR	t/c	0.3151	-0.001	0.0025	7.01×10-1	-0.0014	0.0038	7.19×10-1
rs6959212	7p14.1	STARD3NL	t/c	0.3351	-0.0008	0.0025	7.55×10-1	-0.0006	0.0037	8.69×10-1
rs4727338	7q21.3	SLC25A13	c/g	0.671	0.0006	0.0024	7.90×10-1	0.0065	0.0037	7.39×10-2
rs3801387	7q31.31	WNT16	a/g	0.7406	0.0031	0.0027	2.42×10-1	0.0083	0.004	3.78×10-2
rs7812088	7q36.1	ABCF2	a/g	0.1215	0.0013	0.0036	7.25×10-1	0.004	0.0054	4.57×10-1
rs7017914	8q13.3	XKR9	a/g	0.4945	-0.0054	0.0023	2.04×10-2	-0.0089	0.0034	9.74×10-3
rs2062377	8q24.12	TNFRSF11B	a/t	0.5865	-0.0048	0.0024	4.47×10-2	-0.0003	0.0035	9.31×10-1
rs7851693	9q34.11	FUBP3	c/g	0.6327	0.0002	0.0025	9.46×10-1	-0.0025	0.0038	5.12×10-1
rs3905706	10p11.23	MPP7	t/c	0.232	0.0001	0.003	9.62×10-1	-0.0034	0.0043	4.36×10-1
rs1373004	10q21.1	MBL2	t/g	0.1246	0.0062	0.0039	1.07×10-1	0.0139	0.0057	1.44×10-2
rs7071206	10q22.3_1	KCNMA1	t/c	0.7576	0.0025	0.0029	3.92×10-1	0.0031	0.0043	4.65×10-1
rs7084921	10q24.2	CPN1	t/c	0.3985	-0.0039	0.0025	1.15×10-1	0.0023	0.0035	5.06×10-1
rs7932354	11p11.2	ARHGAP1	t/c	0.3277	-0.0012	0.0027	6.50×10-1	-0.0025	0.0039	5.22×10-1
rs10835187	11p14.1_1	LIN7C	t/c	0.5401	0.0072	0.0024	2.34×10-3	0.007	0.0035	4.53×10-2
rs163879	11p14.1_2	DCDC5	t/c	0.658	0.0009	0.0025	7.10×10-1	0.0056	0.0038	1.39×10-1
rs7108738	11p15.2	SOX6	t/g	0.8224	0.0013	0.003	6.76×10-1	0.0012	0.0045	7.96×10-1
rs3736228	11q13.2	LRP5	t/c	0.1539	-0.0045	0.0033	1.72×10-1	-0.005	0.0048	3.00×10-1
rs7953528	12p11.22	KLHDC5	a/t	0.1837	0.0034	0.0031	2.78×10-1	0.0087	0.0048	6.89×10-2
rs2887571	12p13.33	ERC1	a/g	0.7621	0.0024	0.0027	3.84×10-1	0	0.004	9.96×10-1
rs12821008	12q13.12	DHH	t/c	0.3981	0.0019	0.0025	4.50×10-1	-0.0035	0.0037	3.48×10-1
rs2016266	12q13.13	SP7	a/g	0.6853	-0.0002	0.0025	9.27×10-1	0.0062	0.0038	1.01×10-1
rs1053051	12q23.3	C12orf23	t/c	0.5072	0.0013	0.0023	5.78×10-1	-0.0029	0.0035	4.06×10-1
rs9533090	13q14.11	AKAP11/RAN	t/c	0.4809	0.0026	0.0024	2.71×10-1	-0.001	0.0036	7.86×10-1
rs1286083	14q32.12	RPS6KA5	t/c	0.8009	-0.0041	0.0031	1.82×10-1	-0.006	0.0045	1.83×10-1
rs11623869	14q32.32	MARK3	t/g	0.3399	0.0019	0.0025	4.49×10-1	0.0035	0.0037	3.42×10-1
rs4985155	16p13.11	NTAN1	a/g	0.6532	-0.0077	0.0025	1.96×10-3	-0.0041	0.0038	2.80×10-1
rs9921222	16p13.3_1	AXIN1	t/c	0.4554	0.001	0.0025	6.80×10-1	-0.0012	0.0036	7.42×10-1
rs13336428	16p13.3_2	LOC390667	a/g	0.436	0.0006	0.0026	8.03×10-1	0.0041	0.0038	2.80×10-1
rs1566045	16q12.1	SALL1	t/c	0.8015	-0.0031	0.0035	3.80×10-1	0.0025	0.0052	6.32×10-1
rs10048146	16q24.1	FOXL1	a/g	0.7974	0.0017	0.0032	5.93×10-1	0.0014	0.0048	7.66×10-1
rs4790881	17p13.3	SMG6	a/c	0.6732	0.0033	0.0026	2.05×10-1	-0.0034	0.0039	3.77×10-1
rs227584	17q21.31_1	C17orf53	a/c	0.6694	-0.0014	0.0027	5.95×10-1	-0.0041	0.0039	2.98×10-1
rs1864325	17q21.31_2	MAPT	t/c	0.2204	0.0003	0.0028	9.11×10-1	0.0004	0.0042	9.24×10-1
rs7217932	17q24.3	SOX9	a/g	0.4574	0.0016	0.0024	4.96×10-1	0.0003	0.0036	9.39×10-1
rs4796995	18p11.21	C18orf19	a/g	0.6083	0.0004	0.0024	8.68×10-1	0.0045	0.0035	2.09×10-1
rs884205	18q21.33	TNFRSF11A	a/c	0.2533	0.0002	0.0028	9.36×10-1	0.0014	0.004	7.36×10-1
rs10416218	19q13.11	GPATCH1	t/c	0.7204	0.0022	0.0027	4.27×10-1	0.0044	0.004	2.76×10-1
rs3790160	20p12.2	JAG1	t/c	0.4954	0.0007	0.0024	7.67×10-1	-0.0009	0.0036	8.07×10-1
rs5934507	Xp22.31	FAM9B	a/g	0.7436	-0.008	0.0044	6.67×10-2	-0.0159	0.0065	1.47×10-2

Supplementary Table 1: Study design

Studies with GWAS data

	Study							Sample QC			
	Short name	Full name	Study design	Study base	Country of origin	City/region of origin	Ethnicity (es)	Call rate	Other exclusions	Short Study Description	References
1	**DeCODE**	DeCODE Genetics Study	Cross-sectional	Population-based, clinical-based	Iceland	NA	North-western European	≥91%	1) missing BMD measurement; 2) missing body weight and height.	The study includes 40,000 individuals taking part in various disease projects	[PMID: 18445777] {Styrkarsdottir, 2008 Multiple genetic loci for bone mineral density and fractures}; [PMID: 19079262] {Styrkarsdottir, 2009 Multiple genetic loci for bone mineral density and fractures}
2	**FHS**	Framingham Heart Study	Cohort	Population-based, family-based	United States of America	Framingham	European American	≥ 97%	1. autosomal heterozygosity <0.26 or > 0.30; 2. ethnic outliers (using Eigenstrata).	The Framingham Osteoporosis Study is an ancillary study of the parent, Framingham Study. The Framingham Study is a family-based, multigenerational cohort study initiated originally to study the risk factors for cardiovascular disease	[PMID: 14819398] {Dawber, 1951 Epidemiological approaches to heart disease: the Framingham Study}; [PMID: 474565] {Kannel, 1979 An investigation of coronary heart disease in families. The Framingham offspring study} [PMID: 17372189] {Splansky, 2007 The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination}
3	**RS-I**	Rotterdam Study-I	Cohort	Population-based	The Netherlands	Rotterdam	North-western European	≥ 97.5%	1) missing DNA; 2) gender mismatch with typed X-linked markers; 3) excess autosomal heterozygosity > 0.336~FDR>0.1%; 4) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 5) ethnic outliers using IBS distances > 3SD from PLINK; 6) missing body weight and height.	A prospective population-based cohort study of chronic disabling conditions in Dutch elderly individuals aged 55 years and over. The RS-III cohort included individuals aged 45 years and over.	[PMID:19700477] {Estrada, 2009 GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data}; [PMID:19728115] {Hofman, 2009 The Rotterdam Study: 2010 objectives and design update}; [PMID:1833235] {Hofman, 1991 Determinants of disease and disability in the elderly: the Rotterdam Elderly Study};
4	**RS-II**	Rotterdam Study-II	Cohort	Population-based	The Netherlands	Rotterdam	North-western European	≥ 97.5%			
5	**PROSPER/ PHASE**	The PROpective Study of Pravastatin in the Elderly at Risk	Cohort, randomized-controled trial	Clinical-based	The Netherlands/ United Kingdom /Ireland	Leiden/ Glasgow/ Cork	European	≥ 97.5%	1) missing DNA; 2) gender mismatch with typed X-linked markers; 3) excess autosomal heterozygosity > 0.336~FDR>0.1%; 4) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 5) ethnic outliers using IBS distances > 3SD from PLINK; 6) Missing body weight and height.	A randomized controlled clinical trial to test the effect of pravastatin on cardiovascular outcomes in the elderly at risk.	[PMID: 12457784] {Shepherd, 2002 Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial}

Supplementary Table 2: Study design

Studies that were de-novo genotyped for replication

Study								
Short name	Full name	Study design	Study base	Country of origin	City/region of origin	Ethnicity	Short Study Description	References
1 APOSS	Aberdeen Prospective Osteoporosis Screening Study	Cohort	Population-based	UK	Aberdeen	North-western European	APOSS is a longitudinal population-based study of osteoporotic fracture risk assessment in Caucasian women aged 45-54 years of age at baseline.	[PMID: 16355284] {Macdonald, 2006 Large-scale population-based study shows no evidence of association between common polymorphism of the VDR gene and BMD in British women}
2 CABRIO-C	Cantabria-Camargo	Cross-sectional	Community-based	Spain	Santander	Caucasian (Spanish)	Community-based study designed to evaluate the prevalence of metabolic bone diseases in postmenopausal women and men older than 50 attended at a primary care center in Northern Spain.	[PMID: 20594548] {Olmos, 2010 Bone turnover markers in Spanish adult men The Camargo Cohort Study}; [PMID: 19737549] {Martinez, 2009 Bone turnover markers in Spanish postmenopausal women: the Camargo cohort study}
3 CABRIO-CC	Cantabria Osteoporosis Case-control	Case-control, cross-sectional	Clinic-based plus volunteers	Spain	Santander	Caucasian (Spanish)	Clinic-based study of control individuals and patients with osteoporosis living in Cantabria, a region in Northern Spain	[PMID: 17118999] {Riancho, 2007 Identification of an aromatase haplotype that is associated with gene expression and postmenopausal osteoporosis}; [PMID: 17218734] {Zarrabeitia, 2007 Adiposity, estradiol, and genetic variants of steroid-metabolizing enzymes as determinants of bone mineral density}
4 CAIFOS	Calcium Intake Fracture Outcome Study	Cohort, randomized-controled trial	Population-based	Australia	Perth	Caucasian	Randomized-controled trial and cohort study	[PMID: 16636212] {Prince, 2006 Effects of calcium supplementation on clinical fracture and bone structure: results of a 5-year, double-blind, placebo-controlled trial in elderly women}
5 CaMos	Canadian Multicentre Osteoporosis Study	Cohort	Population-based	Canada	Vancouver, Calgary, Saskatoon, Hamilton, Toronto, Kingston, Québec City, Halifax, St John's	North-western European	The CaMos Study is a population-based, randomly selected, prospective cohort study from 9 Canadian cities followed for 14 years for osteoporosis-related traits and outcomes.	[PMID: 11199195] {Tenenhouse, 2000 Estimation of the prevalence of low bone density in Canadian women and men using a population-specific DXA reference standard: the Canadian Multicentre Osteoporosis Study (CaMos)}; [PMID: 17129177] {Richards, 2007 Changes to osteoporosis prevalence according to method of risk assessment} [PMID: 17242321] {Richards, 2007 Effect of selective serotonin reuptake inhibitors on the risk of fracture}
6 DOPS	Danish Osteoporosis Prevention Study	Cohort	Population-based	Denmark	Aarhus, Odense, Cope	Northern European	Population-based study of perimenopausal women. The women were followed for 10 years. App. 35% were treated with HRT	[PMID: 10340280] {Mosekilde, 1999 The Danish Osteoporosis Prevention Study (DOPS): project design and inclusion of 2000 normal perimenopausal women}
7 EDOS	Edinburgh Osteoporosis Study	Cross-sectional	Clinical-based	UK	Edinburgh and Lothian	British (white caucasian)	Clinical referral population of patients assessed for evaluation of osteoporosis	None

Supplementary Table 2: Study design

Studies that were de-novo genotyped for replication

Study								
Short name	Full name	Study design	Study base	Country of origin	City/region of origin	Ethnicity	Short Study Description	References
8 EPICNOR	Europen Prospective Investigation into Cancer, Norfolk study	Cohort	Population-based	UK	Norfolk	European	A random sample of 1,511 men and women in the top decade of age in the 25,000 participant EPIC-Norfolk prospective study were recruited into a bone fragility study with DXA measurements. GWA data were available for 249 participants with BMD measurements who had been entered into a GWAS investigation of obesity.	[PMID: 12753873] {Kaptoge, 2003 Effects of gender, anthropometric variables, and aging on the evolution of hip strength in men and women aged over 65}; [PMID: 10466767] {Day, 1999 EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer}; [PMID: 19079261] {Willer, 2009 Six new loci associated with body mass index highlight a neuronal influence on body weight regulation};
9 EPOLOS	Early risk identification and effective prevention of osteoporosis based bone fractures in Polish population.	Cross-sectional	Population-based	Poland	Warsaw, Lodz, Poznan, Krakow, Wroclaw, Bydgoszcz	Central European	The EPOLOS Study is a population-based, cross-sectional study of unrelated men and women aged 19-81 years, initiated to identify early risk and effective prevention of osteoporosis based bone fractures in Polish population.	[PMID: 20502405] {Skowrońska-Jóźwiak, 2010 Comparison of selected methods for fracture risk assessment in postmenopausal women: analysis of the Łódź population in the EPOLOS study}; [PMID: 20502404] {Skowrońska-Jóźwiak, 2010 Effect of sex, age, and anthropometric parameters on the size and shape of vertebrae in densitometric morphometry: results of the EPOLOS study}; [PMID: 19396748] {Skowrońska-Jóźwiak, 2009 Identification of vertebral deformities in the Polish population by morphometric X-ray absorptiometry - results of the EPOLOS study}
10 EPOS	European Prospective Osteoporosis Study	Cohort	Population-based	Europe	18 centres across 13 countries in Europe	European	EPOS was an extension of the European Vertebral Osteoporosis Study (EVOS) study and aimed to quantify incidence of vertebral and non-vertebral fractures. EVOS had recruited some 17,342 men and women aged over 50 years from 36 centres in 19 European countries. Each centre had recruited a random sample of up to 300 men and 300 women from population registers stratified into six 5-year age bands: 50-54, 70-74 and 75+. A total of 7,273 participants from 31 EVOS centres took part in the EPOS follow up study.	[PMID: 8797123] {O'Neill, 1996 The prevalence of vertebral deformity in european men and women: the European Vertebral Osteoporosis Study}; [PMID: 10824241] {Ismail, 2000 Validity of self-report of fractures: results from a prospective study in men and women across Europe. EPOS Study Group. European Prospective Osteoporosis Study Group}; [PMID: 11918229] {EPOS study group, 2002 Incidence of vertebral fracture in europe: results from the European Prospective Osteoporosis Study (EPOS)}

Supplementary Table 2: Study design

Studies that were de-novo genotyped for replication

	Study							
Short name	Full name	Study design	Study base	Country of origin	City/region of origin	Ethnicity	Short Study Description	References
11 **HCS**	Hertfordshire Cohort Study	Cohort	Population-based	UK	Hertfordshire county	Caucasian	The Hertfordshire Cohort Study is a population-based cohort study of men and women born between 1931 and 1939 in the county of Hertfordshire, UK. It was initiated to evaluate interactions between the genome, the intrauterine and early postnatal development, and adult diet and lifestyle in the aetiology of chronic disorders in later life.	[PMID: 15964908] {Syddall, 2005 Cohort Profile: The Hertfordshire Cohort Study}
12 **LASA**	Longitudinal Aging Study Amsterdam	Cohort	Population-based	The Netherlands	Amsterdam, Zwolle, Oss and surroundings	North-western European	LASA is an ongoing multidisciplinary cohort study on predictors and consequences of changes in physical, cognitive, emotional and social functioning in older persons.	[PMID: 11927198] {Deeg, 2002 Attrition in the Longitudinal Aging Study Amsterdam: The effect of differential inclusion in side studies}
13 **ManMc**	Manitoba McGill Fracture Study	Cross-Sectional	Population-based	Canada	Winnipeg	Caucasian (97%)	The Manitoba-McGill Fracture Study is a population-based sample of women experiencing a validated clinical hip or forearm fracture requiring orthopedic intervention.	[PMID: 21124974] {Ladouceur, 2010 An Efficient Paradigm for Genetic Epidemiology Cohort Creation}
14 **NOSOS**	North of Scotland Osteoporosis Study	Cohort	Population-based	UK	Aberdeen, Dingwall	North-western European	NOSOS is a population-based osteoporosis screening programme of postmenopausal females aged 60-82 years of age at baseline.	[PMID: 18633668] {Mavroeidi, 2009 Physical activity and dietary calcium interactions in bone mass in Scottish postmenopausal women}; [PMID: 20966103] {Judson, 2010 The Functional ACTN3 577X Variant Increases the Risk of Falling in Older Females: Results From Two Large Independent Cohort Studies}
15 **UFO**	The Umeå Fracture and Osteoporosis Study	Nested case-cohort	Population-based	Sweden	Umeå	Caucasians	The UFO study is a nested case-control study investigating associations between genes, lifestyle and osteoporotic fractures. The study is based on the prosepctive and populationbased Northern Sweden Health and Disease Study cohort, initiated to assess risk factors for diabetes and cardiovascular disesase.	[PMID: 20464545] {Englund, 2010 Physical activity in middle-aged women and hip fracture risk: the UFO study}; [PMID:14660243] {Hallmans, 2003 Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions}

Supplementary Table 3: CNV Population controls

SAGE controls, Illumina 1M, n= 1287 European, 495 African-Americans from USA

SampleID	Chr	Start	End	Size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
B299756_1007842055	6	4199935	4417586	217,652	96	SAGE (COGA study)	Illumina 1M single v1	PMID:20202923 and PMID:20531469	European	Blood

OHI controls, Affymetrix SNP 6.0, n= 1234 European from Canada

SampleID	Chr	Start	End	Size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
unknown_s1	6	4200639	4416634	215,996	217	(Ottawa Heart Institute (OHI) study)	Affymetrix SNP 6.0	PMID:19371834	European	Blood
unknown_s2	6	4200639	4416634	215,996	217	(Ottawa Heart Institute (OHI) study)	Affymetrix SNP 6.0	PMID:19371834	European	Blood
unknown_s3	6	4200639	4416634	215,996	217	(Ottawa Heart Institute (OHI) study)	Affymetrix SNP 6.0	PMID:19371834	European	Blood
unknown_s4	6	4200639	4416634	215,996	217	(Ottawa Heart Institute (OHI) study)	Affymetrix SNP 6.0	PMID:19371834	European	Blood
unknown_s5	6	4200639	4416634	215,996	217	(Ottawa Heart Institute (OHI) study)	Affymetrix SNP 6.0	PMID:19371834	European	Blood

PopGen controls, Affymetrix SNP 6.0, n= 1123 European from Germany(*)

SampleID	Chr	Start	End	Size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
-	-	-	-	-	-	PopGen	Affymetrix SNP 6.0	PMID:16490960	European	Blood

(*) No CNVs were detected in this region (0/1123)

WTCCC2 controls, Affymetrix SNP 6.0, n= 4783 European from UK

SampleID	Chr	Start	End	Size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
69838_A6_WTCCCT510553	6	4200639	4416634	215,996	217	UK Blood Service control (NBS)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Blood
69848_A4_WTCCCT511553	6	4200639	4416634	215,996	217	UK Blood Service control (NBS)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Blood
69848_H8_WTCCCT511477	6	4200639	4416634	215,996	217	UK Blood Service control (NBS)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Blood
69851_F1_WTCCCT511807	6	4200639	4416634	215,996	217	UK Blood Service control (NBS)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Blood
70196_E2_WTCCCT444878	6	4200639	4416634	215,996	217	1958 British Birth Cohort (58C)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Cell line
71062_A5_BLOOD294470	6	4200639	4416634	215,996	217	UK Blood Service control (NBS)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Blood
78721_A10_WTCCCT543540	6	4200639	4416634	215,996	217	1958 British Birth Cohort (58C)	Affymetrix SNP 6.0	WTCCC; https://www.wtccc.org.uk/ccc2/	European	Cell line

Shaikh et al. 2009, Illumina HumanHap550 v1, n= 1320 European, 694 African-American, 12 Asian from USA

SampleID	chr	start	end	size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
unknown/1560747007_A	6	4200707	4417586	216,880	69	CHOP	Illumina HumanHap550 v1	PMID: 19592680	European	Blood
unknown/1557564308_A	6	4200707	4417586	216,880	69	CHOP	Illumina HumanHap550 v1	PMID: 19592680	European	Blood
unknown/1587851191_A	6	4200707	4417586	216,880	69	CHOP	Illumina HumanHap550 v1	PMID: 19592680	European	Blood
unknown/1613642344_A	6	4200724	4417586	216,863	68	CHOP	Illumina HumanHap550 v1	PMID: 19592680	European	Blood

Itsara et al. 2009, n= 2493 (*)

SampleID	chr	start	end	size	#probes	Cohort_name	array_platform	Study	Ethnicity	DNA_source
unknown	6	4183487	4417586	234,099	31	PARC-CAP	Illumina HumanHap300	PMID: 19166990	European	Cell line
147	6	4200707	4417586	216,879	69	NINDS	Illumina HumanHap317 + 240S SNP Array	PMID: 19166990	European	Cell line
271	6	4200707	4426420	225,713	71	NINDS	Illumina HumanHap317 + 240S SNP Array	PMID: 19166990	European	Cell line
unknown	6	4200724	4417586	216,862	30	PARC-PRINCE	Illumina HumanHap300	PMID: 19166990	European	Blood

(*) A total of 2493 samples after QC from the datasets:

936 samples after QC from the Pharmacogenomics and Risk of Cardiovascular Disease study (PARC). The PARC samples are a subset of the cohorts used in two statin trials, CAP and PRINCE, and consist of 960 middle-age (40–70 years) individuals of European descent living in the United States.

671 samples after QC from Neurologically normal individuals identified at the National Institute for Neurological Disorders and Stroke (NINDS). 790 NINDS samples were obtained from the NINDS Human Genetics Resource Center DNA and Cell Line Repository. Genotype data from NINDS were derived from two sets of neurological disease controls totaling 790 people and consist of individuals of European descent with no family history of or any first-degree relative with amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson disease, or schizophrenia.

886 samples after QC from the Human Genome Diversity Panel (HGDP) that consists of 1064 individuals (lymphoblastoid cell lines) sampled from 51 different world populations.
Note: The HGDP panel was also analysed by Jakobsson et al. 2008 (PMID:18288195) and neither the large ~ 220Kb-DEL or smaller 26Kb-DEL was detected in the screened samples.

ACKNOWLEDGMENTS

This thesis would not been possible without the support of many individuals including all the Rotterdam Study participants who kindly donated their time and blood to conduct the genetic research described in this book.

First, I would like to express my deep gratitude to my promotor, Prof. Andre G. Uitterlinden. Dear Andre, thanks for the opportunity to join your group and for always being supportive to me. In my working experience I have had the luck of having great bosses but you are way on top of the list. I must say that in these four years many have told me how lucky I was to find my way in your group. Despite being one of the most successful professor in Netherlands and a highly recognized figure in the world of osteoporosis, you are always down to earth and easy going with anyone. I hope I can teach my son to follow your steps to make a good use of the name.

I'm deeply in debt with my co-promotor dr. Fernando Rivadeneira. Dear Fernando, since my arrival to Rotterdam you helped me beyond your academic responsibilities, thanks for offering your friendship since the very first day, I can not imagine how hard it would have been these years without your support. From you I learned that it is never too late and that everything is possible. Many people will never realize how much they owe you, but I had the privilege to see the amount of time that you can dedicate to help someone without expecting anything in return. Thank you for helping me joining the beautiful world of the osteoporosis research and for all those long nights reviewing the manuscripts in company of nespresso capuchinos and/or coffee sweets.

I am grateful to the members of the reading committee for their time and work: Professor Manfred Kayser, Professor Robert Hofstra, and Professor Timothy Frayling. Dear Manfred, it was a nice experience working with you for our first paper on adult height. Dear Robert, thanks for accepting being in the committee I hope we can work in future projects and learn from your experience. Dear Tim, thanks for showing me the power of international collaborative efforts and for being always supportive.

My deep thanks also to the members of the plenary committee, Professor Claes Olson, Professor Paul de Bakker, Professor Stuart Ralston and dr. Annelies de Klein for their time. Dear Claes and Stuart, thanks for helping in the discussion and the writing of the BMD paper and for sharing your knowledge about osteoporosis. Dear Paul, thanks for providing your critical view on our papers, I hope we collaborate even more in future projects now that you have a base in Netherlands. Dear Annelies, thanks for sharing together with Bert your expertise in clinical genetics.

I would like to also thank Professor Huib AP Pols. Dear Huib, thanks for acepting being the commission chairman, also thank you for giving us space in your tight agenda every month to receive your feedback in our results. I also thank Professor Albert Hofman, thanks for creating the richest cohorts of Europe: the Rotterdam Study and Generation R. Without your contribution this thesis would not have existed. Also thanks for organizing the NIHES courses and for letting me give something back by teaching the new generation of students.

I thank all the Rotterdam Study PIs who invited me to contribute in their groups, Professor Bruno Stricker, Professor Cornelia van Duijin and Professor Henning Tiemeier. I also thank dr. Abbas Denghan and dr. Arfan Ikram for being great colleagues during the PhD period. Dear Abbas, thanks for helping me in the design of my thesis and for providing me useful information on how to deal with the last steps in a PhD. Dear Arfan, thanks for organizing the course for students and for your support.

All the papers within Chapter 4 were product of collaborative work within the Genetic Factors for Osteoporosis (GEFOS) and the GENOMOS Consortia. I would like to thank to all my co-authors and the people involved in these papers. My special gratitude to Professor Doug Kiel, dr. David Karasik, dr. Brent Richards, Professor John Ioannidis, Professor Matthew Brown, dr. Unnur Styrskarsdottir, dr. Evangelos Evangelou, dr. Yi-Hsiang Hsu, dr. Emma Duncan, dr. Evangelia Ntzani, dr. Ling Oei, dr. Omar Albagha and dr. Ching-Ti Liu for being always available to write and review the NG 2012 GEFOS BMD paper.

I also thank Annette de Deugd and Jolanda den Hollander of the Netherlands Consortium of Healthy Aging (NCHA) for always keeping us updated with the latest news and for all their tips on how to become better scientist. My gratitude also to Eline Slagboom and her group for all the discussions on the NCHA meetings.

I'd like to thank the founders of the CHARGE consortium: Bruce Psaty, Chris O'Donell, Villy Gudnarson, Kathy Lunetta, Andre Uitterlinden, Tammy Harris, Jerome Rotter, Aaron Folsom and Eric Boerwinkle. It has been a great way to do scientific collaboration!. In the same way I want to thank the members of the GIANT consortium, my special gratitude to: Joel Hirschhorn, Tim Frayling, Mark McCarty, Goncalo Abecasis, Ruth Loos, Hana Lango-Allen, Sonja Berndt, Mike Weedon, Guillaume Lettre, Cristen Willer, Cecilia Lindgren and Elizabeth Spelliotes, among many more… it was a great adventure working with you!.

My thanks also extend to those who provided technical support, in particular Marijn Verkerk, Joost, Nano and Lenart. My gratitude also to Yurii, Maksim, Najaf, Ayşe and Natalia for proving support with analysis on ERF. Also thank Mark E. for

introducing the bi-variate method and for teaching me how to change my bike flat tire!. I also thank to Frank van Rooij for organizing the Rotterdam Study data. My deep gratitude to our lab team, especially to Mila and Pascal who ran together with a group of superb technicians (Michael, Sara, Manoushka, Ramazan, Saskia, to name a few) the Rotterdam Study GWAS datasets and many other important projects. Also special thanks to Jeroen, Slavik and Robert for creating the Exome dataset just on time to start distracting me when I was finishing this acknowledgments section. I also thank Maria J. for helping me to pay all those invoices and for organizing our accounts. I'd like to also thank Anis and Tobias. Dear Anis, without your grid support running the analysis for this thesis would have taken much longer. I'd like to thank also the support personal of BC Platforms, in particular Petteri for always solving imputation issues. And also thank Rebecca G. from Kbiosciences for helping us with the GENOMOS genotyping.

My special thanks to my kamergenoten! <Dutch section starts here>Beste Lisette, hartelijk bedankt voor de dagelijkse gesprekken die we samen in het Nederlands hebben gehad. Ik ben geslaagd voor de NT2. Beste Marjolein, van jou heb ik dingen geleerd die typisch Nederlands zijn zoals bijvoorbeeld het thuisblijven als je verkouden bent. Annemieke, jou wil ik alvast bedanken voor de 5 km marathon van Rotterdam die je gaat lopen samen met mijn vrouw.</Dutch section ends here>

I also want to thank my colleagues from the 21st floor: Martha, Hanneke, Marjolein, Natalia, and their respective supervisors Joyce and Carola for their input in my presentations.

Now I want to thank my paranymphs Ling Oei and Carolina Medina. Dear Caro, I cannot imagine how our life would be if you had not arrived on time to our group, thanks a lot for helping us and for accepting being my paranymph and a friend. Your dual background in statistics and biology is a treasure, make good use of it!. Dear Ling, thanks a lot for helping me organizing me the promotion and for the help in translating the summaries to Dutch. You're a very talented person and I'm sure you will succeed in all your goals.

It's time to thank my previous and current friends for the lunchtime. For a Latin person lunch is a very important time of the day. The first group was composed of: Marijn, Leonie, Sander and Martha. Dear Martha, thanks for sharing all your experiences on how to deal with the Dutch system!. The second generation included my dear friend Dasha, (te extrañamos mucho!, gracias por invitarme a ser tu paranimfo!). And the current one has been expanded with Sandrita, Lady, Giovanny, Natalia, Nerea and Adriana. Thanks for all the stories and the good time!.

My gratitude also to my former colleagues from Mexico: my master's promotor Professor Edgar Vallejo and to my colleagues from the INMEGEN: Dr. Jimenez, Alejandra, David, Juan Carlos, Laura U, Alfredo, Enrique and Irma. Dear Irma, thanks for being a good friend and for teaching me how to do science and for always giving your best not matter what.

I would like to thank all my friends outside work who have contributed to make my life much easier. Hector, Lulu, Andres por tener siempre abiertas las puertas de su casa y por las carnitas asadas en el parque con Salvatore y Giuseppe, también Lulú por todas las idas con Liz a Blaak por los preciados limones, indispensable en nuestra dieta mexicana. Monica, Paco y bebes por su grata compañía y charlas positivas, y sobre todo por los tips de bebes. Gladis, Eduardo, Osmar, Manuel y pequeñines por tener el gusto de verlos en el curso de holandés y fiestas de rigor. Gustavo, Aida y Daniel por ponerle el sabor y son cubano a las fiestas, gracias chicos. Esther, Pedro y niñas por compartir lindos momentos y vernos con tanto gusto en el Creche. Alejo, Sandra y Alejito por hacernos sentir como en casa en cada celebración, excelentes anfitriones sin duda. Un especial agradecimiento a Carolina Pardo por darme la bienvenida a su casa, por su inmensa paciencia en cada entrega de artículos y por compartir todas y cada una de las fiestas de Gabriela y Paulina. Y no puedo dejar de agradecer también a la ex-banda de Rotterdam, aunque la mayoría ya se hayan mudado a otros lados: Hanane, Robert, Jaibo, Corrales, Claudia y JC, Natalia, Luisa Luna. También mis agradecimientos a Pascal Wensen por ayudarme con las clases de holandes!, y a Agustín por rifarse con el diseño de la portada de este libro.

Finally, this book represents more than just a compilation of papers on a research topic. For me it also represents a tribute to my family which has always inspired me to continue doing research on the genetic basis of human disease. My eternal gratitude to my late father. Carnal (father), thanks to your guidance I'm now officially a scientific researcher, I hope I will be able to be at least as good father as you were. Mother, time and difficulties pass by but you always stay strong, thanks for showing me that no matter what life goes on. Dear love Liz, I do not have enough space to thank your love and for all you have done for me. Thanks for coming to the unpredictable low lands with me and for adapting to a very different way of living. Since we were college students you showed me how perseverance and discipline was the key for our mutual success. Also thank you for giving me the most well-behaved baby that this side of world has seen. Dear Oscar André, I thank you for filling in that part of my heart every day with your endless laughing. My sisters Geor and Vero and my brother Robert. Dear Geor, Lalo, Lalito and Luis Jorge, thanks for being always there when needed. I have learned from you that one should always

enjoy life as if it was our last day on earth. Dear Vero, Miguel Angel, Mariela and Many, thanks for showing me very early how great is to have family and for your support. Dear Robert, thanks for being always an inspiration. My cousin Mafi, thanks for being supportive and for coming to my thesis defense!. My gratitude extend to my family in law who have always promoted us to study abroad. My father in law, in particular, who enjoys genetics as I do.

It seems like it was yesterday when I arrived to the Rotterdam in 2008. Time flies, people come and go and it is very likely that I missed someone's name in this section, please accept my general gratitude to all who in some way contributed to this work.

LIST OF PUBLICATIONS

Papers published in indexed journals.

2007

1. **Estrada-Gil, J.K.**, Fernandez-Lopez, J.C., Hernandez-Lemus, E., Silva-Zolezzi, I., Hidalgo-Miranda, A., Jimenez-Sanchez, G. & Vallejo-Clemente, E.E. GPDTI: a Genetic Programming Decision Tree induction method to find epistatic effects in common complex diseases. ***Bioinformatics*** 23, i167-74 (2007).

2. Velazquez-Cruz, R., Orozco, L., Espinosa-Rosales, F., Carreno-Manjarrez, R., Solis-Vallejo, E., Lopez-Lara, N.D., Ruiz-Lopez, I.K., Rodriguez-Lozano, A.L., **Estrada-Gil, J.K.**, Jimenez-Sanchez, G. & Baca, V. Association of PDCD1 polymorphisms with childhood-onset systemic lupus erythematosus. ***Eur J Hum Genet*** 15, 336-41 (2007).

2009

3. Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A., Stephens, J., Wheeler, E., Arp, P., Gwilliam, R., Jhamai, P.M., Potter, S., Chaney, A., Ghori, M.J., Ravindrarajah, R., Ermakov, S., **Estrada, K.**, Pols, H.A., Williams, F.M., McArdle, W.L., van Meurs, J.B., Loos, R.J., Dermitzakis, E.T., Ahmadi, K.R., Hart, D.J., Ouwehand, W.H., Wareham, N.J., Barroso, I., Sandhu, M.S., Strachan, D.P., Livshits, G., Spector, T.D., Uitterlinden, A.G. & Deloukas, P. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. ***PLoS Genet*** 5, e1000445 (2009).

4. Newton-Cheh, C., Eijgelsheim, M., Rice, K.M., de Bakker, P.I., Yin, X., **Estrada, K.**, Bis, J.C., Marciante, K., Rivadeneira, F., Noseworthy, P.A., Sotoodehnia, N., Smith, N.L., Rotter, J.I., Kors, J.A., Witteman, J.C., Hofman, A., Heckbert, S.R., O'Donnell, C.J., Uitterlinden, A.G., Psaty, B.M., Lumley, T., Larson, M.G. & Stricker, B.H. Common variants at ten loci influence QT interval duration in the QTGEN Study. ***Nat Genet*** 41, 399-406 (2009).

5. Kwan, T., Grundberg, E., Koka, V., Ge, B., Lam, K.C., Dias, C., Kindmark, A., Mallmin, H., Ljunggren, O., Rivadeneira, F., **Estrada, K.**, van Meurs, J.B., Uitterlinden, A., Karlsson, M., Ohlsson, C., Mellstrom, D., Nilsson, O., Pastinen, T. & Majewski, J. Tissue effect on genetic control of transcript isoform variation. ***PLoS Genet*** 5, e1000608 (2009).

6. Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., Lettre, G., Lim, N., Lyon, H.N., McCarroll, S.A., Papadakis, K., Qi, L., Randall, J.C., Roccasecca, R.M., Sanna, S., Scheet, P., Weedon, M.N., Wheeler, E., Zhao, J.H., Jacobs, L.C., Prokopenko, I., Soranzo, N., Tanaka, T., Timpson, N.J., Almgren, P., Bennett, A., Bergman, R.N., Bingham, S.A., Bonnycastle, L.L., Brown, M., Burtt, N.P., Chines, P., Coin, L., Collins, F.S., Connell, J.M., Cooper, C., Smith, G.D., Dennison, E.M., Deodhar, P., Elliott, P., Erdos, M.R., **Estrada, K.**, Evans,

D.M., Gianniny, L., Gieger, C., Gillson, C.J., Guiducci, C., Hackett, R., Hadley, D., Hall, A.S., Havulinna, A.S., Hebebrand, J., Hofman, A., Isomaa, B., Jacobs, K.B., Johnson, T., Jousilahti, P., Jovanovic, Z., Khaw, K.T., Kraft, P., Kuokkanen, M., Kuusisto, J., Laitinen, J., Lakatta, E.G., Luan, J., Luben, R.N., Mangino, M., McArdle, W.L., Meitinger, T., Mulas, A., Munroe, P.B., Narisu, N., Ness, A.R., Northstone, K., O'Rahilly, S., Purmann, C., Rees, M.G., Ridderstrale, M., Ring, S.M., Rivadeneira, F., Ruokonen, A., Sandhu, M.S., Saramies, J., Scott, L.J., Scuteri, A., Silander, K., Sims, M.A., Song, K., Stephens, J., Stevens, S., Stringham, H.M., Tung, Y.C., Valle, T.T., Van Duijn, C.M., Vimaleswaran, K.S., Vollenweider, P., Waeber, G., Wallace, C., Watanabe, R.M., Waterworth, D.M., Watkins, N., Witteman, J.C., Zeggini, E., Zhai, G., Zillikens, M.C., Altshuler, D., Caulfield, M.J., Chanock, S.J., Farooqi, I.S., Ferrucci, L., Guralnik, J.M., Hattersley, A.T., Hu, F.B., Jarvelin, M.R., Laakso, M., Mooser, V., Ong, K.K., Ouwehand, W.H., Salomaa, V., Samani, N.J., Spector, T.D., Tuomi, T., Tuomilehto, J., Uda, M., Uitterlinden, A.G., Wareham, N.J., Deloukas, P., Frayling, T.M., Groop, L.C., Hayes, R.B., Hunter, D.J., Mohlke, K.L., Peltonen, L., Schlessinger, D., Strachan, D.P., Wichmann, H.E., McCarthy, M.I., Boehnke, M., Barroso, I., Abecasis, G.R. & Hirschhorn, J.N. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. ***Nat Genet*** 41, 25-34 (2009).

7. Heard-Costa, N.L., Zillikens, M.C., Monda, K.L., Johansson, A., Harris, T.B., Fu, M., Haritunians, T., Feitosa, M.F., Aspelund, T., Eiriksdottir, G., Garcia, M., Launer, L.J., Smith, A.V., Mitchell, B.D., McArdle, P.F., Shuldiner, A.R., Bielinski, S.J., Boerwinkle, E., Brancati, F., Demerath, E.W., Pankow, J.S., Arnold, A.M., Chen, Y.D., Glazer, N.L., McKnight, B., Psaty, B.M., Rotter, J.I., Amin, N., Campbell, H., Gyllensten, U., Pattaro, C., Pramstaller, P.P., Rudan, I., Struchalin, M., Vitart, V., Gao, X., Kraja, A., Province, M.A., Zhang, Q., Atwood, L.D., Dupuis, J., Hirschhorn, J.N., Jaquish, C.E., O'Donnell, C.J., Vasan, R.S., White, C.C., Aulchenko, Y.S., **Estrada, K.**, Hofman, A., Rivadeneira, F., Uitterlinden, A.G., Witteman, J.C., Oostra, B.A., Kaplan, R.C., Gudnason, V., O'Connell, J.R., Borecki, I.B., van Duijn, C.M., Cupples, L.A., Fox, C.S. & North, K.E. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. ***PLoS Genet*** 5, e1000539 (2009).

8. Lindgren, C.M., Heid, I.M., Randall, J.C., Lamina, C., Steinthorsdottir, V., Qi, L., Speliotes, E.K., Thorleifsson, G., Willer, C.J., Herrera, B.M., Jackson, A.U., Lim, N., Scheet, P., Soranzo, N., Amin, N., Aulchenko, Y.S., Chambers, J.C., Drong, A., Luan, J., Lyon, H.N., Rivadeneira, F., Sanna, S., Timpson, N.J., Zillikens, M.C., Zhao, J.H., Almgren, P., Bandinelli, S., Bennett, A.J., Bergman, R.N., Bonnycastle, L.L., Bumpstead, S.J., Chanock, S.J., Cherkas, L., Chines, P., Coin, L., Cooper, C., Crawford, G., Doering, A., Dominiczak, A., Doney, A.S., Ebrahim, S., Elliott, P., Erdos, M.R., **Estrada, K.**, Ferrucci, L., Fischer, G., Forouhi, N.G., Gieger, C., Grallert, H., Groves, C.J., Grundy, S., Guiducci, C., Hadley, D., Hamsten, A., Havulinna, A.S., Hofman, A., Holle, R., Holloway, J.W., Illig, T., Isomaa, B., Jacobs, L.C., Jameson, K., Jousilahti, P., Karpe, F., Kuusisto, J., Laitinen, J., Lathrop, G.M., Lawlor, D.A., Mangino, M., McArdle, W.L., Meitinger, T., Morken, M.A., Morris, A.P., Munroe, P., Narisu, N., Nordstrom, A., Nordstrom, P., Oostra, B.A., Palmer, C.N., Payne, F., Peden, J.F., Prokopenko, I., Renstrom, F., Ruokonen, A., Salomaa, V., Sandhu, M.S., Scott, L.J., Scuteri, A., Silander, K., Song, K., Yuan, X., Stringham, H.M., Swift, A.J., Tuomi, T., Uda, M., Vollenweider, P., Waeber, G., Wallace, C., Walters, G.B., Weedon, M.N., Witteman, J.C., Zhang, C., Zhang, W., Caulfield, M.J., Collins, F.S., Davey Smith, G., Day, I.N., Franks, P.W., Hattersley, A.T., Hu, F.B., Jarvelin, M.R., Kong, A., Kooner, J.S., Laakso, M., Lakatta, E., Mooser, V., Morris, A.D., Peltonen, L., Samani, N.J., Spector, T.D., Strachan, D.P., Tanaka, T., Tuomilehto, J., Uitterlinden, A.G., van Duijn, C.M., Wareham, N.J., Hugh, W., Waterworth, D.M., Boehnke, M., Deloukas, P., Groop, L., Hunter, D.J., Thorsteinsdottir, U., Schlessinger, D., Wichmann, H.E., Frayling, T.M., Abecasis, G.R., Hirschhorn, J.N., Loos, R.J., Stefansson, K., Mohlke, K.L., Barroso, I. & McCarthy, M.I. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. ***PLoS Genet*** 5, e1000508 (2009).

9. Perry, J.R., Stolk, L., Franceschini, N., Lunetta, K.L., Zhai, G., McArdle, P.F., Smith, A.V., Aspelund, T., Bandinelli, S., Boerwinkle, E., Cherkas, L., Eiriksdottir, G., **Estrada, K.**, Ferrucci, L., Folsom, A.R., Garcia, M., Gudnason, V., Hofman, A., Karasik, D., Kiel, D.P., Launer, L.J., van Meurs, J., Nalls, M.A., Rivadeneira, F., Shuldiner, A.R., Singleton, A., Soranzo, N., Tanaka, T., Visser, J.A., Weedon, M.N., Wilson, S.G., Zhuang, V., Streeten, E.A., Harris, T.B., Murray, A., Spector, T.D., Demerath, E.W., Uitterlinden, A.G. & Murabito, J.M. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. ***Nat Genet*** (2009).

10. Stolk, L., Zhai, G., van Meurs, J.B., Verbiest, M.M., Visser, J.A., **Estrada, K.**, Rivadeneira, F., Williams, F.M., Cherkas, L., Deloukas, P., Soranzo, N., de Keyzer, J.J., Pop, V.J., Lips, P., Lebrun, C.E., van der Schouw, Y.T., Grobbee, D.E., Witteman, J., Hofman, A., Pols, H.A., Laven, J.S., Spector, T.D. & Uitterlinden, A.G. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. ***Nat Genet*** (2009).

11. Silva-Zolezzi, I., Hidalgo-Miranda, A., **Estrada-Gil, J.**, Fernandez-Lopez, J.C., Uribe-Figueroa, L., Contreras, A., Balam-Ortiz, E., del Bosque-Plata, L., Velazquez-Fernandez, D., Lara, C., Goya, R., Hernandez-Lemus, E., Davila, C., Barrientos, E., March, S. & Jimenez-Sanchez, G. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. ***Proc Natl Acad Sci U S A*** 106, 8611-6 (2009).

12. Rivadeneira, F., Styrkarsdottir, U., **Estrada, K.**, Halldorsson, B.V., Hsu, Y.H., Richards, J.B., Zillikens, M.C., Kavvoura, F.K., Amin, N., Aulchenko, Y.S., Cupples, L.A., Deloukas, P., Demissie, S., Grundberg, E., Hofman, A., Kong, A., Karasik, D., van Meurs, J.B., Oostra, B., Pastinen, T., Pols, H.A., Sigurdsson, G., Soranzo, N., Thorleifsson, G., Thorsteinsdottir, U., Williams, F.M., Wilson, S.G., Zhou, Y., Ralston, S.H., van Duijn, C.M., Spector, T., Kiel, D.P., Stefansson, K., Ioannidis, J.P. & Uitterlinden, A.G. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. ***Nat Genet*** 41, 1199-206 (2009).

13. Grundberg, E., Kwan, T., Ge, B., Lam, K.C., Koka, V., Kindmark, A., Mallmin, H., Dias, J., Verlaan, D.J., Ouimet, M., Sinnett, D., Rivadeneira, F., **Estrada, K.**, Hofman, A., van Meurs, J.M., Uitterlinden, A., Beaulieu, P., Graziani, A., Harmsen, E., Ljunggren, O., Ohlsson, C., Mellstrom, D., Karlsson, M.K., Nilsson, O. & Pastinen, T. Population genomics in a disease targeted primary cell model. ***Genome Res*** 19, 1942-52 (2009).

14. van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H., van der Kooi, A.J., de Visser, M., Dahlberg, C., **Estrada, K.**, Rivadeneira, F., Hofman, A., Zwarts, M.J., van Doormaal, P.T., Rujescu, D., Strengman, E., Giegling, I., Muglia, P., Tomik, B., Slowik, A., Uitterlinden, A.G., Hendrich, C., Waibel, S., Meyer, T., Ludolph, A.C., Glass, J.D., Purcell, S., Cichon, S., Nothen, M.M., Wichmann, H.E., Schreiber, S., Vermeulen, S.H., Kiemeney, L.A., Wokke, J.H., Cronin, S., McLaughlin, R.L., Hardiman, O., Fumoto, K., Pasterkamp, R.J., Meininger, V., Melki, J., Leigh, P.N., Shaw, C.E., Landers, J.E., Al-Chalabi, A., Brown, R.H., Jr., Robberecht, W., Andersen, P.M., Ophoff, R.A. & van den Berg, L.H. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. ***Nat Genet*** 41, 1083-7 (2009).

15. **Estrada, K.**, Abuseiris, A., Grosveld, F.G., Uitterlinden, A.G., Knoch, T.A. & Rivadeneira, F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. ***Bioinformatics*** 25, 2750-2 (2009).

16. Richards, J.B., Kavvoura, F.K., Rivadeneira, F., Styrkarsdottir, U., **Estrada, K.**, Halldorsson, B.V., Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Mullin, B.H., Amin, N., Aulchenko, Y.S., Cupples, L.A., Deloukas, P., Demissie, S., Hofman, A., Kong, A., Karasik, D., van Meurs, J.B., Oostra, B.A., Pols, H.A., Sigurdsson, G., Thorsteinsdottir, U., Soranzo, N., Williams, F.M., Zhou, Y., Ralston, S.H., Thorleifsson, G., van Duijn, C.M., Kiel, D.P., Stefansson, K., Uitterlinden, A.G., Ioannidis, J.P. & Spector, T.D. Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. ***Ann Intern Med*** 151, 528-37 (2009).

17. **Estrada, K.**, Krawczak, M., Schreiber, S., van Duijn, K., Stolk, L., van Meurs, J.B., Liu, F., Penninx, B.W., Smit, J.H., Vogelzangs, N., Hottenga, J.J., Willemsen, G., de Geus, E.J., Lorentzon, M., von Eller-Eberstein, H., Lips, P., Schoor, N., Pop, V., de Keijzer, J., Hofman, A., Aulchenko, Y.S., Oostra, B.A., Ohlsson, C., Boomsma, D.I., Uitterlinden, A.G., van Duijn, C.M., Rivadeneira, F. & Kayser, M. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. ***Hum Mol Genet*** 18, 3516-24 (2009).

2010

18. Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Farber, C.R., Demissie, S., Soranzo, N., Bianchi, E.N., Grundberg, E., Liang, L., Richards, J.B., **Estrada, K.**, Zhou, Y., van Nas, A., Moffatt, M.F., Zhai, G., Hofman, A., van Meurs, J.B., Pols, H.A., Price, R.I., Nilsson, O., Pastinen, T., Cupples, L.A., Lusis, A.J., Schadt, E.E., Ferrari, S., Uitterlinden, A.G., Rivadeneira, F., Spector, T.D., Karasik, D. & Kiel, D.P. An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. ***PLoS Genet*** 6, e1000977 (2010).

19. Sotoodehnia, N., Isaacs, A., de Bakker, P.I., Dorr, M., Newton-Cheh, C., Nolte, I.M., van der Harst, P., Muller, M., Eijgelsheim, M., Alonso, A., Hicks, A.A., Padmanabhan, S., Hayward, C., Smith, A.V., Polasek, O., Giovannone, S., Fu, J., Magnani, J.W., Marciante, K.D., Pfeufer, A., Gharib, S.A., Teumer, A., Li, M., Bis, J.C., Rivadeneira, F., Aspelund, T., Kottgen, A., Johnson, T., Rice, K., Sie, M.P., Wang, Y.A., Klopp, N., Fuchsberger, C., Wild, S.H., Mateo Leach, I., **Estrada, K.**, Volker, U., Wright, A.F., Asselbergs, F.W., Qu, J., Chakravarti, A., Sinner, M.F., Kors, J.A., Petersmann, A., Harris, T.B., Soliman, E.Z., Munroe, P.B., Psaty, B.M., Oostra, B.A., Cupples, L.A., Perz, S., de Boer, R.A., Uitterlinden, A.G., Volzke, H., Spector, T.D., Liu, F.Y., Boerwinkle, E., Dominiczak, A.F., Rotter, J.I., van Herpen, G., Levy, D., Wichmann, H.E., van Gilst, W.H., Witteman, J.C., Kroemer, H.K., Kao, W.H., Heckbert, S.R., Meitinger, T., Hofman, A., Campbell, H., Folsom, A.R., van Veldhuisen, D.J., Schwienbacher, C., O'Donnell, C.J., Volpato, C.B., Caulfield, M.J., Connell, J.M., Launer, L., Lu, X., Franke, L., Fehrmann, R.S., te Meerman, G., Groen, H.J., Weersma, R.K., van den Berg, L.H., Wijmenga, C., Ophoff, R.A., Navis, G., Rudan, I., Snieder, H., Wilson, J.F., Pramstaller, P.P., Siscovick, D.S., Wang, T.J., Gudnason, V., van Duijn, C.M., Felix, S.B., Fishman, G.I., Jamshidi, Y., Stricker, B.H., Samani, N.J., Kaab, S. & Arking, D.E. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. ***Nat Genet*** 42, 1068-76 (2010).

20. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R., Workalemahu, T., White, C.C., Bouatia-Naji, N., Harris, T.B., Berndt, S.I., Ingelsson, E., Willer, C.J., Weedon, M.N., Luan, J., Vedantam, S., Esko, T., Kilpelainen, T.O., Kutalik, Z., Li, S., Monda, K.L., Dixon, A.L., Holmes, C.C., Kaplan, L.M., Liang, L., Min, J.L., Moffatt, M.F., Molony, C., Nicholson, G., Schadt, E.E., Zondervan, K.T., Feitosa, M.F., Ferreira, T., Allen, H.L., Weyant, R.J., Wheeler, E., Wood, A.R., **Estrada, K.**, Goddard, M.E., Lettre, G., Mangino, M., Nyholt, D.R., Purcell, S., Smith, A.V., Visscher, P.M., Yang, J., McCarroll, S.A., Nemesh, J., Voight, B.F., Absher, D., Amin, N., Aspelund, T., Coin, L., Glazer, N.L., Hayward, C., Heard-Costa, N.L., Hottenga, J.J., Johansson, A., Johnson, T., Kaakinen, M., Kapur, K., Ketkar, S., Knowles, J.W., Kraft, P., Kraja, A.T., Lamina, C., Leitzmann, M.F., McKnight, B., Morris, A.P., Ong, K.K., Perry, J.R., Peters, M.J., Polasek, O., Prokopenko, I., Rayner, N.W., Ripatti, S., Rivadeneira, F., Robertson, N.R., Sanna, S., Sovio, U., Surakka, I., Teumer, A., van Wingerden, S., Vitart, V., Zhao, J.H., Cavalcanti-Proenca, C., Chines, P.S., Fisher, E., Kulzer, J.R., Lecoeur, C., Narisu, N., Sandholt, C., Scott, L.J., Silander, K., Stark, K., Tammesoo, M.L., Teslovich, T.M., Timpson, N.J., Watanabe, R.M., Welch, R., Chasman, D.I., Cooper, M.N., Jansson, J.O., Kettunen, J., Lawrence, R.W., Pellikka, N., Perola, M., Vandenput, L., Alavere, H., Almgren, P., Atwood, L.D., Bennett, A.J., Biffar, R., Bonnycastle, L.L., Bornstein, S.R., Buchanan, T.A., Campbell, H., Day, I.N., Dei, M., Dorr, M., Elliott, P., Erdos, M.R., Eriksson, J.G., Freimer, N.B., Fu, M., Gaget, S., Geus, E.J., Gjesing, A.P., Grallert, H., Grassler, J., Groves, C.J., Guiducci, C., Hartikainen, A.L., Hassanali, N., Havulinna, A.S., Herzig, K.H., Hicks, A.A., Hui, J., Igl, W., Jousilahti, P., Jula, A., Kajantie, E., Kinnunen, L., Kolcic, I., Koskinen, S., Kovacs, P., Kroemer, H.K., Krzelj, V., Kuusisto, J., Kvaloy, K., Laitinen, J., Lantieri, O., Lathrop, G.M., Lokki, M.L., Luben, R.N., Ludwig, B., McArdle, W.L., McCarthy, A., Morken, M.A., Nelis, M., Neville, M.J., Pare, G., Parker, A.N., Peden, J.F., Pichler, I., Pietilainen, K.H., Platou, C.G., Pouta, A., Ridderstrale, M., Samani, N.J., Saramies, J., Sinisalo, J., Smit, J.H., Strawbridge, R.J., Stringham, H.M., Swift, A.J., Teder-Laving, M., Thomson, B., Usala, G., van Meurs, J.B., van Ommen, G.J., Vatin, V., Volpato, C.B., Wallaschofski, H., Walters, G.B., Widen, E., Wild, S.H., Willemsen, G., Witte, D.R., Zgaga, L., Zitting, P., Beilby, J.P., James, A.L., Kahonen, M., Lehtimaki, T., Nieminen, M.S., Ohlsson, C., Palmer, L.J., Raitakari, O., Ridker, P.M., Stumvoll, M., Tonjes, A., Viikari, J., Balkau, B., Ben-Shlomo, Y., Bergman, R.N., Boeing, H., Smith, G.D., Ebrahim, S., Froguel, P., Hansen, T., Hengstenberg, C., Hveem, K., Isomaa, B., Jorgensen, T., Karpe, F., Khaw, K.T., Laakso, M., Lawlor, D.A., Marre, M., Meitinger, T., Metspalu, A., Midthjell, K., Pedersen, O., Salomaa, V., Schwarz, P.E., Tuomi, T., Tuomilehto, J., Valle, T.T., Wareham, N.J., Arnold, A.M., Beckmann, J.S., Bergmann, S., Boerwinkle, E., Boomsma, D.I., Caulfield, M.J., Collins, F.S., Eiriksdottir, G., Gudnason, V., Gyllensten, U., Hamsten, A., Hattersley, A.T., Hofman, A., Hu, F.B., Illig, T., Iribarren, C., Jarvelin, M.R., Kao, W.H., Kaprio, J., Launer, L.J., Munroe, P.B., Oostra, B., Penninx, B.W., Pramstaller, P.P., Psaty, B.M., Quertermous, T., Rissanen, A., Rudan, I., Shuldiner, A.R., Soranzo, N., Spector, T.D., Syvanen, A.C., Uda, M., Uitterlinden, A., Volzke, H., Vollenweider, P., Wilson, J.F., Witteman, J.C., Wright, A.F., Abecasis, G.R., Boehnke, M., Borecki, I.B., Deloukas, P., Frayling, T.M., Groop, L.C., Haritunians, T., Hunter, D.J., Kaplan, R.C., North, K.E., O'Connell, J.R., Peltonen, L., Schlessinger, D., Strachan, D.P., Hirschhorn, J.N., Assimes, T.L., Wichmann, H.E., Thorsteinsdottir, U., van Duijn, C.M., Stefansson, K., Cupples, L.A., Loos, R.J., Barroso, I., McCarthy, M.I., Fox, C.S., Mohlke, K.L. & Lindgren, C.M. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. ***Nat Genet*** 42, 949-60 (2010).

21. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Allen, H.L., Lindgren, C.M., Luan, J., Magi, R., Randall, J.C., Vedantam, S., Winkler, T.W., Qi, L., Workalemahu, T., Heid, I.M., Steinthorsdottir, V., Stringham, H.M., Weedon, M.N., Wheeler, E., Wood, A.R., Ferreira, T., Weyant, R.J., Segre, A.V., **Estrada, K.**, Liang, L., Nemesh, J., Park, J.H., Gustafsson, S., Kilpelainen, T.O., Yang, J., Bouatia-Naji, N., Esko, T., Feitosa, M.F., Kutalik, Z., Mangino, M., Raychaudhuri, S., Scherag, A., Smith, A.V., Welch, R., Zhao, J.H., Aben, K.K., Absher, D.M., Amin,

N., Dixon, A.L., Fisher, E., Glazer, N.L., Goddard, M.E., Heard-Costa, N.L., Hoesel, V., Hottenga, J.J., Johansson, A., Johnson, T., Ketkar, S., Lamina, C., Li, S., Moffatt, M.F., Myers, R.H., Narisu, N., Perry, J.R., Peters, M.J., Preuss, M., Ripatti, S., Rivadeneira, F., Sandholt, C., Scott, L.J., Timpson, N.J., Tyrer, J.P., van Wingerden, S., Watanabe, R.M., White, C.C., Wiklund, F., Barlassina, C., Chasman, D.I., Cooper, M.N., Jansson, J.O., Lawrence, R.W., Pellikka, N., Prokopenko, I., Shi, J., Thiering, E., Alavere, H., Alibrandi, M.T., Almgren, P., Arnold, A.M., Aspelund, T., Atwood, L.D., Balkau, B., Balmforth, A.J., Bennett, A.J., Ben-Shlomo, Y., Bergman, R.N., Bergmann, S., Biebermann, H., Blakemore, A.I., Boes, T., Bonnycastle, L.L., Bornstein, S.R., Brown, M.J., Buchanan, T.A., Busonero, F., Campbell, H., Cappuccio, F.P., Cavalcanti-Proenca, C., Chen, Y.D., Chen, C.M., Chines, P.S., Clarke, R., Coin, L., Connell, J., Day, I.N., Heijer, M., Duan, J., Ebrahim, S., Elliott, P., Elosua, R., Eiriksdottir, G., Erdos, M.R., Eriksson, J.G., Facheris, M.F., Felix, S.B., Fischer-Posovszky, P., Folsom, A.R., Friedrich, N., Freimer, N.B., Fu, M., Gaget, S., Gejman, P.V., Geus, E.J., Gieger, C., Gjesing, A.P., Goel, A., Goyette, P., Grallert, H., Grassler, J., Greenawalt, D.M., Groves, C.J., Gudnason, V., Guiducci, C., Hartikainen, A.L., Hassanali, N., Hall, A.S., Havulinna, A.S., Hayward, C., Heath, A.C., Hengstenberg, C., Hicks, A.A., Hinney, A., Hofman, A., Homuth, G., Hui, J., Igl, W., Iribarren, C., Isomaa, B., Jacobs, K.B., Jarick, I., Jewell, E., John, U., Jorgensen, T., Jousilahti, P., Jula, A., Kaakinen, M., Kajantie, E., Kaplan, L.M., Kathiresan, S., Kettunen, J., Kinnunen, L., Knowles, J.W., Kolcic, I., Konig, I.R., Koskinen, S., Kovacs, P., Kuusisto, J., Kraft, P., Kvaloy, K., Laitinen, J., Lantieri, O., Lanzani, C., Launer, L.J., Lecoeur, C., Lehtimaki, T., Lettre, G., Liu, J., Lokki, M.L., Lorentzon, M., Luben, R.N., Ludwig, B., Manunta, P., Marek, D., Marre, M., Martin, N.G., McArdle, W.L., McCarthy, A., McKnight, B., Meitinger, T., Melander, O., Meyre, D., Midthjell, K., Montgomery, G.W., Morken, M.A., Morris, A.P., Mulic, R., Ngwa, J.S., Nelis, M., Neville, M.J., Nyholt, D.R., O'Donnell, C.J., O'Rahilly, S., Ong, K.K., Oostra, B., Pare, G., Parker, A.N., Perola, M., Pichler, I., Pietilainen, K.H., Platou, C.G., Polasek, O., Pouta, A., Rafelt, S., Raitakari, O., Rayner, N.W., Ridderstrale, M., Rief, W., Ruokonen, A., Robertson, N.R., Rzehak, P., Salomaa, V., Sanders, A.R., Sandhu, M.S., Sanna, S., Saramies, J., Savolainen, M.J., Scherag, S., Schipf, S., Schreiber, S., Schunkert, H., Silander, K., Sinisalo, J., Siscovick, D.S., Smit, J.H., Soranzo, N., Sovio, U., Stephens, J., Surakka, I., Swift, A.J., Tammesoo, M.L., Tardif, J.C., Teder-Laving, M., Teslovich, T.M., Thompson, J.R., Thomson, B., Tonjes, A., Tuomi, T., van Meurs, J.B., van Ommen, G.J., Vatin, V., Viikari, J., Visvikis-Siest, S., Vitart, V., Vogel, C.I., Voight, B.F., Waite, L.L., Wallaschofski, H., Walters, G.B., Widen, E., Wiegand, S., Wild, S.H., Willemsen, G., Witte, D.R., Witteman, J.C., Xu, J., Zhang, Q., Zgaga, L., Ziegler, A., Zitting, P., Beilby, J.P., Farooqi, I.S., Hebebrand, J., Huikuri, H.V., James, A.L., Kahonen, M., Levinson, D.F., Macciardi, F., Nieminen, M.S., Ohlsson, C., Palmer, L.J., Ridker, P.M., Stumvoll, M., Beckmann, J.S., Boeing, H., Boerwinkle, E., Boomsma, D.I., Caulfield, M.J., Chanock, S.J., Collins, F.S., Cupples, L.A., Smith, G.D., Erdmann, J., Froguel, P., Gronberg, H., Gyllensten, U., Hall, P., Hansen, T., Harris, T.B., Hattersley, A.T., Hayes, R.B., Heinrich, J., Hu, F.B., Hveem, K., Illig, T., Jarvelin, M.R., Kaprio, J., Karpe, F., Khaw, K.T., Kiemeney, L.A., Krude, H., Laakso, M., Lawlor, D.A., Metspalu, A., Munroe, P.B., Ouwehand, W.H., Pedersen, O., Penninx, B.W., Peters, A., Pramstaller, P.P., Quertermous, T., Reinehr, T., Rissanen, A., Rudan, I., Samani, N.J., Schwarz, P.E., Shuldiner, A.R., Spector, T.D., Tuomilehto, J., Uda, M., Uitterlinden, A., Valle, T.T., Wabitsch, M., Waeber, G., Wareham, N.J., Watkins, H., Wilson, J.F., Wright, A.F., Zillikens, M.C., Chatterjee, N., McCarroll, S.A., Purcell, S., Schadt, E.E., Visscher, P.M., Assimes, T.L., Borecki, I.B., Deloukas, P., Fox, C.S., Groop, L.C., Haritunians, T., Hunter, D.J., Kaplan, R.C., Mohlke, K.L., O'Connell, J.R., Peltonen, L., Schlessinger, D., Strachan, D.P., van Duijn, C.M., Wichmann, H.E., Frayling, T.M., Thorsteinsdottir, U., Abecasis, G.R., Barroso, I., Boehnke, M., Stefansson, K., North, K.E., McCarthy, M.I., Hirschhorn, J.N., Ingelsson, E. & Loos, R.J. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. ***Nat Genet*** 42, 937-48 (2010).

22. Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P.I., Muller, M., Morrison, A.C., Smith, A.V., Isaacs, A., Sanna, S., Dorr, M., Navarro, P., Fuchsberger, C., Nolte, I.M., de Geus, E.J., **Estrada, K.**, Hwang, S.J., Bis, J.C., Ruckert, I.M., Alonso,

A., Launer, L.J., Hottenga, J.J., Rivadeneira, F., Noseworthy, P.A., Rice, K.M., Perz, S., Arking, D.E., Spector, T.D., Kors, J.A., Aulchenko, Y.S., Tarasov, K.V., Homuth, G., Wild, S.H., Marroni, F., Gieger, C., Licht, C.M., Prineas, R.J., Hofman, A., Rotter, J.I., Hicks, A.A., Ernst, F., Najjar, S.S., Wright, A.F., Peters, A., Fox, E.R., Oostra, B.A., Kroemer, H.K., Couper, D., Volzke, H., Campbell, H., Meitinger, T., Uda, M., Witteman, J.C., Psaty, B.M., Wichmann, H.E., Harris, T.B., Kaab, S., Siscovick, D.S., Jamshidi, Y., Uitterlinden, A.G., Folsom, A.R., Larson, M.G., Wilson, J.F., Penninx, B.W., Snieder, H., Pramstaller, P.P., van Duijn, C.M., Lakatta, E.G., Felix, S.B., Gudnason, V., Pfeufer, A., Heckbert, S.R., Stricker, B.H., Boerwinkle, E. & O'Donnell, C.J. Genome-wide association analysis identifies multiple loci related to resting heart rate. ***Hum Mol Genet*** 19, 3885-94 (2010).

23. 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., Ferreira, T., Wood, A.R., Weyant, R.J., Segre, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Magi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Hua Zhao, J., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R., Surakka, I., Tammesoo, M.L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., Konig, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Muller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J., Glorioso, N., Haiqing, S., Hartikainen, A.L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietilainen, K.H., Pouta, A., Ridderstrale, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Oliver Schmidt, C., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimaki, T., Melander, O., Mosley, T.H., Jr., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G.J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Mark Lathrop, G., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Adrienne Cupples, L., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Volzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis,

G.R., Stefansson, K., Frayling, T.M. & Hirschhorn, J.N. Hundreds of variants clustered in genomic loci and biological pathways affect human height. ***Nature*** 467, 832-8 (2010).

24. Blauw, H.M., Al-Chalabi, A., Andersen, P.M., van Vught, P.W., Diekstra, F.P., van Es, M.A., Saris, C.G., Groen, E.J., van Rheenen, W., Koppers, M., Van't Slot, R., Strengman, E., **Estrada, K.**, Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Kiemeney, L.A., Vermeulen, S.H., Birve, A., Waibel, S., Meyer, T., Cronin, S., McLaughlin, R.L., Hardiman, O., Sapp, P.C., Tobin, M.D., Wain, L.V., Tomik, B., Slowik, A., Lemmens, R., Rujescu, D., Schulte, C., Gasser, T., Brown, R.H., Jr., Landers, J.E., Robberecht, W., Ludolph, A.C., Ophoff, R.A., Veldink, J.H. & van den Berg, L.H. A large genome scan for rare CNVs in amyotrophic lateral sclerosis. ***Hum Mol Genet*** 19, 4091-9 (2010).

2011

25. Duncan, E.L., Danoy, P., Kemp, J.P., Leo, P.J., McCloskey, E., Nicholson, G.C., Eastell, R., Prince, R.L., Eisman, J.A., Jones, G., Sambrook, P.N., Reid, I.R., Dennison, E.M., Wark, J., Richards, J.B., Uitterlinden, A.G., Spector, T.D., Esapa, C., Cox, R.D., Brown, S.D., Thakker, R.V., Addison, K.A., Bradbury, L.A., Center, J.R., Cooper, C., Cremin, C., **Estrada, K.**, Felsenberg, D., Gluer, C.C., Hadler, J., Henry, M.J., Hofman, A., Kotowicz, M.A., Makovey, J., Nguyen, S.C., Nguyen, T.V., Pasco, J.A., Pryce, K., Reid, D.M., Rivadeneira, F., Roux, C., Stefansson, K., Styrkarsdottir, U., Thorleifsson, G., Tichawangana, R., Evans, D.M. & Brown, M.A. Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. ***PLoS Genet*** 7, e1001372 (2011).

26. Kilpelainen, T.O., Zillikens, M.C., Stancakova, A., Finucane, F.M., Ried, J.S., Langenberg, C., Zhang, W., Beckmann, J.S., Luan, J., Vandenput, L., Styrkarsdottir, U., Zhou, Y., Smith, A.V., Zhao, J.H., Amin, N., Vedantam, S., Shin, S.Y., Haritunians, T., Fu, M., Feitosa, M.F., Kumari, M., Halldorsson, B.V., Tikkanen, E., Mangino, M., Hayward, C., Song, C., Arnold, A.M., Aulchenko, Y.S., Oostra, B.A., Campbell, H., Cupples, L.A., Davis, K.E., Doring, A., Eiriksdottir, G., **Estrada, K.**, Fernandez-Real, J.M., Garcia, M., Gieger, C., Glazer, N.L., Guiducci, C., Hofman, A., Humphries, S.E., Isomaa, B., Jacobs, L.C., Jula, A., Karasik, D., Karlsson, M.K., Khaw, K.T., Kim, L.J., Kivimaki, M., Klopp, N., Kuhnel, B., Kuusisto, J., Liu, Y., Ljunggren, O., Lorentzon, M., Luben, R.N., McKnight, B., Mellstrom, D., Mitchell, B.D., Mooser, V., Moreno, J.M., Mannisto, S., O'Connell, J.R., Pascoe, L., Peltonen, L., Peral, B., Perola, M., Psaty, B.M., Salomaa, V., Savage, D.B., Semple, R.K., Skaric-Juric, T., Sigurdsson, G., Song, K.S., Spector, T.D., Syvanen, A.C., Talmud, P.J., Thorleifsson, G., Thorsteinsdottir, U., Uitterlinden, A.G., van Duijn, C.M., Vidal-Puig, A., Wild, S.H., Wright, A.F., Clegg, D.J., Schadt, E., Wilson, J.F., Rudan, I., Ripatti, S., Borecki, I.B., Shuldiner, A.R., Ingelsson, E., Jansson, J.O., Kaplan, R.C., Gudnason, V., Harris, T.B., Groop, L., Kiel, D.P., Rivadeneira, F., Walker, M., Barroso, I., Vollenweider, P., Waeber, G., Chambers, J.C., Kooner, J.S., Soranzo, N., Hirschhorn, J.N., Stefansson, K., Wichmann, H.E., Ohlsson, C., O'Rahilly, S., Wareham, N.J., Speliotes, E.K., Fox, C.S., Laakso, M. & Loos, R.J. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. ***Nat Genet*** 43, 753-60 (2011).

27. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., **Estrada, K.**, Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., Magi, R., Madden, P.A., Heath, A.C., Nyholt, D.R., Martin, N.G., Montgomery, G.W., Frayling, T.M., Hirschhorn, J.N., McCarthy, M.I., Goddard, M.E. & Visscher, P.M. Genomic inflation factors under polygenic inheritance. ***Eur J Hum Genet*** 19, 807-12 (2011).

28. Kuningas, M., **Estrada, K.**, Hsu, Y.H., Nandakumar, K., Uitterlinden, A.G., Lunetta, K.L., van Duijn, C.M., Karasik, D., Hofman, A., Murabito, J., Rivadeneira, F., Kiel, D.P. & Tiemeier, H. Large common deletions associate with mortality at old age. ***Hum Mol Genet*** 20, 4290-6 (2011).

29. Day-Williams, A.G., Southam, L., Panoutsopoulou, K., Rayner, N.W., Esko, T., **Estrada, K.**, Helgadottir, H.T., Hofman, A., Ingvarsson, T., Jonsson, H., Keis, A., Kerkhof, H.J., Thorleifsson, G., Arden, N.K., Carr, A., Chapman, K., Deloukas, P., Loughlin, J., McCaskie, A., Ollier, W.E., Ralston, S.H., Spector, T.D., Wallis, G.A., Wilkinson, J.M., Aslam, N., Birell, F., Carluke, I., Joseph, J., Rai, A., Reed, M., Walker, K., Doherty, S.A., Jonsdottir, I., Maciewicz, R.A., Muir, K.R., Metspalu, A., Rivadeneira, F., Stefansson, K., Styrkarsdottir, U., Uitterlinden, A.G., van Meurs, J.B., Zhang, W., Valdes, A.M., Doherty, M. & Zeggini, E. A variant in MCF2L is associated with osteoarthritis. ***Am J Hum Genet*** 89, 446-50 (2011).

2012

30. **Estrada, K.**, Styrkarsdottir, U., Evangelou, E., Hsu, Y., Duncan, E.L., Ntzani, E.E., Oei, L., Albagha, O.M.E., Amin, N., Kemp, J.P., Koller, D.L., Li, G., Liu, C., Minster, R.L., Moayyeri, A., Vandenput, L., Willner, D., Xiao, S., Yerges-Armstrong, L.M., Zheng, H., Alonso, N., Eriksson, J., Kammerer, C.M., Kaptoge, S.K., Leo, P.J., Thorleifsson, G., Wilson, J.F., Wilson, S.G., Aalto, V., Alen, M., Aragaki, A.K., Aspelund, T., Center, J.R., Dailiana, Z., Duggan, D.J., Garcia, M., García-giralt, N., Giroux, S., Hallmans, G., Hocking, L.J., Husted, L.B., Jameson, K.A., Khusainova, R., Kim, G., Kooperberg, C., Koromila, T., Kruk, M., Laaksonen, M., Lacroix, A.Z., Lee, S., Leung, P.C., Lewis, J.R., Masi, L., Mencej-Bedrac, S., Nguyen, T.V., Nogues, X., Patel, M.S., Prezelj, J., Rose, L.M., Scollen, S., Siggeirsdottir, K., Smith, A.V., Svensson, O., Trompet, S., Trummer, O., van Schoor, N.M., Woo, J., Zhu, K., Balcells, S., Brandi, M., Buckley, B.M., Cheng, S., Christiansen, C., Cooper, C., Dedoussis, G., Ford, I., Frost, M., Goltzman, D., González-Macías, J., Kähönen, M., Karlsson, M., Khusnutdinova, E., Koh, J., Kollia, P., Langdahl, B.L., Leslie, W.D., Lips, P., Ljunggren, Ö., Lorenc, R.S., Marc, J., Mellström, D., Obermayer-Pietsch, B., Olmos, J.M., Pettersson, U., Reid, D.M., Riancho, J.A., Ridker, P.M., Rousseau, F., Slagboom, P.E., Tang, N.L.S., Urreizti, R., van Hul, W., Viikari, J., Zarrabeitia, M.T., Aulchenko , Y.S., Castano-Betancourt, M., Grundberg, E., Herrera, L., Ingvarsson, T., Johannsdottir, H., Kwan, T., Li, R., Luben, R., Medina-Gómez, C., Palsson, S.Th., Reppe, S., Rotter, J.I., Sigurdsson, G., van Meurs, J.B.J., Verlaan, D., Williams, F.M.K., Wood, A.R., Zhou, Y., Gautvik, K.M., Pastinen, T., Cauley, J.A., Chasman, D.I., Clark, G.R., Cummings, S.R., Danoy, P., Dennison, E.M., Eastell, R., Eisman, J.A., Gudnason, V., Hofman, A., Jackson, R.D., Jones, G., Jukema, J.W., Khaw, K., Lehtimäki, T., Liu, Y., Lorentzon, M., McCloskey, E., Mitchell, B.D., Nandakumar, K., Nicholson, G.C., Oostra, B.A., Peacock, M., Pols, H.A.P., Prince, R.L., Raitakari, O., Reid, I.R., Robbins, J., Sambrook, P.N., Sham, P.C., Shuldiner , A.R., Tylavsky, F.A., van Duijn, C.M., Wareham, N.J., Cupples, L.A., Econs, M.J., Evans, D.M., Harris, T.B., Kung, A.W.C., Psaty, B.M., Reeve, J., Spector, T.D., Streeten, E.A., Zillikens, M., Thorsteinsdottir, U., Ohlsson, C., Karasik, D., Richards, J.B., Brown, M.A., Stefansson, K., Uitterlinden, A.G., Ralston, S.H., Ioannidis, J.P.A., Kiel, D.P., Rivadeneira, F. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. ***Nature Genetics*** advance online publication, 15 April 2012; doi:10.1038/ng.2249.(2012).

PHD-PORTFOLIO

Name: Karol Estrada
Erasmus MC Department: Internal Medicine – Genetic Laboratory
Research School: Netherlands Institute for Health Sciences (NIHES) and
School of Molecular Medicine (MolMed)
PhD Period: April 2008 – April 2012
Promotor: Prof. dr. A.G. Uitterlinden
Supervisor: Dr. F. Rivadeneira

PhD training

Name	Year	Workload
General research skills		
Biomedical English Writing and Communication	2008	13x3 hours
InDesign Course	2011	1 day
Research courses		
BC SNPMax course	2008	1 week
R-course	2008	1 week
Genome-Wide Association Studies	2008	1 week
Causal Inference	2009	1 week
Regression Analysis	2009	1 week
Survival Analysis	2009	1 week
Analysis of repeated measurements	2010	1 week
Mendelian randomization	2011	1 week
Seminars and workshops		
Weekly scientific seminars Genetics Lab	2008-2012	2 ECTS
Next generation Sequencing data analysis	2010	1 week
Local Presentations		
CHARGE meeting: Large-scale Copy Number Analysis in population-based cohorts	2009	April 16th
Dept. of Epidemiology: "Identification of Common Polygenic Variation from many Loci Highlights the Complex Genetic Architecture Underlying Adult Height"	2009	November 17th
Dept. Internal Medicine: "Studies of adult height, Understanding the Architecture of Polygenic Traits"	2010	February 25th

Dept. Internal Medicine: “Genetic Factors of Osteoporosis”	2011	April 7th
(Inter)national conferences and presentations		
Illumina Users Group Meeting, Munich.	2008	3 days
European Society of Calcified Tissues, Barcelona	2008	4 days
European Society of Human Genetics, Barcelona	2008	4 days
Genomic Disorders Meeting, Hinxton	2009	3 days (poster)
Wetenschappendagen, Antwerpen	2009	2 days (poster)
Illumina Users Group Meeting, Crete.	2009	3 days (poster)
European Society of Calcified Tissues, Vienna	2009	4 days (oral)
European Society of Human Genetics, Vienna	2009	4 days (poster)
American Society of Bone and Mineral Research, Denver	2009	5 days (poster)
Genomics of Common Diseases, Hinxton	2009	4 days (plenary)
International Genetic Epidemiology Society Meeting, Honolulu	2009	1 day
American Society of Human Genetics, Honolulu	2009	5 days (oral)
CHARGE Meeting, Washington	2010	2 days (oral)
Wetenschappendagen, Antwerpen	2010	2 days (oral)
American Society of Bone and Mineral Research, Toronto	2010	5 days (oral)
CHARGE sequencing meeting	2010	2 days
American Society of Human Genetics, Washington	2010	4 days (oral)
Dutch Society of Calcium and Bone	2010	2 days (oral)
European Society of Human Genetics, Gothenburg	2010	4 days (plenary)
European Society of Calcified Tissues, Glasgow	2010	4 days (oral-poster)
Next Generation Sequencing and GWAS Symposium, Gothenburg	2010	2 days (plenary)
Wetenschappendagen, Antwerpen	2011	2 days (poster)
CHARGE Meeting, Boston	2011	3 days (oral)
European Society of Calcified Tissues, Athens	2011	4 days (oral)
American Society of Bone and Mineral Research, San Diego	2011	5 days (poster)

Teaching

Course name	Year
5th “SNPs and Human Diseases”, teacher assistant	Nov 2008
6th “SNPs and Human Diseases”, teacher assistant	Nov 2009
Genomics in Molecular Medicine, “Copy Number Variation (CNV) analysis”. “Genetics of Height”	Aug, 2010
7th Course SNPs and Human Diseases	Nov, 2011
Advances in Genome Wide Association Studies of Complex Genetic Disorders, “Literature-based pathway analysis”,	
Genomics in Molecular Medicine, “Genetics of Osteoporosis and Fracture Risk ”, “Copy Number Variation Analysis”	Aug 2011
Advance course of GWAS analysis for PhD Students	Sept 2011
8th Course SNPs and Human Diseases, “Genetic Imputation: Applications to Genome-wide Studies”.	Nov 2011

Other research activities

Name	Year	Workload
Short term visiting scientist at the Prof. John Ioannidis lab in Ioannina, Greece.	2008	2 weeks

About the author

Karol Estrada was born in Mexico City in March 10th, 1979. After obtaining a bachelor degree in Computer Systems Engineering, he followed training in computational genomics as a graduate student at the Monterrey Institute of Technology. His master's thesis was awarded the national prize "Romulo Garza" for the most outstanding research.

In 2005, he founded the Department of Computational Genomics in the National Institute of Genomic Medicine. In 2008 he moved to Rotterdam to work in the Genetic Factors for Osteoporosis project as PhD student.

Between 2008 and 2012, he described 45 novel loci associated with bone mineral density in two separate first-authored *Nature Genetics* papers. He also co-led the meta-analysis of adult height which was published in *Nature*, 2010. He has received multiple international Young Investigator Awards. Karol is married to Lizbeth Herrera. In October 27th, 2011 he received his most important award for genetic research: his son Oscar André.