Genetics of Estrogen-Related Traits;

From Candidate Genes to GWAS

Lisette Stolk

The work described in this thesis was perfromed at the Department of Internal Medicine, Epidemiology, and Clinical Genetics of the Erasmus Medical Center (MC), Rotterdam, the Netherlands. 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 generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). The studies described in this thesis were funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810, and funding from the European Commision (HEALTH-F2-2008-201865 (GEFOS); HEALTH-F2-2008-35627 (TREAT-OA)). The study described in Chapter 7 was performed within the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium.

The publication of this thesis was financially supported by:

Cover design: René Raaphorst
Printed by: Optima Grafische Communicatie, Rotterdam

ISBN: 978-90-8559-588-5

Genetics of Estrogen-Related Traits;

From Candidate Genes to GWAS

Genetica van Oestrogeen-Gerelateerde Eindpunten;

Van Kandidaat Genen tot GWAS

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
dinsdag 1 december 2009 om 15:30 uur

door

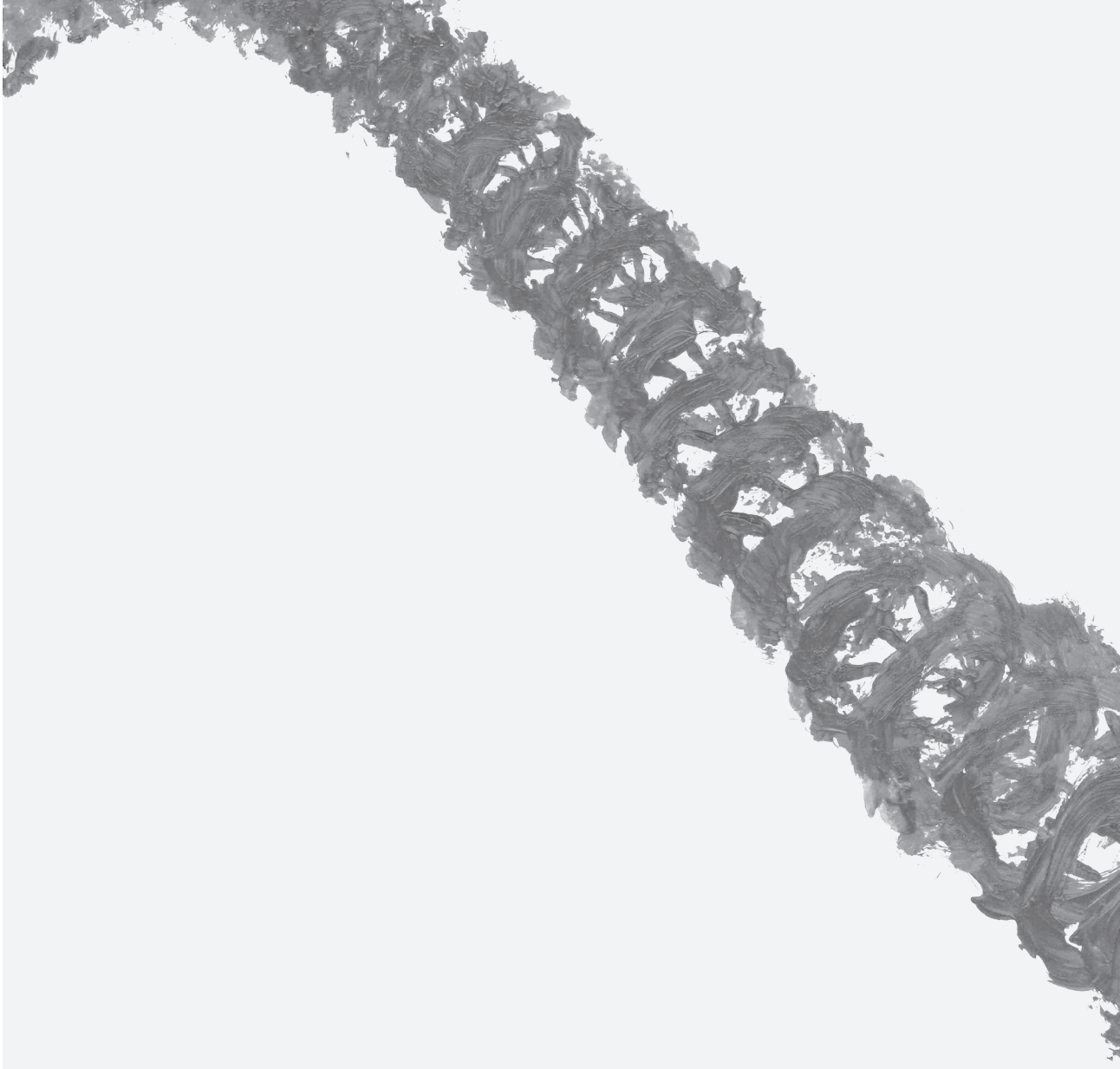
Lisette Stolk

geboren te Dordrecht

Promotiecommissie

Promotor	:	Prof.dr. A.G. Uitterlinden
Overige leden	:	Prof.dr. H.A.P .Pols Prof.dr. T.D. Spector Dr. J.S.E. Laven
Copromotor	:	Dr. J.B.J. van Meurs

CONTENT



Chapter 1

Introduction

Estradiol is one of the reproductive hormones, and has effects not only on reproductive tissue affecting female development and fertility, but also on non-reproductive tissues. This is most clearly visible in women where, after menopause, substantial physiological changes occur. Menopause marks the end of menstrual cycles, and the end of estradiol production by the ovaries. Menopausal symptoms, including hot flashes, insomnia and emotional alterations, are caused by the depletion of estrogens. Postmenopausal women have an increased risk for diseases such as osteoporosis, cardiovascular diseases, stroke and ovarian cancer, which suggests a role for estrogen in the occurrence of the diseases. On the contrary, long estradiol exposure, due to early menarche and/or late menopause, gives rise to a higher risk for breast cancer. In elderly men the risk for these diseases is lower due to higher levels of estrogens compared to post-menopausal women of the same age.

Sex Steroids in Men and Women

In men and premenopausal women estrogens are mainly produced in the gonads under the influence of the Hypothalamus-Pituitary-Gonad (HPG) axis. The hypothalamus secretes gonadotropin releasing hormone (GnRH) which stimulates the pituitary to secrete luteinizing hormone (LH) and follicle stimulating hormone (FSH) (Figure 1). In response to these two gonadotropins the gonads synthesize the sex steroids estradiol (E2) and testosterone (T) from cholesterol. When sex steroid levels

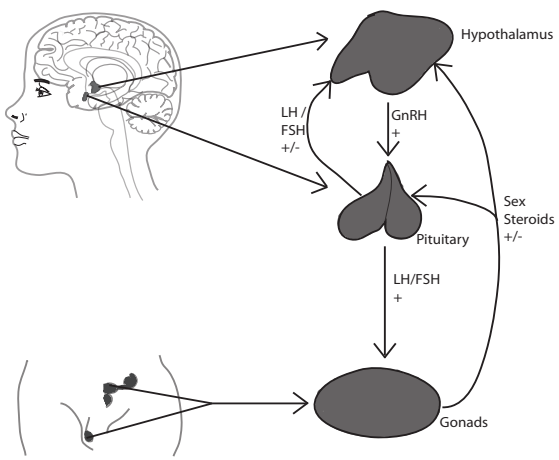


Figure 1. Schematic overview of the Hypothalamus-Pituitary-Gonad (HPG) axis, showing the regulation of sex steroid secretion via a negative feedback loop (GnRH: gonadotrophin releasing hormone, LH: luteinizing hormone, FSH: follicle stimulation hormone).

are high, there is a negative feedback on the hypothalamus and pituitary, while when sex steroid levels are low, there is a positive feedback. In this way sex steroids regulate their own production (Figure 1).

In women menarche, the first menstrual period, marks the start of menstruation and high estrogen levels. The mean age at menarche is 13 years with a normal range between 9 and 17 years, and starts about 2 years after the onset of puberty[1]. Menopause, the end of the menstrual periods, marks the cessation of a women's reproductive lifespan and the decline of estrogen levels. The age-related decrease in ovarian follicle numbers and a decay in the oocyte quality lead to the occurrence of natural loss of fecundity and, ultimately, menopause. It occurs because there are no ovarian follicles to develop anymore, due to successive cycles of ovulation. The rate of this ovarian ageing process is highly variable among women and can range between 40-60 years of age[2]. Because there are no developing follicles anymore, the gonads cannot produce estrogens under the influence of gonadotrophins after menopause.

This lack of female sex steroid hormones results in a loss of negative

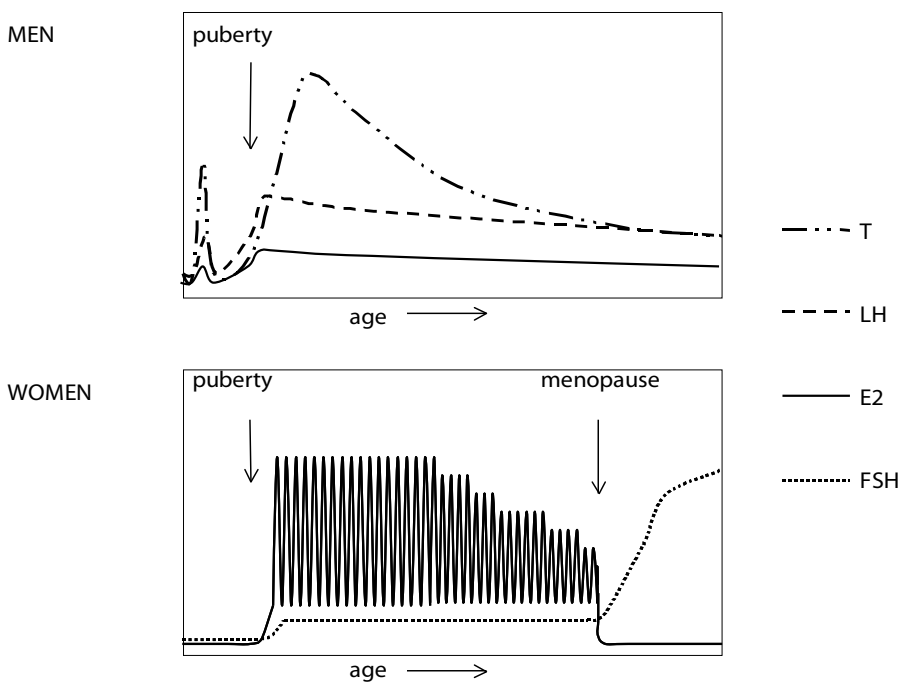


Figure 2. Patterns of hormone levels (estradiol (E2), testosterone (T), luteinizing hormone (LH), and follicle stimulating hormone (FSH)) over life time for men (upper panel) and women (lower panel).

feedback on the hypothalamus and pituitary, and high circulating concentrations of gonadotrophins (Figure 2).

After menopause the only source of female sex hormones is the adrenal androgens converted locally to estrogens. In men androgen and estrogen levels become higher during puberty and during aging the hormone levels get lower, although much more gradually as in women (Figure 2). Over life male testosterone levels are higher and estradiol levels are lower compared to levels in females. However, after menopause female estradiol levels are lower compared to the levels in men.

Sex Steroid Synthesis and Metabolism

All steroid hormones in men and women are derived from cholesterol The complete synthesis and metabolism pathway from cholesterol to estrogens and androgens is shown in Figure 3. In women estrogens and androgens are produced by the ovaries, where the different cell types in the ovaries have all the enzymes to synthesize estrogens. However, the amounts of the enzymes differ and therefore, the predominant hormone per cell type differs. In the theca cells the predominantly synthesized hormones are androstendione (Adion) and testosterone (T), whereas in the granulosa cells the predominantly produced hormones are estrone (E1) and estradiol (E2).

The conversion of cholesterol to pregnenolone is controlled by LH, while FSH controls the conversion of androgens to estrogens. The four key enzymes in the estrogen synthesis pathway are CYP11A1 – converting cholesterol to pregnenolone; HSD3B2 – converting pregnenolone to progesterone, and dehydroepiandrosterone (DHEA) to androstendione (Adion); CYP17A1 – converting pregnenolone to 17-hydroxypregnenolone (17OHpreg); and aromatase (CYP19A1) – converting androgens to estrogens.

After menopause, when the ovaries are non-functional, women are dependent on the local conversion of androgens produced, by the adrenals, to estrogens by aromatase.

In men the adrenal glands and the testis synthesize androgens and in the peripheral target tissue aromatase converts the androgens to estrogens. In the adrenals and the testis androstendione is formed, that is converted to testosterone in the testis. Testosterone is transported to the peripheral tissue via the circulation, were it is converted to estradiol (by aromatase) or dihydrotestosterone (DHT) via 5-alpha-reductase (SRD5A1 and SRD5A2).

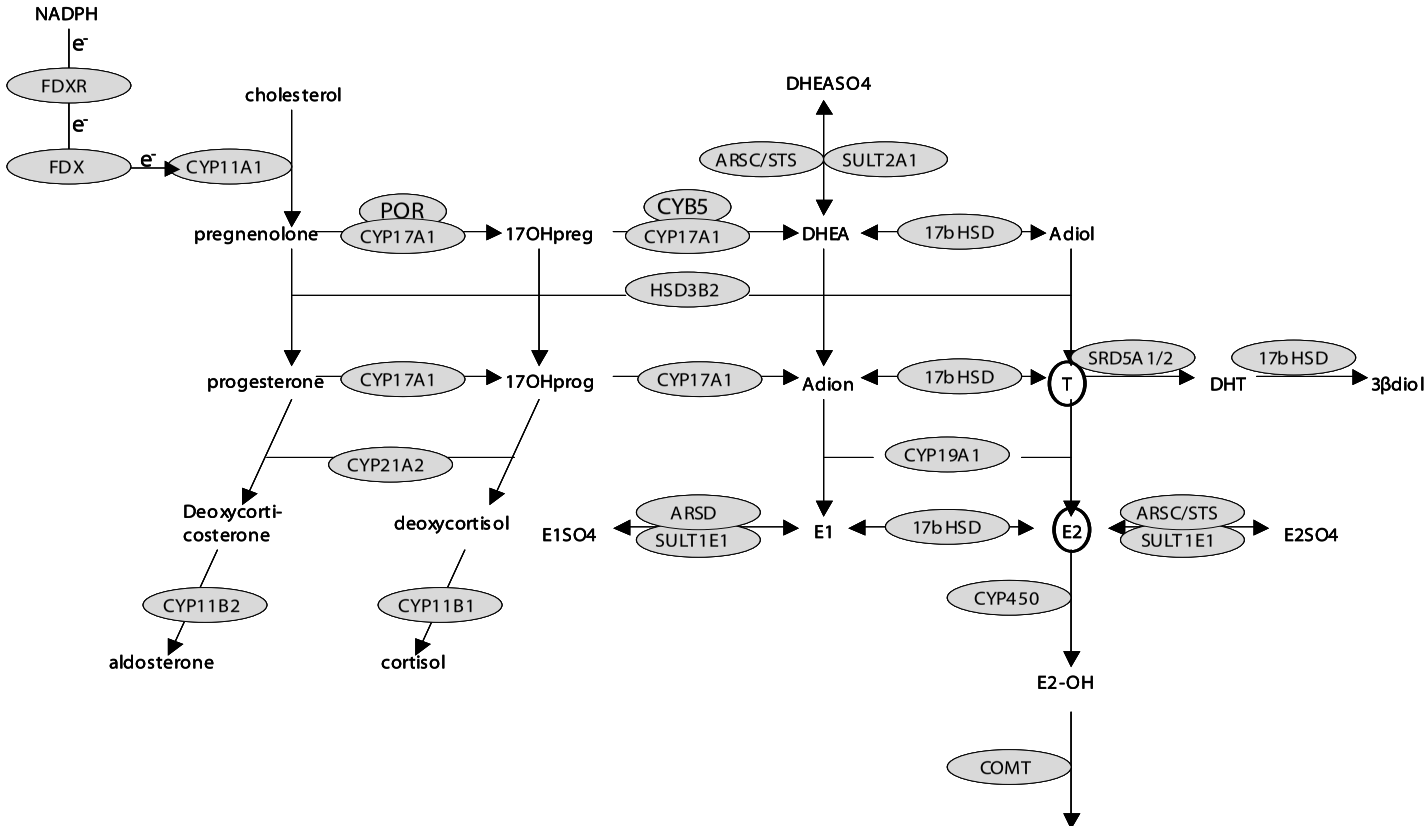


Figure 3. Schematic overview of the synthesis of sex steroids estradiol (E2) and testosterone (T) from cholesterol. In grey all the enzymes involved in this synthesis and metabolism are shown. (17OHpreg: 17-hydroxypregnenolone; 17Ohprog: 17-hydroxyprogesterone; DHEA: dehydroepiandrosterone; DHEASO4: dehydroepiandrosterone sulfate; Adiol: androstenediol; Adion: androstendione; E1: estrone; E1SO4: estrone sulfate E2-OH: hydroxy estradiol; E2SO4: estradiol sulfate; DHT: dihydrotestosterone).

Estrogen signalling

Most of the estrogen produced by the gonads is released into the circulation where almost all of the estrogen binds to sex hormone binding globulin (SHBG). Part of the estrogen diffuses through the cell-membrane and binds the estrogen receptor alpha (ERα; chr 6q25.1) or beta (ERβ; chr 14q23.2), which triggers the receptors to form dimers. After dimerization the receptor-ligand complex diffuses into the cell nucleus where it binds to specific DNA sequences, so called Estrogen Response Elements (EREs), which are located in the promoter region of estrogen target genes. Binding of the receptor-ligand complex to EREs leads to changes in gene expression of estrogen target genes.

Besides these genomic effects estradiol also exerts more rapid actions, which cannot be explained by the nuclear or genomic function of the hormone. These non-genomic effects are thought to be mediated by a membrane bound estrogen receptor. Membrane bound ERα was thought to be this non-genomic receptor[3,4], but more recently the G-protein coupled receptor 30 (GPR30) also known as G-protein coupled estrogen receptor (GPER; chr 7p22.3) was identified as a membrane-bound estrogen receptor[5,6]. In an estrogen receptor negative breast cancer cell line GPER was shown to mediate estrogen-promoted proliferation *in vivo*[7,8].

The function of several organs and systems is affected when one of these receptors is deleted in mice or human as is shown in Table 1. In 1994 a human male with a disruptive mutation of the ERα gene leading to estrogen resistance was reported[22]. This man showed continuous growth due to failure of closing of the growth plate, but also osteoporosis. Furthermore, he had endothelial dysfunction and early atherosclerotic coronary artery disease[23,24]. These results illustrate the pleiotropic effects of estradiol, and the differences and similarities in effect between the two genders.

Genetic variation

The human genome consists of four nucleotides, Adenine (A), Thymine (T), Guanine (G) and Cytosine (C), which form basepairs; Adenine with Thymine, and Cytosine with Guanine. The total length of the human DNA is approximately 3,300,000,000 basepairs (bp). Approximately 1% of the genome encodes around 24,000 genes transcribed in human cells. The DNA-sequence is not the same in all individuals, on average 1 in every 300bp is polymorph, i.e., 1 in 300bp can differ between individuals in the population. When comparing two random individuals 1 in 1,000bp is

Table 1. Summary of phenotypes of ERα, ERβ and GPER knockout mice.

	ERα	ERβ	GPER
Reproductive tract			
Male	Normal development[9,10], infertility[11]	Not affected[12]	NS
Female	Normal development[11], infertility[11]	Fertile, but small litter size[12]	NS
Hormone levels			
Male	Not affected[13]	Not affected[13]	NS
Female	Elevated E2, T and LH levels[13]	Not affected[13]	NS
Bone			
Male	Lower femoral length, diameter and BMD[14,15]	Not affected[16]	NS
Female	Lower femoral length, diameter and BMD[14,15]	Slightly higher BMD[16]	No effect of E2 on growth plate[17]
Cardiovascular			
Male	Protective effect on stroke abolished[18]	Not affected[18]	NS
Female	Protective effect on stroke abolished[18]	Not affected[18]	Increased blood pressure[19]
Glucose metabolism			
Male	Impaired glucose tolerance and insulin resistance[20,21]	NS	NS
Female	Impaired glucose tolerance and insulin resistance[20,21]	NS	Hyperglycemia and impaired glucose tolerance[19]

ERα: Estrogen Receptor Alpha; ERβ: Estrogen Receptor Beta; GPER: G-protein Coupled Estrogen Receptor; BMD: Bone Mineral Density; NS: not studied.

different, so that such individuals on average are different for over 3 million positions in their genomes. The most common variants are single nucleotide polymorphisms (SNPs), which are sites in the genome where only one bp of the chromosome differs. For example, a subset of all the chromosomes in the population has an A at a certain location of the DNA, whereas the other chromosomes in the population have a G at this location. This pair of A and G are called the alleles of a SNP.

DNA is transcribed into messenger RNA (mRNA) which codes for the amino acids that build up a protein. If a change in the nucleotide sequence results in a change in mRNA sequence it could change the amino acid incorporated in the protein and in this way it could change the function of the protein, which is illustrated in Figure 4. This change in function of the protein could contribute to the risk of getting a certain disease. This is shown in the lower part of figure 4, when people with the disease (cases) carry a genetic variant more frequently compared to people that do not have the disease (controls), i.e., the genetic variation is associated with the disease.

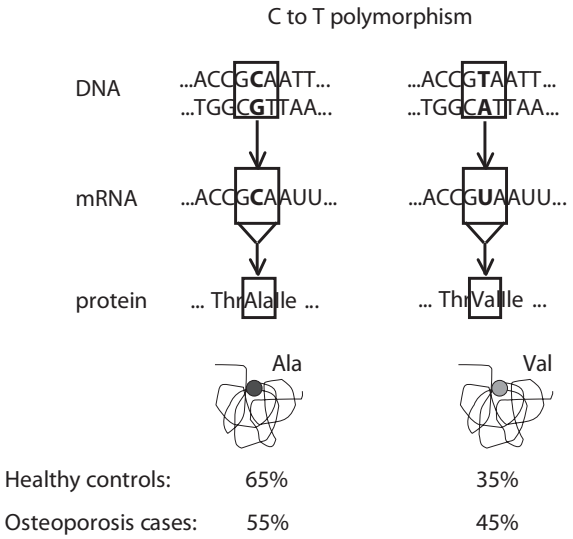


Figure 4. Schematic overview of how a SNP could influence a protein and disease risk. In the upper part the DNA is shown, with two alleles, a C-allele and a T-allele. This is transcribed to the mRNA as a C or an U (uracil). In the protein the GCA is translated into an Alanine amino acid, and the GUA is translated into a Valine amino acid. The lower part shows an association with osteoporosis where cases have a higher frequency of the Valine-allele or T-allele compared to healthy controls.

Studying complex traits

Estrogen-related traits like sex hormone levels, and estrogen-related diseases such as osteoporosis, and cardiovascular disease are so-called complex traits or diseases. Complex traits are traits of which the etiology is explained by the effects of multiple genes and multiple environmental factors, while also interactions between genes itself and genes and the environment play a role. To study polymorphisms modifying these traits or diseases several approaches can be used, including linkage and association analysis. Linkage analysis is used to identify linkage between approximately 500 highly polymorphic markers (microsatellites) throughout the genome and a disease. The stretches of DNA examined with these ~500 markers are long, so the effects of the genetic variants modifying the disease risk have to be relatively large. For such a genome search sibling pairs, a set of family pedigrees, or an isolated population are studied since areas of linkage in the genome are much larger in genetically more closely related individuals than in an out-bred population. This makes it easier to identify regions in the genome that are associated with the phenotype of

interest. Once linkage studies have identified a region of potential linkage, the next step is fine-mapping of the region, because hundreds of genes may lie within the area identified by linkage analysis. For complex traits, however, the effects of the genetic variants contributing to the disease are thought to be very small and the number of genetic variants involved much bigger. Therefore, linkage analysis is unable to identify genetic variants underlying common traits.

For complex traits and diseases a more powerful approach, association analysis, is used, where one seeks to find an association of a disease with a genetic variant which is the causal variant or at least highly correlated to this variant. This correlation between two or more genetic variants is measured in linkage disequilibrium (LD). LD can be depicted with two variables, the first is D', which is a measure for the amount of recombination between two variants. The closer D' is to 1 the lower the amount of recombination between the two variants, D' can range between 0 (no LD) and 1 (complete LD). A second measure of LD is r^2, which is the correlation between two variants, or how well they can predict each other. An example of how this correlation can be used in genetic studies is shown in Figure 5. In this figure alleles at SNP1 and 2 are always present together in the population and are highly correlated (they can predict each other). The same holds for SNP3 and 5, and SNP4 and 6., therefore, only SNP1, 3, and 6 need to be genotyped to obtain all information about the 6 SNPs in a population.

To study associations between a genetic variant and a disease or trait the most widely used approach up to a few years ago was to examine one or more candidate genes based on biological knowledge about traits and diseases. Given the pivotal role of estrogen receptor alpha (*ESR1*) in estrogen signalling this gene is a logical

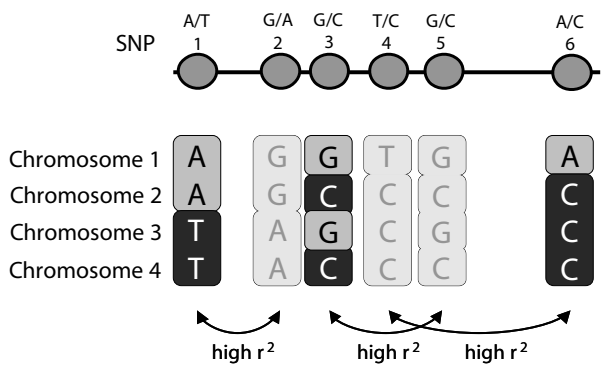


Figure 5. Correlation of six SNPs. SNP1 and SNP2 have the A and G nucleotide at chromosome 1 and 2, and T and A nucleotide at chromosome 3 and 4. Therefore r^2 is high, and they can predict each other. The same holds for SNP3 and 5 and SNP4 and 6 (adapted from Carlson, et al. [25]).

candidate to study for estrogen-related traits and diseases. Within the Rotterdam Study two polymorphisms located in the first intron of the gene were found associated with many disease endpoints[26-36], as is shown in Figure 6. These results are strikingly similar compared to the pleiotropy of phenotypes seen in estrogen receptor knockout mice and man. Recently a new technique to study complex traits in a hypothesis-free manner has become available, the Genome-Wide Association Study (GWAS).

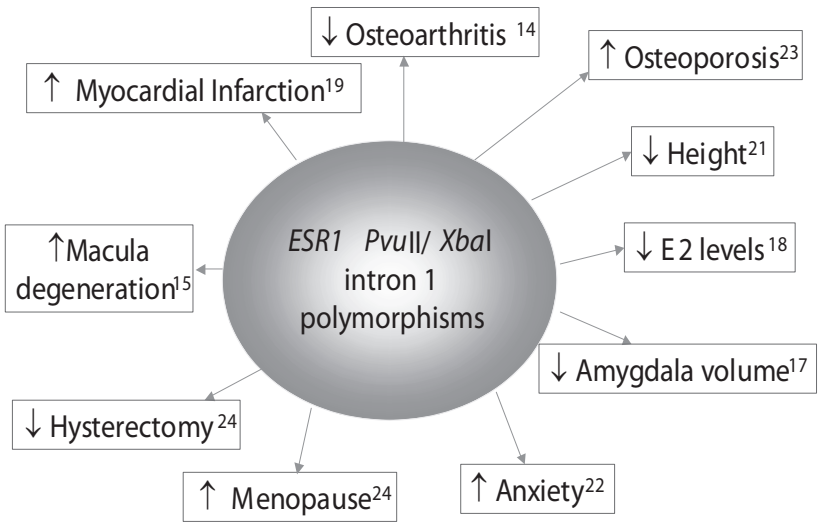


Figure 6. Scheme showing the pleiotropic actions of the Estrogen Receptor alpha (*ESR1*) *Pvu*II and *Xba*I SNPs.

Genome-Wide Association Study (GWAS)

GWAS consists of screening the genome of many hundreds to thousands of subjects in a case-control study or population based cohort study, with >500,000 SNPs and subsequentely perform an association analysis between phenotype and all the genetic markers. Such a GWAS then identifies genetic markers associated to the phenotype of interest with a certain statistical significance. The availability of GWAS is possible because of the knowledge on the common variation present in human (the International HapMap Project[37]) and the development of new techniques making it possible to genotype > 500,000 SNPs in a single DNA sample at once.

Because of the multiplicity of testing so many markers certain thresholds have been considered to declare an association "genome wide significant" (gws). The commonly used gws threshold at this moment is $5x10^{-8}$, which is the nominal p-value

of 0.05 divided by the number of independent common genetic variants (estimated to be ~1 million). Naturally, the chance of seeing a gws association depends on factors such as the effect size of the markers and the phenotype and the size of the study sample. Usually, a typical GWAS consists of a discovery sample with GWA data and a replication sample with GWA data. From the discovery GWAS the "top-hits" (for example, all genetic markers that reach a significance of $1x10^{-5}$) will be analysed in subsequent replication cohorts which do not necessarily have GWA data, but which can be genotyped for the particular genetic markers identified. This approach was shown to be successful for several diseases and traits, amongst others, age-related macular degeneration[38], breast cancer[39], diabetes mellitus type II[40,41], and height[42-46].

The Rotterdam Study

For the studies described in this thesis we used genotype and phenotype information of subjects of the Rotterdam Study I and II (RSI and RSII), a large prospective population-based cohort study of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the elderly. Rationale and design have been described previously[47]. For the entire cohort, information on vital status is obtained continuously from the municipal authorities in Rotterdam. For subjects who moved outside the research area, mortality data are obtained from general practitioners (GPs). GPs in the research area (covering 80% of the cohort) reported all relevant fatal and non-fatal events, such as fractures and myocardial infarctions, through a computerized system. Research physicians verified follow-up information by checking GPs' patients' records. This is possible because in The Netherlands the GP has a gate-keeper function, which means that the only way to access specialist and hospital care is by consulting a GP. The GP retains all medical information of his patients. For the remaining 20% of the cohort, research physicians regularly visited the GPs and collected data from their records. All events were coded independently by two research physicians according to the International Classification of Diseases, 10th revision (ICD-10)[48]. If there was disagreement, consensus was reached in a separate session. A medical expert in the field reviewed all coded events for final classification.

Aim and description of chapters / outline of this thesis

The aim of this thesis is 1) to identify genetic variations in genes of the estrogen pathway that are associated with height, E2-levels, bone mineral density (BMD), and risk for vertebral fractures, and 2) to identify new genes and pathways involved in inter-individual variations in hormone levels, menarche and menopause. In part A, three candidate genes involved in estrogen signalling and estrogen degradation are studied using a candidate gene approach for osteoporosis, height, and estradiol levels. In part B, 61 genes in the estrogen pathway are studied for the association with hormone levels, age at natural menopause, lumbar spine BMD and vertebral fracture risk. Part C describes three hypothesis-free GWAS for hormone levels, age at menarche, and age at natural menopause.

REFERENCES

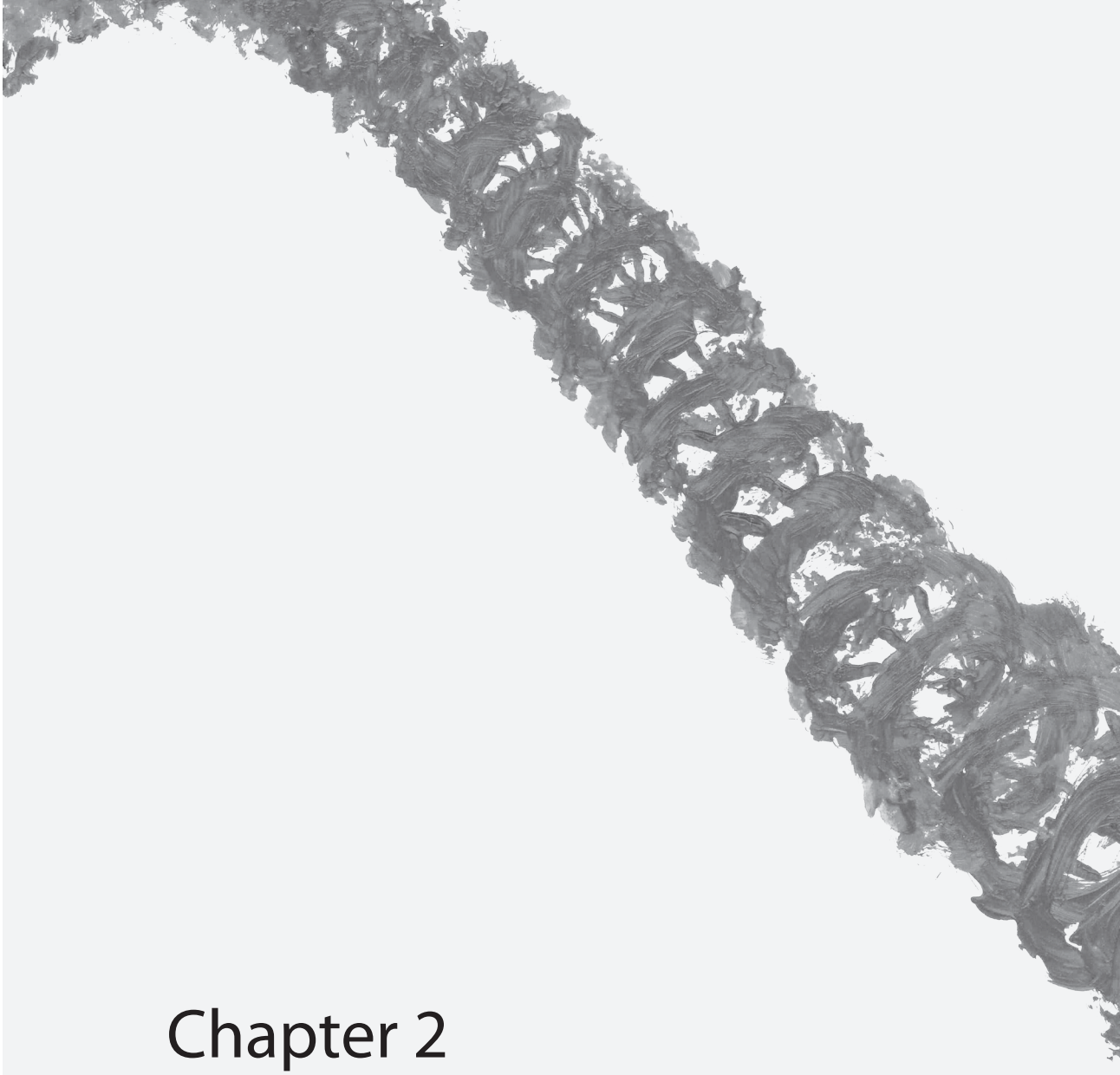
1. Marshall, W.A. & Tanner, J.M. Variations in pattern of pubertal changes in girls. Arch Dis Child 44, 291-303 (1969).
2. te Velde, E.R., Dorland, M. & Broekmans, F.J. Age at menopause as a marker of reproductive ageing. Maturitas 30, 119-25 (1998).
3. Evinger, A.J., 3rd & Levin, E.R. Requirements for estrogen receptor alpha membrane localization and function. Steroids 70, 361-3 (2005).
4. Razandi, M., Pedram, A., Merchenthaler, I., Greene, G.L. & Levin, E.R. Plasma membrane estrogen receptors exist and functions as dimers. Mol Endocrinol 18, 2854-65 (2004).
5. Revankar, C.M., Cimino, D.F., Sklar, L.A., Arterburn, J.B. & Prossnitz, E.R. A transmembrane intracellular estrogen receptor mediates rapid cell signaling. Science 307, 1625-30 (2005).
6. Thomas, P., Pang, Y., Filardo, E.J. & Dong, J. Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. Endocrinology 146, 624-32 (2005).
7. Filardo, E.J., Quinn, J.A., Frackelton, A.R., Jr. & Bland, K.I. Estrogen action via the G protein-coupled receptor, GPR30: stimulation of adenylyl cyclase and cAMP-mediated attenuation of the epidermal growth factor receptor-to-MAPK signaling axis. Mol Endocrinol 16, 70-84 (2002).
8. Maggiolini, M. et al. The G protein-coupled receptor GPR30 mediates c-fos up-regulation by 17beta-estradiol and phytoestrogens in breast cancer cells. J Biol Chem 279, 27008-16 (2004).
9. Donaldson, K.M. et al. Morphometric study of the gubernaculum in male estrogen receptor mutant mice. J Androl 17, 91-5 (1996).
10. Eddy, E.M. et al. Targeted disruption of the estrogen receptor gene in male mice causes alteration of spermatogenesis and infertility. Endocrinology 137, 4796-805 (1996).
11. Lubahn, D.B. et al. Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. Proc Natl Acad Sci U S A 90, 11162-6 (1993).
12. Krege, J.H. et al. Generation and reproductive phenotypes of mice lacking estrogen receptor beta. Proc Natl Acad Sci U S A 95, 15677-82 (1998).
13. Couse, J.F. & Korach, K.S. Estrogen receptor null mice: what have we learned and where will they lead us? Endocr Rev 20, 358-417 (1999).
14. Korach, K.S. Insights from the study of animals lacking functional estrogen receptor. Science 266, 1524-7 (1994).
15. Korach, K.S., Taki, M. & Kimbro, K.S. The effects of estrogen receptor gene disruption on bone. in Women's Health and Menopause (ed. R, P.) 69-73 (Kluwer Academic Publishers and Fondazione Giovanni Lorenzini, Amsterdam, The Netherlands, 1997).
16. Windahl, S.H., Vidal, O., Andersson, G., Gustafsson, J.A. & Ohlsson, C. Increased cortical bone mineral content but unchanged trabecular bone mineral density in female ERbeta(-/-) mice. J Clin Invest 104, 895-901 (1999).
17. Windahl, S.H. et al. The role of the G protein-coupled receptor GPR30 in the effects of estrogen in ovariectomized mice. Am J Physiol Endocrinol Metab 296, E490-6 (2009).
18. Dubal, D.B. et al. Estrogen receptor alpha, not beta, is a critical link in estradiol-mediated protection against brain injury. Proc Natl Acad Sci U S A 98, 1952-7 (2001).
19. Martensson, U.E. et al. Deletion of the G protein-coupled receptor 30 impairs glucose tolerance, reduces bone growth, increases blood pressure, and eliminates estradiol-stimulated insulin release in female mice. Endocrinology 150, 687-98 (2009).
20. Heine, P.A., Taylor, J.A., Iwamoto, G.A., Lubahn, D.B. & Cooke, P.S. Increased adipose

tissue in male and female estrogen receptor-alpha knockout mice. Proc Natl Acad Sci U S A 97, 12729-34 (2000).
21. Taylor, J.A. & Lubahn, D.B. Impaired glucose tolerance in the ER KO mous. 80th Annual Meeting of The Endocrine Society, 257 (1998).
22. Smith, E.P. et al. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 331, 1056-61 (1994).
23. Sudhir, K. et al. Premature coronary artery disease associated with a disruptive mutation in the estrogen receptor gene in a man. Circulation 96, 3774-7 (1997).
24. Sudhir, K. et al. Endothelial dysfunction in a man with disruptive mutation in oestrogen-receptor gene. Lancet 349, 1146-7 (1997).
25. Carlson, C.S. et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74, 106-20 (2004).
26. Bergink, A.P. et al. Estrogen receptor alpha gene haplotype is associated with radiographic osteoarthritis of the knee in elderly men and women. Arthritis Rheum 48, 1913-22 (2003).
27. Boekhoorn, S.S. et al. Estrogen receptor alpha gene polymorphisms associated with incident aging macula disorder. Invest Ophthalmol Vis Sci 48, 1012-7 (2007).
28. Bos, M.J. et al. Variation in the estrogen receptor alpha gene and risk of stroke: the Rotterdam Study. Stroke 39, 1324-6 (2008).
29. den Heijer, T. et al. Variations in estrogen receptor alpha gene and risk of dementia, and brain volumes on MRI. Mol Psychiatry 9, 1129-35 (2004).
30. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. Eur J Endocrinol 153, 327-34 (2005).
31. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms and risk of myocardial infarction. Jama 291, 2969-77 (2004).
32. 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).
33. Schuit, S.C. et al. Height in pre- and postmenopausal women is influenced by estrogen receptor alpha gene polymorphisms. J Clin Endocrinol Metab 89, 303-9 (2004).
34. Tiemeier, H. et al. Estrogen receptor alpha gene polymorphisms and anxiety disorder in an elderly population. Mol Psychiatry 10, 806-7 (2005).
35. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. Hum Mol Genet 12, 1745-54 (2003).
36. Weel, A.E. et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. J Clin Endocrinol Metab 84, 3146-50 (1999).
37. The International HapMap, C. The International HapMap Project. Nature 426, 789-96 (2003).
38. Klein, R.J. et al. Complement factor H polymorphism in age-related macular degeneration. Science 308, 385-9 (2005).
39. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087-93 (2007).
40. Diabetes Genetics Initiative of Broad Institute of Harvard and Mit, L.U.a.N.I.o.B.R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-6 (2007).
41. Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341-5 (2007).
42. Estrada, K. et al. A genome-wide association study of northwestern Europeans involves the CNP signaling pathway in the etiology of human height variation. Human

Molecular Genetics 18, 3516-24 (2009).
43. Gudbjartsson, D.F. et al. Many sequence variants affecting diversity of adult human height. Nat Genet 40, 609-15 (2008).
44. Lettre, G. et al. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet 40, 584-91 (2008).
45. Soranzo, N. et al. 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).
46. Weedon, M.N. et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40, 575-83 (2008).
47. Hofman, A. et al. The Rotterdam Study: objectives and design update. Eur J Epidemiol 22, 819-29 (2007).
48. International Statistical Classification of Diseases and Related Health Problems. Geneva, World Health Organization (1992).

Part A

Candidate Gene Studies



Chapter 2

Fine-mapping of the Estrogen Receptor alpha *Pvu*II-*Xba*I linkage block and the effect of larger samples size

Lisette Stolk, Joyce BJ van Meurs, Fernando Rivadeneira, Carola Zillikens, Huibert AP Pols, Albert Hofman, Frank H de Jong, André G Uitterlinden

Manuscript in preparation

ABSTRACT

The estrogen receptor (*ESR1*) *Pvu*II and *Xba*I SNPs are associated with bone and hormone endpoints, however the linkage region of these SNPs and their effect on function of the gene remains unknown. The aim of this study was to fine-map the association of *Pvu*II-*Xba*I with bone mineral density, fracture risk, estradiol (E2)-levels, and menopausal age in women of the Rotterdam Study I – Bone Cohort (RSI-BC, n=1,070) and the complete Rotterdam Study I (RSI, n=3,547). For this purpose we genotyped five SNPs in the *Pvu*II-*Xba*I linkage-block using Taqman Allelic Discrimination and used genotype data for 9 SNPs from the Illumina HumanHap 550K beadchip in 3,547 postmenopausal women. We found *Pvu*II and *Xba*I associated with lumbar spine BMD (beta=0.02 g/cm^2, SE=0.007; P<0.02), vertebral fracture risk (OR=0.45, 95%CI[0.31-0.65]; P<6x10^{-5}), estradiol-levels (beta=2.3 pmol/L, SE=1.1; P=0.03) and age at natural menopause (beta= -0.6 yr, SE=0.25; P<0.03) in RSI-BC. When the sample size was tripled (RSI) these associations were only significant for E2-levels (beta=2 pmol/L, SE=0.8; P<0.009) and vertebral fracture risk (OR=0.75, 95%CI[0.60-0.95]; P<0.02). Furthermore, we found similar results for five SNPs in the 3′ end of the LD-block, in moderate correlation (r^2≥0.45) with *Pvu*II and *Xba*I, and two SNPs in the 5′ end of the LD-block with slightly lower correlation (r^2≥0.24). Effect sizes for these seven SNPs were similar to those for *Pvu*II and *Xba*I. Haplotype analysis showed that the rs488133 SNP in the 5′ end of the LD-block could be the functional variant, or one highly correlated with that SNP. In conclusion, fine-mapping of the *Pvu*II-*Xba*I LD-block showed that the causal variant for the associations with E2-levels and vertebral fracture risk in RSI is located in the promoter region of the gene.

INTRODUCTION

The estrogen receptor alpha (*ESR1*) gene is a well studied gene for associations of DNA variations with bone mineral density (BMD), fracture risk, estradiol (E2) levels, and age at natural menopause[1-3]. In these analyses the *Pvu*II (rs2234693) and *Xba*I (rs9340799) polymorphisms, located in the first intron, were studied. In some studies these SNPs were examined together with a (TA)n VNTR approximately 1kb upstream of exon 1[2]. A meta-analysis on 18,000 women did not confirm the previously reported association with BMD, but the meta-analysis of 5,834 women showed a protective effect of the *Xba*I SNP on fracture risk independent of BMD[4]. Studies examining functionality found some indication for differences in enhancer activity for the *Xba*I SNP[5] and binding of B-myb for the *Pvu*II SNP[6]. Still, the impact of these polymorphisms on the transcription or activity of the *ESR1* gene has not been established. In fact, polymorphisms linked to *Pvu*II and *Xba*I could be the functional variants underlying the reported associations.

In this study we examined additional common variants in the genomic area to fine-map the association of *Pvu*II-*Xba*I with BMD, fracture risk, E2-levels, and menopausal age in women of the Rotterdam Study I. For this purpose we genotyped 14 SNPs in the *Pvu*II-*Xba*I linkage-block in 3,547 postmenopausal women.

MATERIALS and METHODS

Study population

Subjects were participants of the Rotterdam Study I (RSI), a large prospective population-based cohort study of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the elderly. Rationale and design have been described previously[7]. All 10,275 inhabitants aged 55 years and over were invited for baseline examination between August 1990 and June 1993. Of those, 7,983 participated. Among the subjects living independently, the overall response rate was 78 percent for home interview and 71 percent for examination in the research centre, where anthropometric characteristics and bone mineral density (BMD) were measured, and blood samples were taken. The Rotterdam Study was approved by the medical ethics committee of the Erasmus University Medical School, and written informed consent was obtained from each subject. In the current study

the complete cohort of 3,547 postmenopausal women for which genome wide SNP genotyping data was available was used. In addition, we also examined a subset of RSI, which was used before in several association studies[1-3]. This subset, the Rotterdam Study I – Bone Cohort (RSI-BC, n=1,070 women) was described previously[2]. Selection of subjects in this subset was based on the following selection criteria: age < 80 years, no use of a walking aid, no known diabetes mellitus and no use of diuretics, estrogen, thyroid hormone or cytostatic drug therapy.

Clinical examination

Height and weight were measured at baseline examination with the subject in a standing position with indoor clothing without shoes. BMD (in grams per square centimeter) was determined by dual-energy x-ray absorptiometry (DEXA, DPX-L densitometer, Lunar, Madison, WI) at the lumbar spine (LS) (vertebrae L2, L3, L4), as described elsewhere[8]. Age at menopause, defined as being menopausal for 12 continuous months of amenorrhea, and smoking habits were assessed by a questionnaire, as described previously[3].

Hormone assays

E2-levels were measured in plasma of a subgroup of 709 females, as described earlier[1]. The minimal detection limit for the E2-assay was 4.8 pmol/L, therefore, all non-detectable levels were set to 2.4 pmol/L.

Assessment of Vertebral Fractures

Both at baseline and at follow-up visits, between 1997 and 2001, thoracolumbar radiographs of the spine were obtained. The follow-up radiographs were available for 3,469 individuals (61% of the genotyped population, 1,971 women) who survived an average of 6.4 (SD 0.4) years after the baseline centre visit and who were still able to come to our research centre. All follow-up radiographs were scored for the presence of vertebral fracture by the McCloskey/Kanis method as described earlier[9]. If a vertebral fracture was detected, the baseline radiograph was evaluated as well. If the vertebral fracture was already present at baseline, it was considered a prevalent fracture. If it was not present at baseline, the fracture was defined to be incident.

Genotyping

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. 1-2 ng genomic DNA was dispensed into 384-wells

plates using a Caliper Sciclone ALH3000 pipetting robot (Caliper LS, Mountain View, CA, USA). Six tagging SNPs in the *Pvu*II-*Xba*I LD block, spanning 43kb were selected using the tagger option in Haploview v3.1, from HapMap release #21. Genotypes were determined using Taqman allelic discrimination Assay-on-Demand (Applied Biosystems Inc., Foster City, CA, USA). The PCR reaction mixture included 1-2 ng of genomic DNA in a 2 μl volume and the following reagents: FAM and VIC probes (200 nM), primers (0.9 μM), 2x Taqman PCR master mix (ABgene, Epsom, UK). Reagents were dispensed in the 384-well plates using the Deerac Equator NS808 (Deerac Fluidics, Dublin, Ireland). PCR cycling reactions were performed in 384 wells PCR plates in an ABI 9700 PCR system (Applied Biosystems Inc., Foster City, CA, USA) and consisted of initial denaturation for 15 minutes at 95°C, and 40 cycles with denaturation of 15 seconds at 95°C and annealing and extension for 60 seconds at 60°C. Results were analysed by the ABI Taqman 7900HT using the sequence detection system 2.22 software (Applied Biosystems Inc., Foster City, CA, USA). To confirm the accuracy of genotyping results, 332 (5%) randomly selected samples were re-genotyped with the same method. No inconsistencies were observed.

Additionaly, all DNA samples were genotyped in the framework of a GWAS database using the Illumina HumanHap 550K bead array according to the manufacturer´s protocol, and quality control was performed as described previously[10]. This array contains nine additional SNPs in the *ESR1* LD-block of interest.

Statistical analysis

All analyses were performed using PLINK v1.05[11] with a minor allele frequency (MAF) cut off of 5%, and a Hardy-Weinberg equilibrium (HWE) p-value cut off of 0.0001. Association analyses of lumbar spine BMD, E2-levels, and age at natural menopause were performed using (adjusted) linear regression. For vertebral fracture risk the Odds Ratio (OR) was calculated using adjusted logistic regression. The association of haplotypes of all 15 SNPs with vertebral fracture risk in RSI women was calculated using a simple χ^2-test. The haplotype association with E2-levels was calculated using linear regression. Associations were considered significant if the p-value was below 0.05.

RESULTS

Study-groups and genotyping

In this study only women with genome-wide genotyping data were used, so

in total we studied 1,070 women of the original bone cohort subset (RSI-BC) and 3,547 women from the total RSI cohort (including the 1,070 women from RSI-BC). Baseline characteristics for the two study-groups are shown in Table 1. Overall the women in the RSI-BC subset are younger (-1.6 years, P<0.001), lighter (-1.3kg, P=0.02), and they had a slightly lower lsBMD (-0.03 g/cm^2, P=0.006).

Table 1. Characteristics per study group.

	RSI-BC	RSI	P
N	1,070	3,547	
Age (yrs)	68.7 (8.1)	70.3 (9.6)	< 0.001
Height (cm)	161.6 (6.5)	161.3 (6.6)	NS
Weight (kg)	68.3 (10.2)	69.6 (11.3)	0.001
lsBMD (g/cm^2)	1.01 (0.17)	1.04 (0.18)	0.001
N_vert fx/total	82/578 (14%)	192/1626 (12%)	NS
E2-levels (pmol/L)	15.6 (14.1)	16.2 (14.8)	NS
ANM40-60 (yrs)	49.7 (3.8)	49.9 (3.9)	NS

Data are mean (SD); RSI-BC: Rotterdam Study I – Bone Cohort; RSI: Rotterdam Study I; lsBMD: lumbar spine Bone Mineral Density; N_vert fx: Number of vertebral fractures; E2: estradiol; ANM40-60: Age at Natural Menopause between age 40 and 60; NS: non-significant.

Figure 1 shows a schematic overview of the *ESR1* gene with the LD-blocks based on HapMap phase II release 22. In total 14 SNPs in the LD-block were studied (*Pvu*II, *Xba*I, nine SNPs from the Illumina Human Hap 550K beadarray, and three additional tagging SNPs genotyped using Taqman allelic discrimination). The 14 SNPs and their correlation (r^2) are shown in the lower part of the figure, with their relative locations compared to the previously studied *Pvu*II/*Xba*I polymorphisms and the TA(n) VNTR. Minor allele frequencies of all studied SNPs were similar over the two study groups (Table 2).

Association of PvuII – XbaI in the two cohorts

Table 3 shows the results for the *Pvu*II (rs2234693) and *Xba*I (rs9340799) polymorphisms for lumbar spine BMD (lsBMD), vertebral fracture risk, E2-levels and age at natural menopause for the RSI-BC and RSI study groups. In RSI-BC *Pvu*II and *Xba*I are both associated with lsBMD, with *Pvu*II being the most significantly associated SNP with an effect of 0.02 g/cm^2 higher BMD per C-allele (P=1.7x10^{-3}). The effect size for

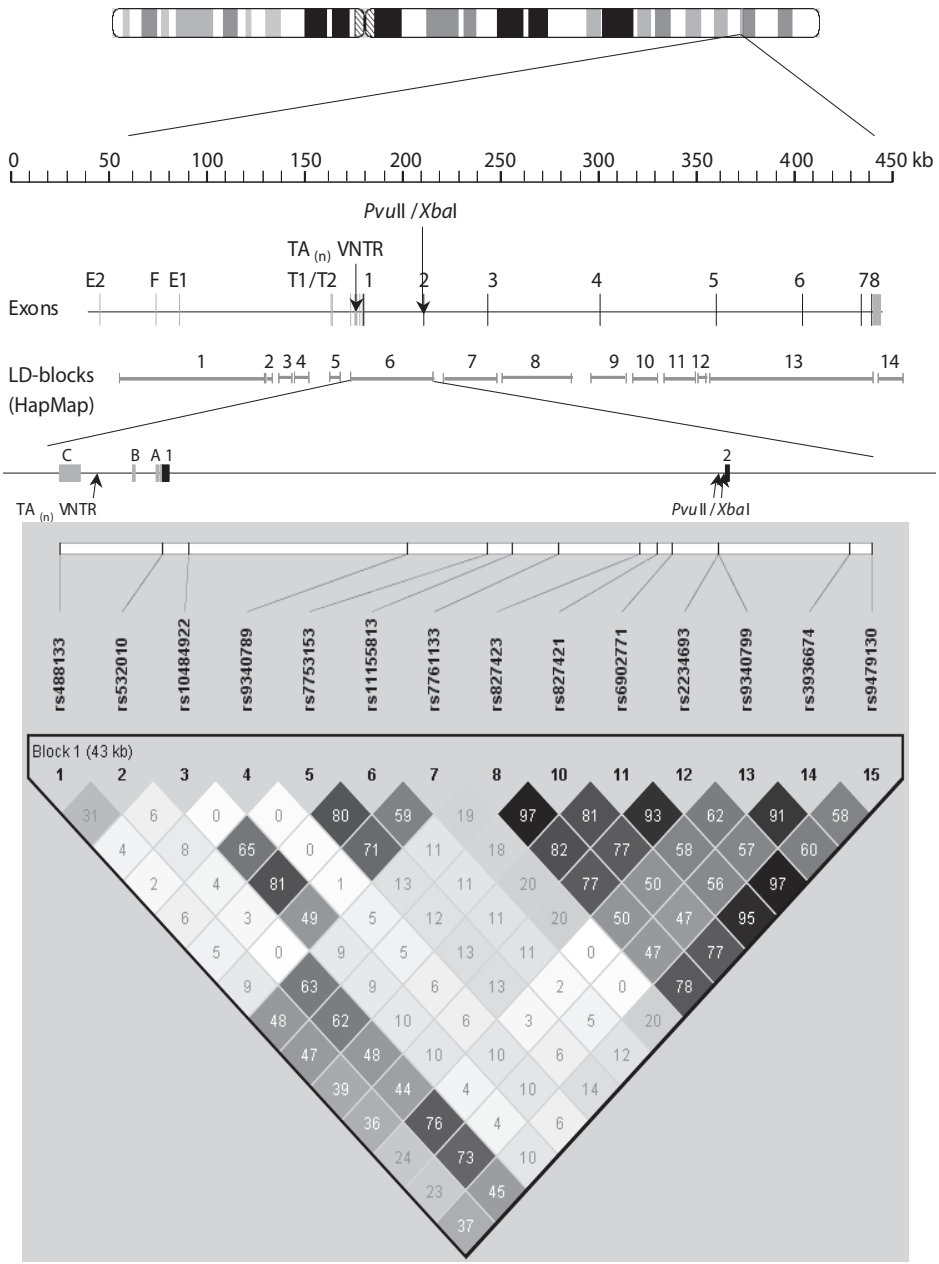


Figure 1. Schematic overview of the *ESR1* gene region located on Chr 6q25.1. Grey vertical lines indicate promoter exons (E2-A), and black vertical lines coding exons (1- 8). The HapMap Phase II data shows the LD-block pattern based on D' for the gene region. The LD-region including the *Pvu*II/*Xba*I SNPs and the $(TA)_n$ VNTR is shown on the middle part of the figure. The lower part of the figure shows the r^2 LD-plot for the women of the RSI cohort.

Table 2. Information on study subgroups and the 14 SNPs in this study.

	RSI-BC	RSI
N	1,070	3,547
SNP	**MAF**	**MAF**
rs488133	T:0.31	T:0.33
rs532010	G:0.41	G:0.39
rs10484922	A:0.09	A:0.09
rs9340789	G:0.05	G:0.05
rs7753153	A:0.12	A:0.12
rs11155813	C:0.11	C:0.10
rs7761133	C:0.16	C:0.16
rs827423	A:0.49	A:0.50
rs827421	A:0.48	A:0.50
rs6902771	T:0.48	T:0.47
rs2234693	C:0.47	C:0.46
rs9340799	G:0.36	G:0.35
rs3936674	A:0.35	A:0.33
rs9479130	C:0.47	C:0.46

RSI-BC: Rotterdam Study I – Bone Cohort; RSI: Rotterdam Study I; MAF: Minor Allele Frequency.

*Xba*I is +0.02 g/cm^2 (P=0.01). In RSI these two polymorphisms are not associated with lsBMD. For vertebral fracture risk these two SNPs are associated in both RSI-BC and RSI.

In RSI-BC the C-allele of *Pvu*II was associated with a 54% lower odds (P=2.2x10^{-5}) and *Xba*I with a 56% lower odds (P=6.4x10^{-5}). While in the complete cohort the ORs for those SNPs are 0.73 for *Pvu*II (P=0.02) and 0.77 for *Xba*I (P=0.01).

*Pvu*II is not associated with E2-levels in RSI-BC (n=300), whereas *Xba*I was associated with an effect size of +2.4 pmol/L per copy of the minor allele (SE=1.16;

Table 3. *Pvu*II and *Xba*I association results for both study groups for lumbar spine BMD, vertebral fracture risk, estradiol-levels and age at natural menopause.

		RSI-BC		RSI	
		***Pvu*II**	***Xba*I**	***Pvu*II**	***Xba*I**
lsBMD	beta (g/cm^2)	0.022 (0.007)	0.019 (0.008)	0.003 (0.005)	0.002 (0.005)
	P	1.7x10^{-3}	0.01	0.52	0.71
vert fx	OR [95% CI]	0.46 [0.32-0.64]	0.44 [0.30-0.65]	0.77 [0.62-0.95]	0.74 [0.59-0.94]
	P	2.2x10^{-5}	6.4x10^{-5}	0.02	0.01
E2-levels	beta (pmol/L)	1.69 (1.11)	2.35 (1.16)	1.98 (0.80)	2.20 (0.82)
	P	0.11	0.03	8.9x10^{-3}	4.6x10^{-3}
ANM40-60	beta (yr)	-0.55 (0.25)	-0.65 (0.26)	-0.23 (0.14)	-0.24 (0.15)
	P	0.03	0.01	0.11	0.11

Values are mean (SD) for lsBMD, E2-levels and ANM40-60; RSI-BC: Rotterdam Study I – Bone Cohort; RSI: Rotterdam Study I; lsBMD: lumbar spine Bone Mineral Density; vert fx: vertebral fracture risk; E2: estradiol; ANM40-60: Age at Natural Menopause between age 40 and 60.

P=0.03). In RSI (n=600) both SNPs are significantly associated with E2-levels, for *Pvu*II the effect was +2.0 pmol/L (SE=0.80; P=0.009) and for *Xba*I the effect was +2.2 pmol/L (SE=0.82; P=0.005). Age at natural menopause was associated with both *Pvu*II (beta=-0.55 yr, SE=0.14; P=0.03) and *Xba*I (beta=-0.65 yr, SE=0.15; P=0.01). In the complete RSI cohort these associations are no longer significant (P=0.11, beta=-0.23 yr, SE= 0.15 for both SNPs).

Fine-mapping of the association signal

In the complete cohort RSI we performed fine-mapping of the associations of *Pvu*II and *Xba*I with vertebral fracture risk and E2-levels. As shown in Figure 2, besides the *Pvu*II and *Xba*I SNPs, rs488133 (OR=1.2, P=0.06) and rs532010 (OR=0.8, P=0.02) are also associated with vertebral fracture risk. Furthermore, the SNPs at the 3' end of the LD-block are significantly associated with vertebral fracture risk with similar ORs (OR=0.8).

Estradiol levels are associated with rs488133 (beta=-1.7 pmol/L, P=0.04) and rs532010 (beta=1.8 pmol/L, P=0.02), but also with three additional SNPs in the 3' end of the LD-block (Figure 2).

The associations for vertebral fracture risk and estradiol levels are found with several SNPs in the LD-block. To determine if they arise from the same signal, we examined the correlation between these SNPs and *Pvu*II and *Xba*I. The correlation of *Pvu*II and *Xba*I with rs488133 is around 24%, and with rs532010 it is approximately 75%, the correlation of these two SNPs itself is 31%. The correlation of the SNPs in the 3' end of the LD-block, where *Pvu*II and *Xba*I are located, is between 47% and 97%. This shows that all associations arise from the same signal, because all have an $r^2>0.2$.

Additionally, we performed haplotype analysis for vertebral fx risk and E2-

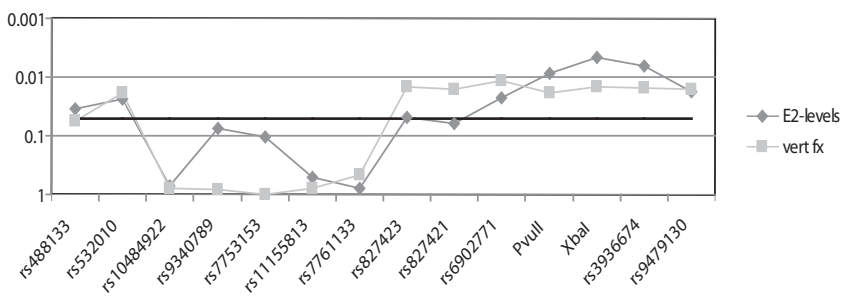


Figure 2. Plot showing the p-values of the 14 SNPs in the *Pvu*II-*Xba*I LD-block for estradiol-levels (E2-levels) and vertebral fracture risk (vert fx). The black horizontal line indicates a p-value of 0.05.

levels in the RSI women, of which the results are shown in Table 4. As can be seen from this table, only haplotype 1 is associated with these endpoints. For vertebral fx risk χ^2=4.83 (P=0.03), with an increased vert fx risk in carriers of this haplotype. For E2-levels this haplotype has an effect of -1.9 pmol/L per copy of the haplotype (P=0.03). The frequency of this haplotype is around 32% in this population.

Table 4. Results from the haplotype analysis for vertebral fracture risk and estradiol levels in RSI women.

			vertebral fx		E2-levels (pmol/L)	
No	**Haplotype**	**Freq**	**χ^2**	**P**	**Beta**	**P**
1	TAGCGTTAACTAGA	0.32	4.83	0.03	-1.91	0.03
2	CGGCGTTGGTCGAC	0.26	3.59	0.09	1.51	0.12
3	CAGCGTTAACTAGA	0.17	0.77	0.38	1.01	0.34
4	CAACACCGGTCAGC	0.10	0.0002	0.99	-0.82	0.57
5	CGGGGTTGGTCGAC	0.05	0.22	0.64	1.90	0.32
6	CGGCGTTGGCTAGA	0.04	0.16	0.69	0.09	0.97
7	CGGCGTCGGTCGAC	0.04	2.20	0.14	0.91	0.65
8	CAGCATCGGTCAGC	0.02	0.02	0.90	-0.42	0.88
9	CGGCACCGGTCGGC	0.01	0.40	0.53	0.36	0.91

Freq: haplotype frequency; χ^2: Chi^2 test statistic; vertebral fx: vertebral fracture risk; E2-levels: estradiol-levels.

DISCUSSION

In this study we analysed 14 tagging SNPs in the *ESR1 Pvu*II-*Xba*I LD-block including *Pvu*II and *Xba*I in both a subset of the RSI cohort (RSI-BC) and in the total RSI cohort, which is three times the size of the RSI-BC. We found two SNPs in the 5' end of the LD-block (rs488133 and rs532010) to be associated with vertebral fracture risk and estradiol levels, together with seven SNPs in the 3' end of the LD-block, where *Pvu*II and *Xba*I are also located. This is consistent with the correlation pattern we see in Figure 1, which shows moderate to high correlation between the two 5' end SNPs and the SNPs in the 3' end of the LD-block.

As previously described the *Pvu*II and *Xba*I polymorphisms were associated with lsBMD, vertebral fracture risk, estradiol levels and age at natural menopause in RSI-BC. However, in the complete cohort (RSI) with tripled sample size, these SNPs were only associated with vertebral fracture risk and estradiol levels, but not with lsBMD and age at natural menopause. For lsBMD this result is in concordance with results from the GENOMOS consortium which showed that both *Pvu*II and *Xba*I are not

associated with lsBMD in a large meta-analysis of 18.000 women[4]. The same study also showed that *Xba*I was associated with vertebral fracture risk, while *Pvu*II was not. In the present study we do find an association of *Pvu*II with vertebral fracture risk in the total female population (n=3,574), although the effect size was attenuated compared to RSI-BC (n=1,070) (RSI-BC: OR=0.5, RSI: OR=0.7). We do not expect very large effects of these polymorphisms. This is based on the common variant common disease hypothesis, where we expect a large number of genetic variants with all small effect sizes and combinations of these variants will cause a (disease) phenotype.

Fine-mapping of the previously found associations revealed an LD-block of 43kb, of which the SNPs in the 3' end are moderately correlated to *Pvu*II and *Xba*I (r^2>0.45). Of these SNPs several are associated with vertebral fracture risk and estradiol levels, due to the high correlation these associations likely represent the same signal. Next to these signals in the 3' end of the LD-block we also found rs488133 and rs532010 at the 5' end of the LD-block to be significantly associated with vertebral fractures and estradiol levels. The correlation between these two SNPs is 31%, and could therefore point to the same association that was seen for *Pvu*II and *Xba*I (r^2≥0.24). The effects seen for rs488133 on estradiol-levels and vertebral fracture risk are consistent, in the sense that estradiol-levels are lower for carriers of the T-allele, and vertebral fracture risk is higher in carriers of the T-allele. The effect of genetic variation in the estrogen receptor alpha gene is more direct on E2-levels than on the other phenotypes, for which E2-levels could be an intermediate. The absence of an effect of the *ESR1* polymorphisms on lsBMD could indicate that the effect on vert fx risk is independent of BMD.

The haplotype analysis showed one significantly associated haplotype with vertebral fx risk and E2-levels. This haplotype was part of the previously defined *Pvu*II-*Xba*I haplotype 1 (TA), which was associated with higher vertebral fx risk and lower E2-levels[1,2] in the RSI-BC. So the results we find for this new haplotype are as we expected. The other frequent (>5%) haplotype that originates from the former haplotype 1 showed higher E2-levels in carriers, although this was not significant. The only difference between the two haplotypes is the T-allele of rs488133, which is present in the most frequent haplotype. This T-allele was associated with an increased fx risk and lower E2-levels itself, and could explain the difference we observe. From this it can be concluded that probably not *Pvu*II and *Xba*I are the functional or causal variants, but rs488133 or a SNP in high correlation with this SNP.

rs488133 was previously identified to be associated with HDL levels and functional studies showed that this polymorphism changed a glucocorticoid receptor binding site in the promoter of *ESR1*[12,13]. The polymorphism disrupts the glucocorticoid-

induced down-regulation of transcription of the estrogen receptor. No association or functionality studies for rs532010 have been performed until now. However, the mechanism by which this polymorphism affects the function or level of estrogen receptors in the cell still needs more research.

In conclusion, we identified seven additional SNPs in the *Pvu*II-*Xba*I LD-block of the *ESR1* gene to be associated with estradiol levels and vertebral fracture risk in the complete RSI cohort. One of these SNPs was previously identified to be functional and associated with HDL-levels. Because this study is based on tagging SNPs, further research is needed to identify the functional SNP(s) driving the associations.

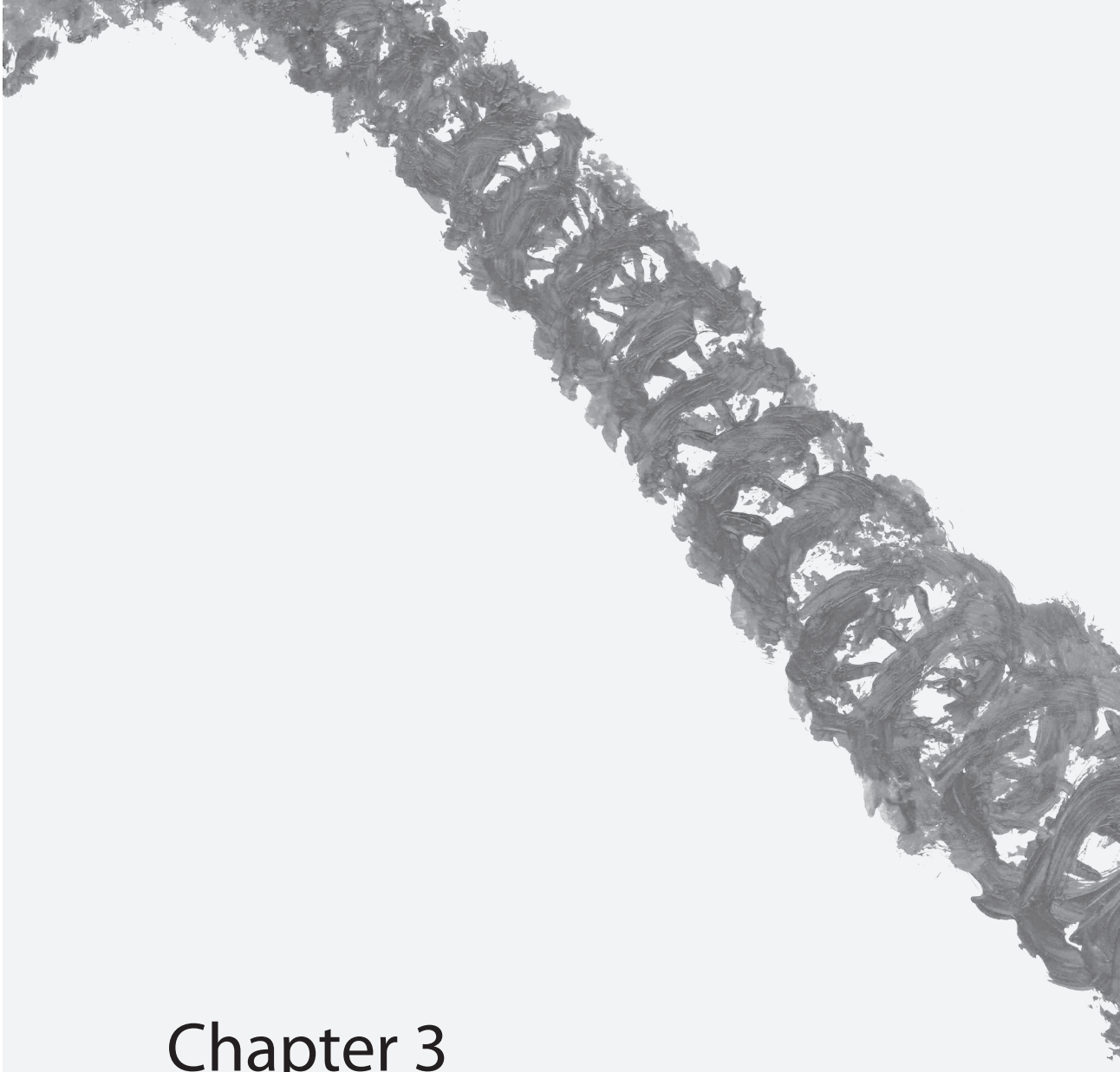
ACKNOWLEDGEMENTS

This study was funded by the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA), Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation 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 authors are very grateful to the participants of the Rotterdam Study and to the DXA and radiograph technicians, L Buist and HWM Mathot. Furthermore, we acknowledge all participating general practitioners and the many field workers in the research center in Ommoord, Rotterdam, The Netherlands. 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 Commision (DG XII), and the Municipality of Rotterdam.

REFERENCES

1. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. Eur J Endocrinol 153, 327-34 (2005).
2. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. Hum Mol Genet 12, 1745-54 (2003).
3. Weel, A.E. et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. J Clin Endocrinol Metab 84, 3146-50 (1999).
4. Ioannidis, J.P. et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. Jama 292, 2105-14 (2004).
5. Maruyama, H. et al. Lack of an association of estrogen receptor alpha gene polymorphisms and transcriptional activity with Alzheimer disease. Arch Neurol 57, 236-40 (2000).
6. Herrington, D.M. et al. Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. Circulation 105, 1879-82 (2002).
7. Hofman, A. et al. The Rotterdam Study: objectives and design update. Eur J Epidemiol 22, 819-29 (2007).
8. Burger, H. et al. The association between age and bone mineral density in men and women aged 55 years and over: the Rotterdam Study. Bone Miner 25, 1-13 (1994).
9. Van der Klift, M., De Laet, C.E., McCloskey, E.V., Hofman, A. & Pols, H.A. The incidence of vertebral fractures in men and women: the Rotterdam Study. J Bone Miner Res 17, 1051-6 (2002).
10. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 371, 1505-12 (2008).
11. 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).
12. Figtree, G.A. et al. A commonly occurring polymorphism upstream of the estrogen receptor alpha alters transcription and is associated with increased HDL. Atherosclerosis 199, 354-61 (2008).
13. Figtree, G.A., Guzik, T., Robinson, B.G., Channon, K.M. & Watkins, H. Functional estrogen receptor alpha promoter polymorphism is associated with improved endothelial-dependent vasolidation. Int J Cardiol (2009).



Chapter 3

The *RIZ* Pro704 insertion-deletion polymorphism, bone mineral density and fracture risk: the Rotterdam Study

Lisette Stolk, Joyce BJ van Meurs, Pascal P Arp, Albert Hofman, Huibert AP Pols, André G Uitterlinden

Bone 2008 ***2****(2):286-93.*

ABSTRACT

Estrogens play a major role in the maintenance of bone and bone strength, and they exert their effects via estrogen receptors. Recently, an estrogen receptor alpha (ERα) specific co-activator, Retinoblastoma-Interacting Zinc-finger protein (RIZ1, 1p36), was shown to strongly enhance ERα function in vitro. The same study showed that a proline insertion-deletion polymorphism at amino acid position 704 (Pro704 ins/del) in the *RIZ1* gene was associated with heel BMD in young Swedish women. We tested the relation between the *RIZ1* Pro704 ins/del polymorphism and BMD and fracture risk in Caucasian elderly men and women of the Rotterdam Study. We also examined whether estradiol levels (measured in a subset) or genetic variation in the estrogen receptor alpha (*ESR1*) gene influenced this relation.

In 2,424 men and 3,517 women from the Rotterdam Study, *RIZ1* genotypes were determined and associations with BMD (lumbar spine and femoral neck) and fracture risk were analysed. We recorded 374 vertebral fractures at baseline and during 6.4 (SD=0.4) years of follow-up, and 1,219 incident non-vertebral fractures.

The allele-frequency of the Pro704 insertion was 41%, the genotype distribution was in Hardy-Weinberg Equilibrium (P=0.94). We found no association of this polymorphism with BMD or fracture risk. Stratification for gender, estradiol levels or interaction with *ESR1* risk haplotype did not change these results.

In conclusion, in this large study we observed no association of the *RIZ1* Pro704 insertion-deletion polymorphism with BMD or fracture risk. This suggests this polymorphism to play a minor role, if any, as a genetic determinant of osteoporosis in elderly subjects.

INTRODUCTION

Osteoporosis is a skeletal disorder characterized by low bone mineral density (BMD) and micro architectural deterioration of bone, resulting in an increase in bone fragility and fracture risk[1]. Osteoporotic fractures are associated with substantial morbidity and mortality, and are therefore a major health care problem both in postmenopausal women and elderly men[2]. Twin studies have estimated a high heritability of BMD (up to 80%), bone turnover (63%), and bone geometry (62%), while heritability of risk for fractures is lower (25%-35%)[3]. The genes involved in these traits, however, are largely unknown.

It is known that estrogens play a major role in the development and maintenance of bone[4] and therefore genes regulating effects of sex steroid signalling are good candidates to study for contributing to explaining genetic variance of osteoporosis.

Estradiol (E2) exerts its effect via binding to the Estrogen Receptor alpha (*ESR1*, 6q25.1) and beta (*ESR2*, 14q22-24). Retinoblastoma-interacting zinc finger protein (RIZ) was previously characterized as a tumorsuppressor gene[5], but it was also shown to act as a coactivator of *ESR1*[6]. It forms a complex with ERα and the SRC1 or p300 complexes, thereby enhancing the activity of the receptor in the presence of estradiol[7]. When there is a low level of E2 or no E2 at all, RIZ acts as a repressor of E2-target genes[7].

There are several reasons why RIZ1 is a plausible candidate gene. First, the *RIZ*-gene is located on chromosomal region 1p36, a region that has been identified as a QTL for BMD[8]. Second, genetic variation in other genes involved in E2 signalling lead to inter-individual variation in BMD or fracture risk[9,10]. Third, Grundberg et al. showed that a 3bp deletion insertion polymorphism in the coding region of the gene is associated with heel BMD in young Swedish women (n=298)[11,12]. This polymorphism leads to an insertion/deletion of a Proline in exon 7, at amino acid location 704 (Pro704 ins/del). They also showed that the deletion had a lower enhancing effect on ESR1 coactivation compared to the insertion. In a study of elderly Swedish men and women, the same research group showed that this polymorphism was associated with femoral, lumbar spine and total body BMD in women (n=1,044). In elderly men (n=3,014) they found only an association with lumbar spine BMD. It was shown that this relation was influenced by plasma estradiol levels[11].

The aim of this study was to investigate if this polymorphism is associated with BMD and fracture risk in elderly women and men in the Rotterdam study. We also

tested for interaction of this polymorphism with the risk haplotype of the *Pvu*II and *Xba*I SNPs of the *ESR1* gene in the same study population[10]. Since action of RIZ was shown to be dependent on E2-levels, we also examined whether E2-levels influenced the relation between the *RIZ* variant and BMD and fracture risk.

MATERIAL and METHODS

Study populations

Subjects were participants of the Rotterdam Study, a large prospective population-based cohort study of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the elderly. Rationale and design have been described previously[13]. All 10,275 inhabitants aged 55 years and over were invited for baseline examination between August 1990 and June 1993. Of those, 7,983 participated. Among the subjects living independently, the overall response rate was 78 percent for home interview and 71 percent for examination in the research centre, where anthropometric characteristics and bone mineral density were measured, and blood samples were taken. The Rotterdam Study was approved by the medical ethics committee of the Erasmus University Medical School, and written informed consent was obtained from each subject. The current study is based on 5,720 subjects (3,348 postmenopausal females) for whom genotype data was available for the RIZ Pro704 deletion polymorphism.

In addition, we determined allele frequencies in a panel of subjects of different ethnic background from the NIGMS Human Diversity Panel (Coriell Institute, Camden, New Jersey, U.S.A.). The panel consisted of 60 African Americans (HD04 and HD50) and 110 Han-Chinese (HD02 and HD100).

Clinical examination

Height and weight were measured at baseline examination with the subject in a standing position with indoor clothing without shoes. BMD (in grams per square centimeter) was determined by dual-energy x-ray absorptiometry (DEXA, DPX-L densitometer, Lunar, Madison, WI) at the femoral neck (fnBMD) and lumbar spine (lsBMD) (vertebrae L2, L3, L4), as described elsewhere[14]. Age at menopause and smoking habits were assessed by a questionnaire.

Hormone assays

Levels of steroid hormones were measured in plasma of a subgroup of 528 males and 631 females, as described earlier[15]. Of these subjects 492 males (median E2-level: 44.02 pmol/L, range 0.0-179.6 pmol/L) and 548 females (median E2-level: 11.94 pmol/L, range: 0.0-85.8 pmol/L) were used in the analysis.

Assessment of Incident Non-vertebral Fractures and Vertebral Fractures

Follow-up for non-vertebral fractures started either the 1st of January 1991 or, if later, at the time of inclusion into the study. For this analysis, follow-up ended either at January 1st, 2002 or, when earlier, at the participant's death, comprising an average follow-up period of 7.4 (SD 3.3) years for non-vertebral fractures. For approximately 80% of the study population, medical events were reported through computerized general practitioner diagnosis registers. For the remaining 20%, research physicians collected data from the general practitioners' medical records of the study participants. All collected fractures were verified by reviewing discharge reports and letters from medical specialists. Fracture events were coded independently by two research physicians according to the International Classification of Diseases, 10th revision (ICD-10). Finally, an expert in osteoporosis reviewed all coded events for final classification. Any fracture was used as an outcome measure, to have sufficient power. All fractures which were considered not osteoporotic (fractures caused by cancer and all hand, foot, skull and face fractures) were excluded. In addition, we considered separately all fragility fractures that occur at older age, which included hip, proximal humerus and pelvis fractures.

Gene structure

The structure of the gene was determined using the NCBI gene database, The polymorphisms in the region were retained from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/ (10/03/2006)) and the HapMap genome browser (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B35/ (10/03/2006)). The haploblock structure of the gene was determined using Haploview v3.32[17] according to the method by Gabriel et al.[18].

Genotyping

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. 1-2 ng genomic DNA was dispensed into 384-wells plates using a Caliper Sciclone ALH3000 pipetting robot (Caliper LS, Mountain View, CA,

USA). Genotypes were determined using the Taqman allelic discrimination assay. The Assay-by-Design service (www.appliedbio-systems.com) was used to set up a Taqman allelic discrimination assay for the RIZ Pro704 insertion/deletion polymorphism (Primers Fw: TCAAAGCTCAAACAACTTCTTC-AAACC Rv: GCAGACACACAAACAGGACCTAATT, probes FAM: TCCCTGCA-GGAGTT, VIC: CTGCAGGAGGAGTT). The PCR reaction mixture included 1-2 ng of genomic DNA in a 2 μl volume and the following reagents: FAM and VIC probes (200 nM), primers (0.9 uM), 2x Taqman PCR master mix (ABgene, Epsom, UK). Reagents were dispensed in the 384-well plates using the Deerac Equator NS808 (Deerac Fluidics, Dublin, Ireland). PCR cycling reactions were performed in 384 wells PCR plates in an ABI 9700 PCR system (Applied Biosystems Inc., Foster City, CA, USA) and consisted of initial denaturation for 15 minutes at 95°C, and 40 cycles with denaturation of 15 seconds at 95°C and annealing and extension for 60 seconds at 60°C. Results were analysed by the ABI Taqman 7900HT using the sequence detection system 2.22 software (Applied Biosystems Inc., Foster City, CA, USA). To confirm the accuracy of genotyping results, 332 (5%) randomly selected samples were re-genotyped with the same method. No inconsistencies were observed. Genotyping of the *Pvu*II and *Xba*I polymorphisms in the *ESR1*-gene was described elsewhere[15].

Statistical analysis

Hardy Weinberg Equilibrium (HWE) was calculated according to standard procedures using Chi-square analysis. Power for BMD and non-vertebral fracture risk was calculated using PS Power and Sample size calculations v2.1.30[19] for an α of 0.05, for both a dominant and a recessive model.

For each allele subjects were grouped according to genotype by allele copy number (0, 1, and 2, corresponding to non-carriers, heterozygote carriers and homozygote carriers, respectively). We allowed for three possible genetic models to explain differences between groups, i.e., an allele-dose effect, a dominant effect or a recessive effect. Allele dose was defined as the number of copies of a certain allele in the genotype. In case of a consistent trend reflected as an allele-dose effect, we performed a linear regression analysis to quantify the association. In case of a dominant or a recessive effect of the test allele, analysis of (co)variance AN(C)OVA was performed to test for differences between two genotype groups. For dominant alleles, we compared test-allele carriers versus non-carriers.

Hazard ratios (HRs) with 95% confidence intervals (95%CI) were calculated by logistic regression analyses to estimate the relative risk of fractures by genotype for the risk allele, with no copies of the risk allele as the reference group. We first calculated

crude hazard ratios and then adjusted for potentially confounding factors (e.g., age, height, weight, and BMD). To estimate non-vertebral fracture risk by genotype, we used Cox proportional hazard models, thereby taking potential differences in time-to-event into account. To estimate the risk of vertebral fractures, odds ratios (ORs) with 95% confidence intervals (95% CI) were calculated using logistic regression models. We were not able to use Cox proportional hazard models for vertebral fractures since the exact time of the event was not known. All statistical analyses were performed using SPSS version 11.0.1 (SPSS Inc., Chicago, USA). P-values are two-sided, and P=0.05 or less was considered significant.

RESULTS

Genotyping and gene structure

In our Caucasian study population the allele frequency of the Proline-insertion was 41%, the genotype distribution was in Hardy-Weinberg equilibrium (P=0.94). Figure 1 shows a schematic overview of the *RIZ1*-gene with the polymorphisms found in two databases, dbSNP (182 polymorphisms) and HapMap (133 polymorphisms, all present in the dbSNP database), indicated with vertical lines. The Proline insertion deletion polymorphism is located in exon 7 at amino acid position 704. We analysed the LD-block structure across the gene based on HapMap-data (depicted as horizontal lines above the gene structure). The Pro704 ins/del variant is situated in haploblock I in the gene, reaching from the 5'UTR until intron 7. Figure 1 also shows the allele frequency of the Pro704 insertion allele in three different ethnic groups, in African Americans it was 55%, and in Han-Chinese 60%. The frequency of the insertion allele is higher in both ethnic groups compared to the frequency in Caucasians in our study population (41%).

Table 1. Baseline anthropometric measures according to RIZ Pro704 Ins/Del genotype.

	MEN				WOMEN			
	Del/Del	**Del/Ins**	**Ins/Ins**	**P**	**Del/Del**	**Del/Ins**	**Ins/Ins**	**P**
N (%)	839 (35)	1,111 (47)	419 (18)		1,174 (35)	1,628 (49)	546 (16)	
Age (yrs)	68.2 (8.2)	67.9 (7.9)	68.1 (8.3)	0.68	69.4 (9.1)	70.0 (9.4)	69.6 (9.3)	0.32
Height (cm)[1]	174.7 (6.6)	174.8 (6.5)	174.8 (6.6)	0.94	161.4 (6.2)	161.2 (6.2)	161.4 (6.2)	0.70
Weight (kg)[1]	78.6 (10.3)	78.6 (10.3)	78.2 (10.3)	0.71	69.7 (11.3)	69.6 (11.3)	69.3 (11.3)	0.73
BMI (kg/m^2)[1]	25.7 (2.9)	25.7 (2.9)	25.6 (2.9)	0.55	26.8 (4.1)	26.8 (4.1)	26.6 (4.1)	0.56

[1] values (SD) are adjusted for age; p-values are for AN(C)OVA.

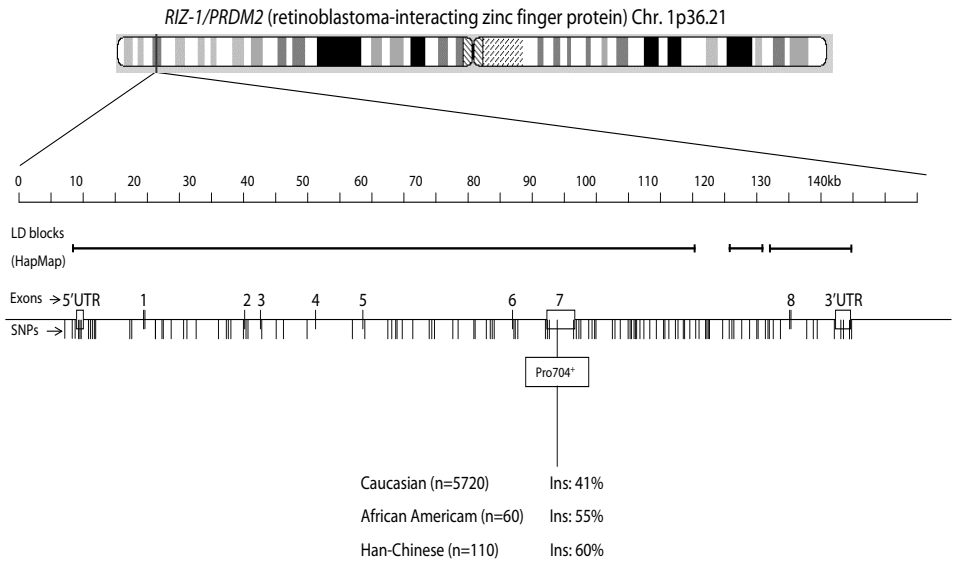


Figure 1. Schematic overview of the *RIZ1*/*PRDM2* gene. The gene spans about 125 kb on chromosome 1p36.21. In light grey the 5'UTR and 3'UTR are depicted. The exons (1 through 8) are shown in dark grey. The vertical lines represent the polymorphisms in the gene described in the dbSNP and HapMap databases. The horizontal lines above the gene structure are a schematic representation of the three LD-blocks according to HapMap data. The Pro704 insertion deletion polymorphism is located in exon 7. The allele frequency of the insertion allele is also shown for a Caucasian, an African American and a Han-Chinese population.

RIZ1 Pro704 ins/del and BMD and fracture risk

Table 1 depicts the baseline measures according to the *RIZ1* Pro704 insertion deletion polymorphism. In both men and women there was no relation between the polymorphism and baseline measures.

We examined the association of the polymorphism with lumbar spine (ls) and femoral neck (fn) BMD in men and women. As shown in Figure 2, there is no association with these BMD measures in this study. Furthermore there was no association of the *RIZ* Pro704 insertion deletion polymorphism with either vertebral or incident non-vertebral fracture risk in elderly men and women (Figure 3).

Power calculations showed that we could detect a difference in BMD of 0.02 g/cm^2 for a dominant model, and 0.005 g/cm^2 for a recessive model with 80% power. For non-vertebral fractures we have 80% power to detect a relative risk of 1.4 for males and 1.2 for females in a dominant model, while in a recessive model these relative risks were 1.5 and 1.3 respectively.

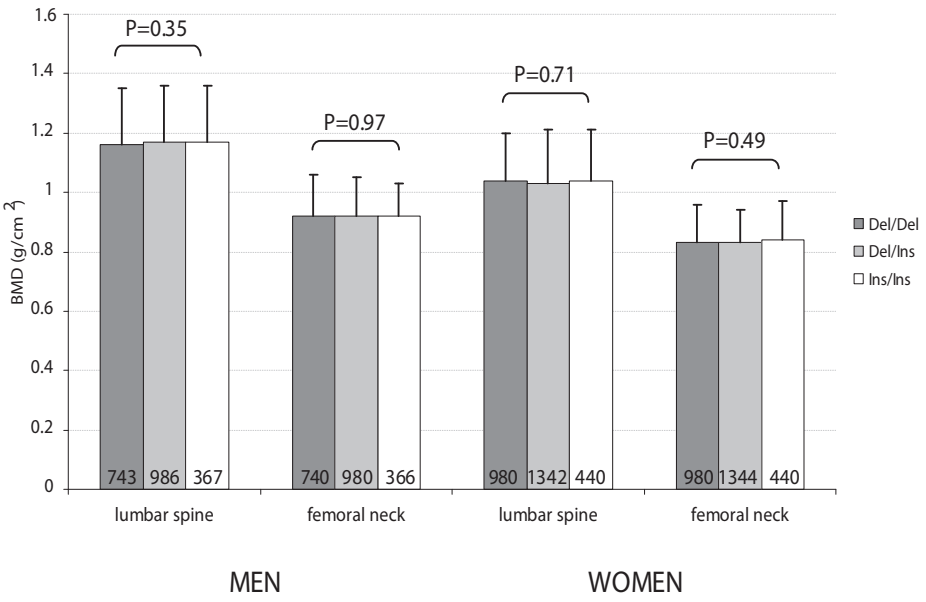


Figure 2. Lumbar spine and femoral neck BMD in men and women according to *RIZ* Pro704 insertion deletion polymorphism. Homozygous for the deletion is depicted in dark grey, heterozygous in light grey and homozygous insertion in white. The error bars depict SDs. P-values are adjusted for age, height, and weight. In each bar of the graphs the number of subjects in that group is depicted.

No effect of E2-levels or interaction with ESR1 polymorphisms

In a subset of individuals (492 males and 548 females) we stratified the analysis according to levels of estradiol in two equal groups. Low and high were defined using the median estradiol level in men and women. For men this was 44.02 pmol/L and for women the median level was 11.94 pmol/L. Stratification for estradiol levels did not change the result for both lsBMD and non-vertebral fracture risks (Figure 4). For fnBMD and vertebral fracture risk similar results were obtained (data not shown). In women we also stratified for detectable vs. non-detectable levels of E2, but this did not change the results.

Next we studied the interaction of the *RIZ1* Pro704 insertion deletion polymorphism with the risk haplotype of *ESR1*, haplotype 1 of the *Pvu*II/*Xba*I polymorphisms. For lsBMD and non-vertebral fracture risk of the *RIZ1* polymorphism in two different genotype groups of *ESR1* haplotype 1 (non-carriers, and carriers of haplotype 1, Figure 5), no association was observed. For fnBMD and vertebral fracture risk we observed similar results.

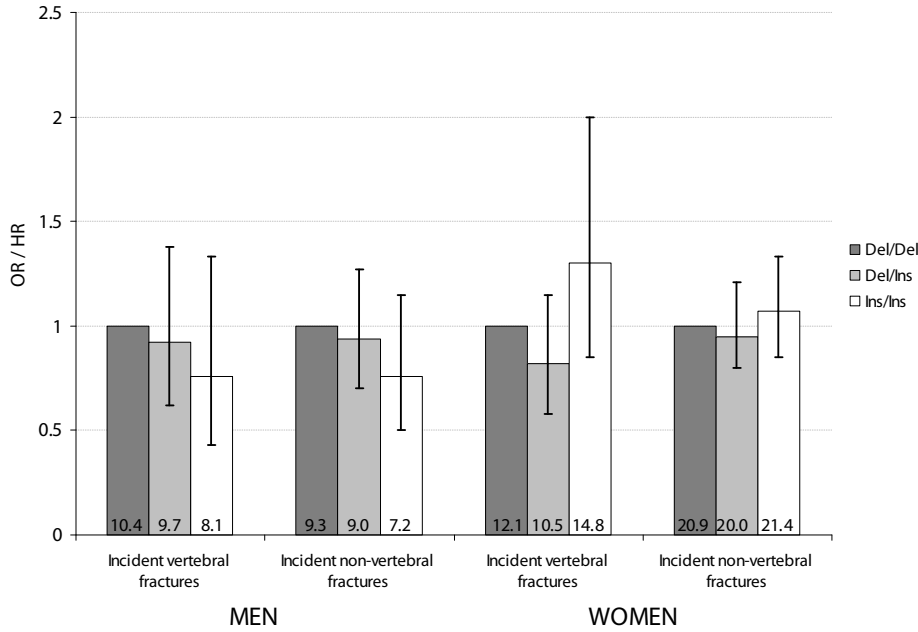


Figure 3. Incident vertebral fractures and non-vertebral fractures in men and women according to *RIZ* Pro704 insertion deletion polymorphism. For vertebral fracture risks ORs are shown, for nonvertebral fractures HRs. Homozygous for the deletion is depicted in dark grey, heterozygous in light grey and homozygous insertion in white. The error bars depict 95% confidence intervals. P-values are adjusted for age, height, and weight. In each bar the percentage of subjects with a fracture in that group is depicted.

DISCUSSION

In this study of elderly women and men of the Rotterdam study we observed no association of the *RIZ* Pro704 insertion deletion polymorphism with neither lumbar spine BMD or femoral neck BMD, nor with fracture risk. Stratification according to high and low estradiol levels yielded similar results. In addition, no influence of the risk haplotype of the *Pvu*II and *Xba*I SNPs of the *ESR1* gene was found on the relation between the *RIZ1* polymorphism and bone-related endpoints.

The aim of this study was to replicate and extend the findings published earlier by Grundberg et al., who showed a significant association with heel BMD in young Swedish women, with lumbar spine BMD in elderly men and with femoral, lumbar spine and total body BMD in elderly women, but no association with self-reported osteoporotic fractures[11,12].

Our study population is large and we have sufficient statistical power to

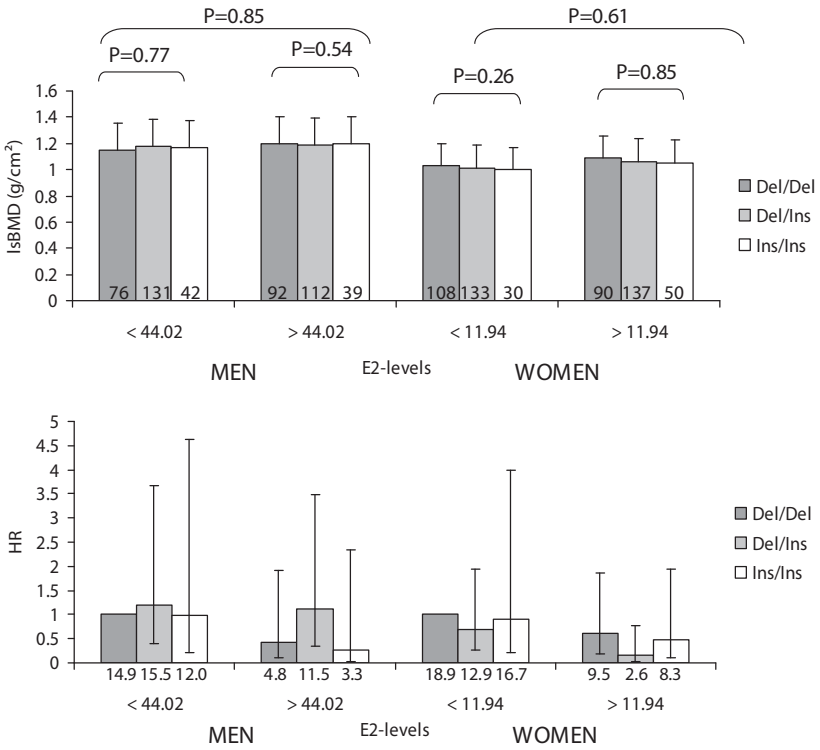


Figure 4. Lumbar spine bone mineral density and vertebral fracture risks in men and women of the Rotterdam Study according to *RIZ* Pro704 insertion deletion polymorphism, stratified for estradiol levels. Homozygous for the deletion is depicted in dark grey, heterozygous in light grey and homozygous insertion in white. For BMD the error bars depict SDs, for fracture risks the error bars depict 95% confidence intervals. P-values are adjusted for age, height, and weight (and BMD in case of fracture risks). In each bar of the BMD graphs the number of subjects in that genotype group is depicted, for the fracture risk the number is the percentage of subjects with a fracture in that group.

detect the previously reported genetic effects on BMD by Grundberg et al. While lack of statistical power does not seem to be the explanation for our lack of replication, there are several differences in the study population between the two studies.

First, Dutch Caucasians can differ from Swedish Caucasians in their genetic architecture and in environmental factors interacting with them. Although allele frequencies between the 2 study groups are similar (58% vs. 56% for the Pro704 del allele), the genetic background might be different and genetic interactions could contribute to explain the lack of effect of this polymorphism. We looked at the interaction of the *RIZ* polymorphism (Pro 704 deletion) with *ESR1* polymorphisms (*ESR1* haplotype 1), because the RIZ1 and ERα proteins interact in estradiol-signalling. While

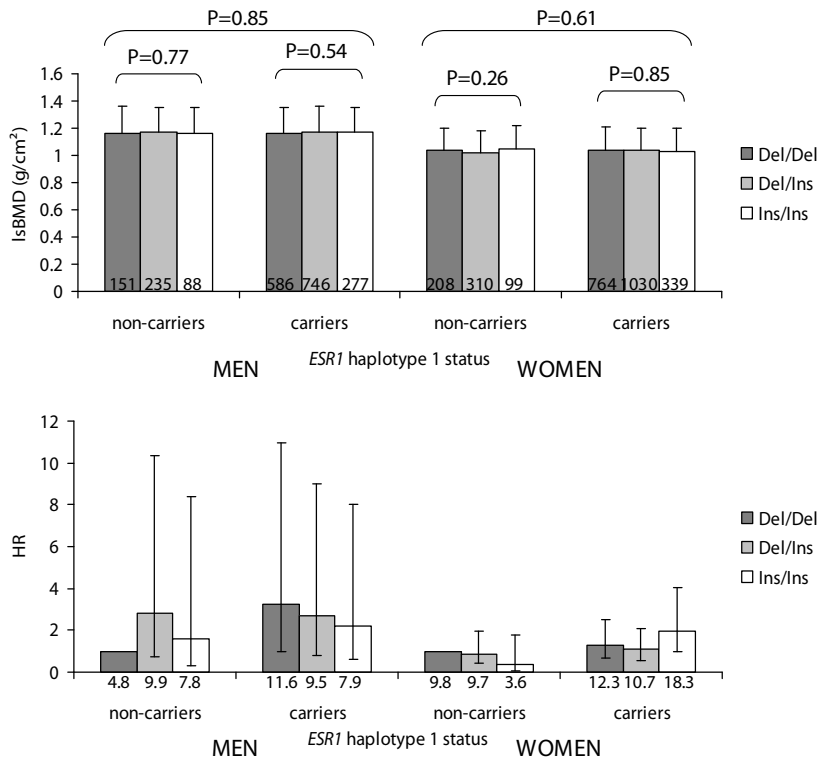


Figure 5. Lumbar spine bone mineral density and vertebral fracture risks in men and women of the Rotterdam Study according to *RIZ* Pro704 insertion deletion polymorphism, stratified for ESR1 haplotype 1 carrier-status. Homozygous for the deletion is depicted in dark grey, heterozygous in light grey and homozygous insertion in white. For BMD the error bars depict SDs, for fracture risks the error bars depict 95% confidence intervals. P-values are adjusted for age, height, and weight (and BMD in case of fracture risks). In each bar of the BMD graphs the number of subjects in that genotype group is depicted, for the fracture risk the number is the percentage of subjects with a fracture in that group.

having risk variants in both these genes could lead to lower BMD and higher fracture risk, we did not observe such interaction. Yet, this is only one possible interaction and there are many more potential ones left untested.

Second, in a cell-line model Carling et al. previously observed the action of RIZ1 to be dependent on E2-levels[7]. Grundberg et al. showed that the Pro704 deletion-allele had a lower enhancing effect in the presence of E2 compared to the insertion-allele[12], and they observed an opposite effect on BMD of the *RIZ1* polymorphism in high vs. low levels of E2[11]. We therefore also stratified the analysis for estradiol levels using a sex-adjusted median of E2-levels, but we did not see an effect of stratification for E2-levels on BMD or fracture risk. Therefore, although, there is a positive correlation

between estradiol levels and BMD in this study population which we described earlier[20], this does not seem to be *RIZ* genotype dependent. The discrepancy between this study and the Swedish study could be due to differences in the E2-levels between populations. E2-levels are 50% lower in our population compared to the levels in the Swedish men, with a median of 44.02 pmol/L and 92.6 pmol/L, respectively. This difference could be due to method of measurement of the E2-levels . In our study a kit from a different manufacturer was used. Although both kits had the same detection limit, in our study the intra- and inter-assay CV was higher (18% and 20%) compared to the Swedish study (3% and 6%). There was also a difference between the two studies in the albumin levels used in the calculation of E2-levels. In our population albumin levels were measured whereas in the Swedish population a fixed level of 43 g/L albumin was applied for all subjects. Furthermore, we did only single measurements due to small volumes of plasma available, where the Swedish study used duplicate measurements. Another explanation for this difference could be our lack of power due to the limited number of men in our study. We had only data on E2-levels available for 528 men, whereas the Swedish study used 2,914 men in their analysis. However, for an alpha of 0.05, we do have 80% power to pick up a difference of 0.33 SD in lsBMD between the two extreme genotypes in our study population.

Third, the *RIZ1*-gene is located on chromosome 1p36, a frequently observed QTL for BMD[8]. Other candidate genes, like methylene tetrahydrofolate reductase (*MTHFR*), lysyl hydroxylase (*PLOD1*) and tumor necrosis factor receptor superfamily, member 1B (*TNFRSF1B*), are also located in this area and have all been previously found to be associated with BMD[21,22]. It could be that one or more of these genes and even other genes in this region all play a role in determining BMD, and that these effects might differ across Dutch and Swedish populations due to differences in LD structure. We also examined only a single SNP in this QTL instead of haplotypes, while it has been suggested that studying haplotypes can increase power to detect rare causal alleles[23]. Yet, we studied a polymorphism with well-established functional effects, and therefore we did not pursue a haplotype tagging approach in this study. Based on the LD-block structure shown in Figure 1 we can, however, not exclude that other variations in the gene could contribute to the effects on BMD seen in the study of Grundberg et al.[11,12].

Finally, we cannot exclude that either one of the two studies represents a false positive (Grundberg et al.) or a false negative (this study). In any case, the *RIZ1* Pro704 del variant does not seem to have an universal association with either BMD and/or fracture risk. Additional well-powered association studies will therefore remain necessary to solve this issue.

In conclusion, in this study of Dutch Caucasian elderly men and women we were not able to replicate the previous findings in Swedish Caucasian men and women on the association of the *RIZ1* Pro704 insertion deletion polymorphism with BMD and fracture risk.

ACKNOWLEDGEMENTS

This study is supported by the Netherlands Organization of Scientific Research (NWO)-Research Institute for Diseases in the Elderly (Grant 014-93-015; RIDE2) and the European Commision (Grant QLK6-CT-2002-02629 (GENOMOS)). The authors are very grateful to the participants of the Rotterdam study and to the DXA and radiograph technicians, L Buist and HWM Mathot. Furthermore, we acknowledge all participating general practitioners and the many field workers in the research center in Ommoord, Rotterdam, The Netherlands.

WEBRESOURCES

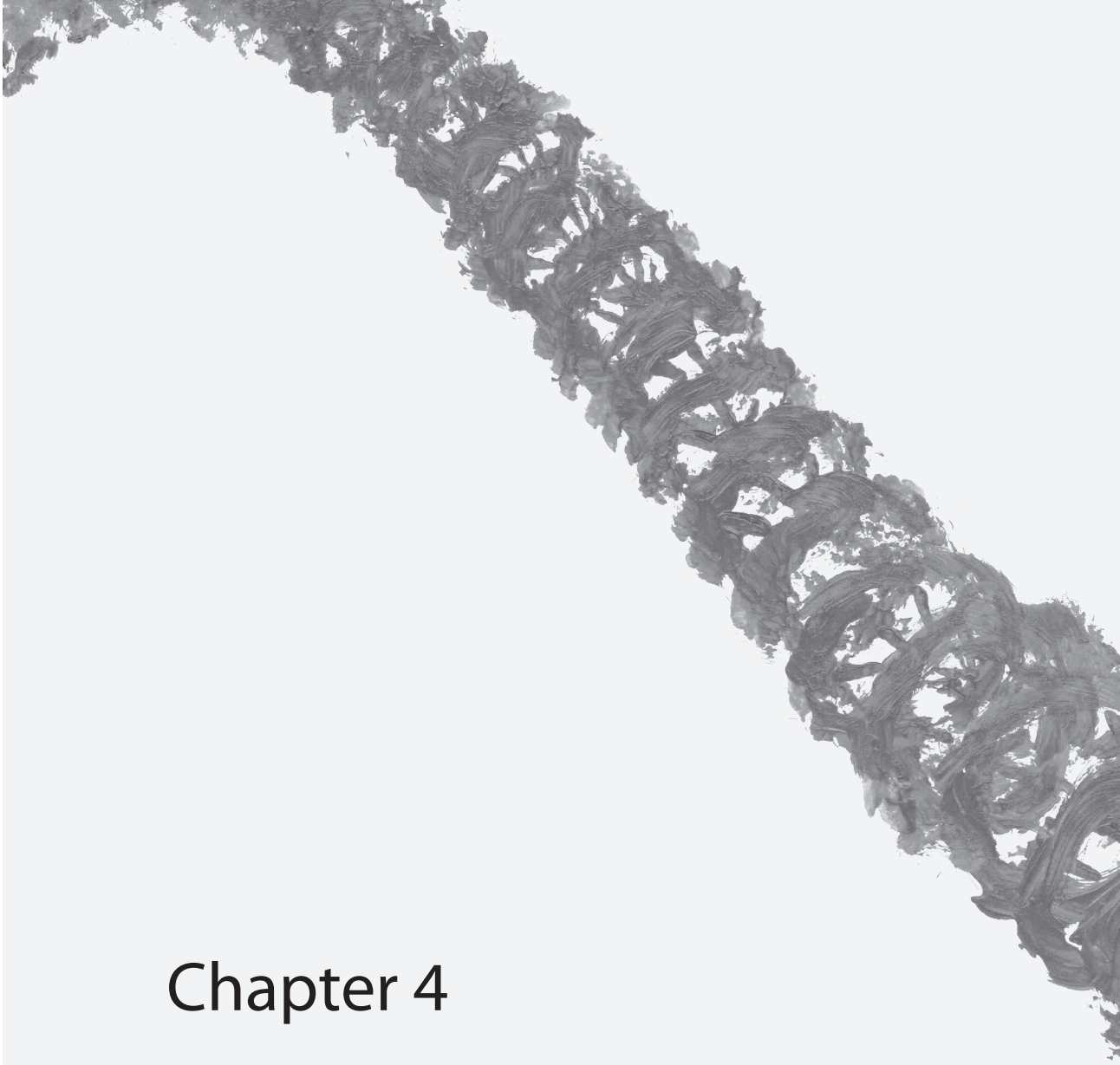
dbSNP: http://www.ncbi.nlm.nih.gov/SNP/ (10/03/2006)
HapMap release #21: http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B35/ (10/03/2006)

REFERENCES

1. Cooper, C. Epidemiology of osteoporosis. *Osteoporos Int* 9 Suppl 2, S2-8 (1999).
2. Melton, L.J., 3rd. Hip fractures: a worldwide problem today and tomorrow. *Bone* 14 Suppl 1, S1-8 (1993).
3. Ralston, S.H. Genetic determinants of susceptibility to osteoporosis. *Curr Opin Pharmacol* 3, 286-90 (2003).
4. Riggs, B.L., Khosla, S. & Melton, L.J., 3rd. Sex steroids and the construction and conservation of the adult skeleton. *Endocr Rev* 23, 279-302 (2002).
5. Huang, S. The retinoblastoma protein-interacting zinc finger gene RIZ in 1p36-linked cancers. *Front Biosci* 4, D528-32 (1999).
6. Abbondanza, C. et al. The retinoblastoma-interacting zinc-finger protein RIZ is a downstream effector of estrogen action. *Proc Natl Acad Sci U S A* 97, 3130-5 (2000).
7. Carling, T. et al. A histone methyltransferase is required for maximal response to female sex hormones. *Mol Cell Biol* 24, 7032-42 (2004).
8. Wilson, S.G. et al. Comparison of genome screens for two independent cohorts provides replication of suggestive linkage of bone mineral density to 3p21 and 1p36. *Am J Hum Genet* 72, 144-55 (2003).
9. 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).
10. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. *Hum Mol Genet* 12, 1745-54 (2003).
11. Grundberg, E. et al. The impact of estradiol on bone mineral density is modulated by the specific estrogen receptor-alpha cofactor retinoblastoma-interacting zinc finger protein-1 insertion/deletion polymorphism. *J Clin Endocrinol Metab* 92, 2300-6 (2007).
12. Grundberg, E. et al. A deletion polymorphism in the RIZ gene, a female sex steroid hormone receptor coactivator, exhibits decreased response to estrogen in vitro and associates with low bone mineral density in young Swedish women. *J Clin Endocrinol Metab* 89, 6173-8 (2004).
13. 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).
14. Burger, H. et al. The association between age and bone mineral density in men and women aged 55 years and over: the Rotterdam Study. *Bone Miner* 25, 1-13 (1994).
15. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. *Eur J Endocrinol* 153, 327-34 (2005).
16. Van der Klift, M., De Laet, C.E., McCloskey, E.V., Hofman, A. & Pols, H.A. The incidence of vertebral fractures in men and women: the Rotterdam Study. *J Bone Miner Res* 17, 1051-6 (2002).
17. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-5 (2005).
18. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. *Science* 296, 2225-9 (2002).
19. Dupont, W.D. & Plummer, W.D., Jr. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 19, 589-601 (1998).
20. Goderie-Plomp, H.W. et al. Endogenous sex hormones, sex hormone-binding globulin, and the risk of incident vertebral fractures in elderly men and women: the Rotterdam

Study. *J Clin Endocrinol Metab* 89, 3261-9 (2004).

21. Abrahamsen, B. et al. A common methylenetetrahydrofolate reductase (C677T) polymorphism is associated with low bone mineral density and increased fracture incidence after menopause: longitudinal data from the Danish osteoporosis prevention study. *J Bone Miner Res* 18, 723-9 (2003).
22. Spotila, L.D. et al. Association analysis of bone mineral density and single nucleotide polymorphisms in two candidate genes on chromosome 1p36. *Calcif Tissue Int* 73, 140-6 (2003).
23. de Bakker, P.I. et al. Efficiency and power in genetic association studies. *Nat Genet* 37, 1217-23 (2005).



Chapter 4

The Catechol-O-Methyltransferase Met158 Low-Activity Allele and Association with Nonvertebral Fracture Risk in Elderly Men

Lisette Stolk, Joyce BJ van Meurs, Mila Jhamai, Pascal P Arp, Johannes PT van Leeuwen, Albert Hofman, Frank H de Jong, Huibert AP Pols, and André G Uitterlinden

J Clin Endocrinol Metab. 2007 ***92****(8):3206-12*

ABSTRACT

Because sex steroids play an important role in bone development, variants in genes encoding proteins involved in estrogen synthesis and metabolism could contribute to interindividual variation in bone parameters and fracture risk. An example is catechol-*O*-methyltransferase (COMT), an estrogen-degrading enzyme involved in inactivation of catechol-estrogens. Its gene contains a functional valine to methionine substitution at codon 158. The aim of our study was to determine whether this polymorphism is associated with bone parameters and fracture risk in elderly subjects.

COMT genotypes were determined using TaqMan allelic discrimination in 2,515 men and 3,554 women from the Rotterdam Study, a population-based cohort study of individuals aged 55 and older. Associations with bone mineral density (BMD) and bone loss were analyzed using ANOVA or analysis of covariance, whereas fracture risk was analyzed using Cox's proportional hazard regression analysis. *COMT* mRNA expression in three osteoblastic cell lines (SaOS, MG63, and SVHFO) was analyzed by RT-PCR.

Male carriers of the Met158 allele had an increased risk for osteoporotic fractures (hazard ratio (HR)=1.6; 95% confidence interval (95%CI[1.0 –2.4])) and for fragility fractures (HR=2.7; 95%CI[1.3–5.9]), with evidence for a dominant effect. Adjustments for age, height, weight, and BMD did not change the risk estimates. In women, this association was weaker and not significant. BMD was not significantly associated with the variant in either men or women. *COMT* mRNA was expressed in all three osteoblastic cell lines tested.

The *COMT* Val158Met polymorphism is associated with fracture risk in elderly men, through a mechanism independent of BMD.

INTRODUCTION

Osteoporosis is a skeletal disorder characterized by low bone mineral density (BMD) and microarchitectural deterioration of bone, resulting in an increase in bone fragility and fracture risk[1]. Osteoporotic fractures are associated with substantial morbidity and mortality and are therefore a major health care problem both in postmenopausal women and elderly men[2]. Twin studies have estimated a high heritability of BMD (up to 80%), bone turnover (63%), and bone geometry (62%), whereas heritability of risk for fractures is lower (25%–35%)[3]. The genes involved in these traits, however, are largely unknown.

Sex steroids play a major role in bone development[6], and therefore, genes regulating sex steroid production and metabolism are good candidates to study for involvement in the development of osteoporosis. There are several candidate genes in the estrogen pathway. We previously reported on genetic variations in the estrogen receptor α (*ESR1*) gene to be associated with BMD, height, and fracture risk[7-9]. Variations in genes involved in estrogen synthesis, like aromatase (*CYP19*), are also associated with bone parameters[10]. Until now, not much attention has been given to genetic variation of genes involved in estrogen degradation. Estrogens are degraded in several steps; the two major steps are shown in Figure 1. The first step is oxidation by cytochrome P450s (e.g,. CYP1A1, -1B1, and -3A4), resulting in catechol-estrogens. One study examining *CYP1A1* observed a nonsynonymous single-nucleotide polymorphism (SNP) in the gene to be associated with BMD in postmenopausal Caucasian women[11].

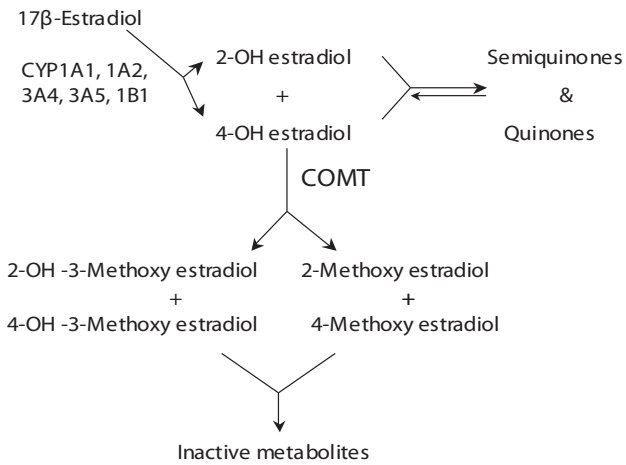


Figure 1. Pathway for oxidative metabolism of estradiol (modified from[4,5])

This suggests that there might also be a link between genetic variation in estrogen-degrading enzymes and differences in BMD.

COMT is the gene coding for catechol-*O*-methyltransferase (COMT), which inactivates circulating catechol-estrogens by catalyzing *O*-methylation of 2-hydroxylated and 4-hydroxylated estrogens to their methoxy derivatives 2-OH-methoxy-estrogen and 4-OH-methoxy-estrogen[5]. The best-known polymorphism in the *COMT* gene is a functional G to A substitution, leading to a valine (G) to methionine (A) substitution at codon 158. The methionine variant results in thermolability of the enzyme[12] and a 3- to 4-fold lower enzyme activity compared with the valine variant[13]. The alleles are therefore known as a low-activity ($COMT^L$) and a high-activity ($COMT^H$) allele. The frequency of the methionine variant ($COMT^L$) is around 50% in Caucasian populations.

In middle-aged men, the low-activity allele was shown to be associated with increased serum estradiol levels[14]. This finding indicates a link between COMT polymorphisms and estrogens. In postmenopausal women, however, there was no association found between the $COMT^L$ genotype and normal estradiol levels[15].

With regard to bone metabolism, not many studies have been published on COMT. In 458 young adult Swedish men, the $COMT^L$-allele was associated with a 4% lower peak BMD in the femur[16]. Yet, in a cohort of 1,795 postmenopausal women, no association of the $COMT^L$-allele with femoral neck and spine BMD was found[17]. The expression of COMT in bone cells could indicate a more local effect of COMT on bone than a systemic effect on estrogen levels; however, this has not been studied as far as we know.

Here we investigated the effect of the *COMT* Val158Met variant on BMD, fracture risk, and estrogen levels in elderly men and women of the Rotterdam Study. To study the local expression of COMT in bone cells, we looked at mRNA expression in three different osteoblast-like cell lines.

SUBJECTS and METHODS

Study populations

Subjects were participants of the Rotterdam Study, a large prospective population-based cohort study of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the

elderly. Rationale and design have been described previously[18]. All 10,275 inhabitants aged 55 year and over were invited for baseline examination between August 1990 and June 1993. Of those, 7,983 participated. Among the subjects living independently, the overall response rate was 77% for home interview and 71% for examination in the research center, where anthropometric characteristics and BMD were measured and blood samples were taken. The Rotterdam Study was approved by the medical ethics committee of the Erasmus University Medical School, and written informed consent was obtained from each subject. The current study is based on 6,298 subjects (2,570 male) for whom genotype data were available for the *COMT* Val158Met polymorphism. Baseline measurements of BMD were available for 5,156 subjects (82% of the genotyped cohort, 2,217 males) and both height and weight data for 6,069 subjects (96% of the genotyped cohort, 2,515 males). In addition, we determined allele frequencies in a panel of subjects of different ethnic background from the National Institute of General Medical Sciences Human Diversity Panel (Coriell Institute, Camden, NJ). The panel consisted of 60 African-Americans (HD04 and HD50) and 110 Han-Chinese (HD02 and HD100).

Clinical examination

Height and weight were measured at baseline examination with the subject in a standing position with indoor clothing without shoes. BMD (in grams per square centimeter) was determined by dual-energy x-ray absorptiometry (DPX-L densitometer; Lunar, Madison, WI) at the femoral neck and lumbar spine (vertebrae L2, L3, and L4), as described elsewhere[19]. Age at menopause and smoking habits were assessed by a questionnaire.

Hormone assays

Estrone, estradiol, androstenedione, and testosterone levels were determined in a gender-stratified random sample (n = 1,159) as described earlier[20].

Assessment of incident nonvertebral fractures and vertebral fractures

For nonvertebral fractures, follow-up started either January 1, 1991, or, if later, at the time of inclusion into the study. For this analysis, follow-up ended either at January 1, 2002, or, when earlier, at the participant's death, comprising an average follow-up period of 7.4 (SD=3.3) years for nonvertebral fractures. For approximately 80% of the study population, medical events were reported through computerized general practitioner diagnosis registers. For the remaining 20%, research physicians

collected data from the general practitioners' medical records of the study participants. All collected fractures were verified by reviewing discharge reports and letters from medical specialists. Fracture events were coded independently by two research physicians according to the International Classification of Diseases, 10th revision (ICD-10). Finally, an expert in osteoporosis reviewed all coded events for final classification. Any fracture was used as an outcome measure to have sufficient power. All fractures that were considered not osteoporotic fractures caused by cancer and all hand, foot, skull, and face fractures) were excluded. In addition, we considered separately all fragility fractures that occur at older age, which included hip, proximal humerus, and pelvis fractures. Both at baseline and at follow-up visits between 1997 and 2001, thoracolumbar radiographs of the spine were obtained. The follow-up radiographs were available for 3,469 individuals (1,498 men) who survived an average of 6.4 (SD=0.4) years after the baseline center visit and who were still able to come to our research center. All follow-up radiographs were scored for the presence of vertebral fracture by the McCloskey/Kanis method as described earlier[21]. If a vertebral fracture was detected, the baseline radiograph was evaluated as well. If the vertebral fracture was already present at baseline, it was considered a baseline prevalent fracture. If it was not present at baseline, the fracture was defined to be incident.

Genotyping

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. Genomic DNA (1–2 ng) was dispensed into 384-well plates using a Caliper Sciclone ALH3000 pipetting robot (Caliper LS, Mountain View, CA). Genotypes were determined using the TaqMan allelic discrimination assay. The Assay-by- Design service (www.appliedbio-systems.com) was used to set up a TaqMan allelic discrimination assay for the COMT Val158Met polymorphism (primers were CGAGATCAACCCCGACTGT (forward) and CAGGCATGCACACCTTGTC (reverse); probes were FAM, TCGCTGGCATGAAG, and VIC, TTTCGCTGGCGTGAAG). The underlined characters indicate the SNP position. The PCR mixture included 1–2 ng genomic DNA in a 2-μl volume and the following reagents: FAM and VIC probes (200 nM), primers (0.9 μM), 2x Taqman PCR master mix (ABgene, Epsom, UK). Reagents were dispensed in the 384-well plates using the Deerac Equator NS808 (Deerac Fluidics, Dublin, Ireland). PCR cycling reactions were performed in 384-well PCR plates in an ABI 9700 PCR initial denaturation for 15 min at 95°C and 40 cycles with denaturation of 15 sec at 95°C and annealing and extension for 60 sec at 60°C. Results were analyzed by the ABI TaqMan 7900HT using the sequence detection system 2.22 software (Applied Biosystems). To

confirm the accuracy of genotyping results, 332 (5%) randomly selected samples were re-genotyped with the same method. No inconsistencies were observed.

Quantitative real-time PCR

Cell nuclear extracts from SaOS and MG-63 (human osteoblast-like sarcoma cells) and SV-HFO (human preosteoblast cells) at different differentiation stages were pooled per cell line and were prepared as described elsewhere[22]. Quantitative real-time PCR was carried out using an ABI 7700 sequence detection system (Applied Biosystems). Reactions were performed in 25-μl volumes using a qPCR core kit (Eurogentec, Seraing, Belgium). Reaction mixtures contained 20 ng cDNA, 5 mm MgCl2, 200 μM dNTPs, and 0.025 U/μl Hot GoldStar enzyme (QIAGEN, Valencia, CA). Primer and probe sets were designed using the Primer Express software (version 2.0; Applied Biosystems). The amount of human glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA was used as internal control to normalize for possible differences in RNA extraction and degradation as well as efficiency of the cDNA synthesis. The cDNA primers COMT-For (5'-AGGAGTGGGCCATGAACGT-3'), COMT-Rev (5'-GGCTGGTGCTCCTGAATCAC-3'), GAPDH-For (5'-ATGGGGAAGGTGAAGGTCG-3'), and GAPDH-Rev (5'-TAAAAGCAGCCCTGGTGACC-3') were used to amplify the *COMT* and *GAPDH* cDNA, respectively. The probe sequences for *COM*T and *GAPDH* were COMT-FAM (5'-FAM-ACAAGAAAGGCAAGATCGTGGACGAA-3') and GAPDH-FAM (5'-FAM-CGCCCAATACGACCAAATCCGTTGAC-3'). Cycling conditions were 50°C for 2 min and 95°C for 10 min followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Data are presented as relative mRNA levels calculated by the equation: $2^{-\Delta Ct}$ (ΔCt = Ct of *COMT* - Ct of *GAPDH*, where Ct is cycle threshold).

Statistical analysis

Hardy-Weinberg Equilibrium was calculated according to standard procedures using χ^2 analysis. For each allele, subjects were grouped according to genotype by allele copy number (0, 1, and 2, corresponding to noncarriers, heterozygote carriers, and homozygote carriers, respectively). We allowed for three possible genetic models to explain differences between groups, *i.e.* an allele-dose effect, a dominant effect, or a recessive effect. Allele dose was defined as the number of copies of a certain allele in the genotype. In case of a consistent trend reflected as an allele-dose effect, we performed a linear regression analysis to quantify the association. In case of a dominant or a recessive effect of the test allele, ANOVA or analysis of covariance was performed to test for differences between two genotype groups. For dominant alleles,

we compared test-allele carriers *vs.* noncarriers.

Odds ratios (ORs) and hazard ratios (HRs) with 95% confidence intervals (95%CI) were calculated by logistic regression analyses to estimate the relative risk of fractures by genotype for the risk allele, with no copies of the risk allele as the reference group. We first calculated crude odds ratios and hazard ratios and then adjusted for potentially confounding factors (e.g., age, height, weight, and BMD). To estimate nonvertebral fracture risk by genotype, we used Cox proportional hazard models, thereby taking potential differences in time-to-event into account. To estimate the risk of vertebral fractures, ORs with 95%CI were calculated using logistic regression models. We were not able to use Cox proportional hazard models for vertebral fractures because the exact time of the event was not known. All statistical analyses were performed using SPSS v11.0.1 (SPSS Inc., Chicago, IL). P-values are two-sided, and $P \leq 0.05$ was considered significant.

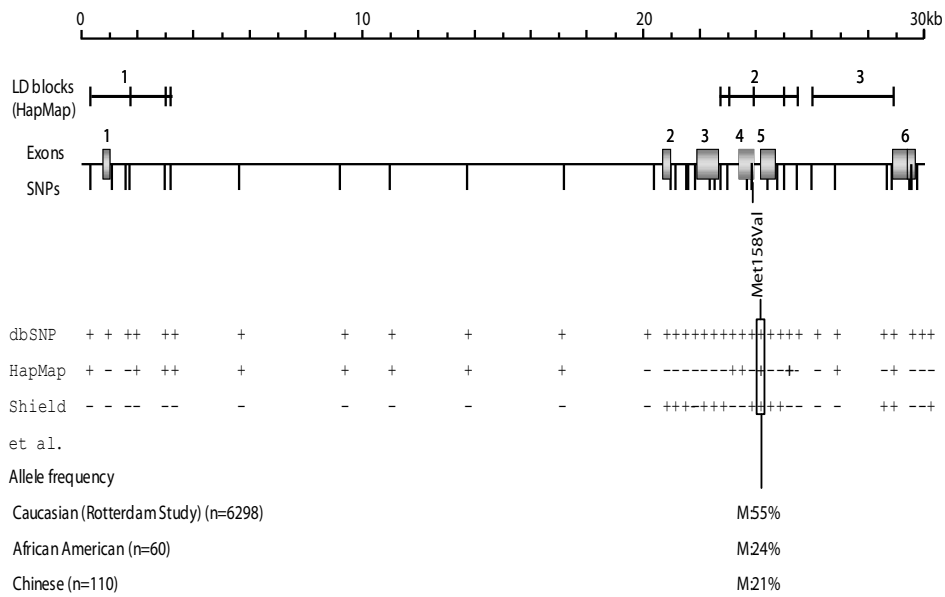


Figure 2. Schematic overview of the *COMT* gene showing the methionine to valine substitution at codon 158. In grey are the six exons (numbered 1–6), and the vertical lines represent SNPs found in this gene. We searched the NCBI database (http://www.ncbi.nlm.nih.-gov/), the HapMap database (Phase II release 19) (http://www.hapmap.org), and a study by Shield et al.[23] in which the COMT gene was resequenced to find SNPs. + indicates that the SNP is present in one of the data resources; – indicates not present. In the lower part of the figure, the allele frequencies of the low activity methionine allele are shown for three different ethnic populations (the Rotterdam Study (Caucasian), African-American, and Han-Chinese).

RESULTS

Genotyping

The allele frequency of the A-allele (methionine, $COMT^{L}$) was 55% in our study population, the genotype distribution was in Hardy-Weinberg equilibrium (P=0.99). Figure 2 shows the *COMT* gene with the exons numbered 1–6 and indicates SNPs found in several data sources: dbSNP, HapMap, and a resequencing study[23]. The methionine to valine substitution is located in exon 4 and was found in both the dbSNP database, the HapMap genome browser (Phase II release 19), and the resequencing study[23]. We analyzed the LD-block structure from publicly available HapMap data across the gene (Figure 2). The Met158Val variant is situated in haploblock 2 in the gene, encompassing intron 3 to intron 5. Figure 2 also shows the allele frequency of the Met158 variant in three different ethnic groups. The allele frequency of the Met allele in African-Americans (24%) and in Han-Chinese (21%) is lower compared with the frequency in Caucasians in our population (55%).

Baseline characteristics

None of the baseline characteristics were significantly different between the three genotype groups in men and women (Table 1). In Table 2, the bone density measures of the study population are shown. Both lumbar spine BMD and femoral neck BMD did not differ by *COMT* Val158Met genotype. No association of the *COMT* polymorphism with serum estrone and estradiol levels was found.

Fracture risks

Table 3 shows the hazard ratios (HR) and 95% CI for osteoporotic and fragility fractures, and the odds ratios (OR) and 95% CI for vertebral fractures for both men

Table 1. Baseline anthropometric measures according to COMT Met158Val genotype.

		MEN				Women			
		$COMT^{LL}$	$COMT^{HL}$	$COMT^{HH}$	P	$COMT^{LL}$	$COMT^{HL}$	$COMT^{HH}$	P
N		751	1,240	524		1,061	1,761	732	
Age (years)		68.1 (8.0)	68.1 (8.2)	67.8 (7.9)	0.78[a]	69.8 (9.2)	69.9 (9.1)	69.3 (9.5)	0.35[a]
Height (cm)	Crude	175.0 (6.8)	174.5 (6.7)	174.7 (7.2)	0.28[a]	161.4 (6.8)	161.3 (6.6)	161.2 (6.7)	0.56[b]
	Adjusted[1]	175.1 (6.5)	174.6 (6.5)	174.6 (6.5)	0.23[a]	161.4 (6.2)	161.3 (6.2)	161.1 (6.2)	0.28[b]
Weight (kg)	Crude	78.3 (10.3)	78.5 (10.9)	78.5 (10.6)	0.96[a]	69.6 (10.8)	69.6 (11.5)	69.3 (11.7)	0.86[a]
	Adjusted[1]	78.4 (10.4)	78.5 (10.4)	78.4 (10.3)	0.96[a]	69.6 (11.2)	69.6 (11.2)	69.3 (11.2)	0.75[a]
BMI (kg/m^2)	Crude	25.6 (3.0)	25.7 (3.0)	25.7 (2.8)	0.45[a]	26.7 (4.0)	26.8 (4.2)	26.7 (4.1)	0.86[a]
	Adjusted[1]	25.6 (3.0)	25.7 (3.0)	25.7 (3.0)	0.46[a]	26.7 (4.1)	26.8 (4.1)	26.7 (4.1)	0.89[a]
Ever smoked (%)		91.6	91.4	93.8	0.24[a]	44.8	47.5	46.0	0.46[a]

[1] values are adjusted for age; [a] p-values were calculated using AN(C)OVA analysis; [b] p-values were calculated using linear regression analysis; values are mean (SD).

Table 2. Bone characteristics according to COMT Met158Val genotype.

		MEN				Women			
		$COMT^{LL}$	**$COMT^{HL}$**	**$COMT^{HH}$**	**P**	**$COMT^{LL}$**	**$COMT^{HL}$**	**$COMT^{HH}$**	**P**
lsBMD (g/cm^2)	Crude	1.16 (0.20)	1.17 (0.20)	1.16 (0.18)	0.62	1.04 (0.18)	1.04 (0.18)	1.03 (0.18)	0.51
	Adjusted[1]	1.16 (0.18)	1.17 (0.20)	1.16 (0.19)	0.74	1.04 (0.18)	1.04 (0.15)	1.03 (0.17)	0.47
fnBMD (g/cm^2)	Crude	0.92 (0.14)	0.92 (0.13)	0.91 (0.14)	0.65	0.83 (0.13)	0.83 (0.14)	0.83 (0.13)	0.97
	Adjusted[1]	0.92 (0.13)	0.92 (0.13)	0.91 (0.13)	0.62	0.83 (0.12)	0.83 (0.11)	0.82 (0.12)	0.28
Bone loss (%/year)	Crude	-0.46 (0.94)	-0.45 (0.87)	-0.40 (0.94)	0.49	-0.84 (1.00)	-0.85 (1.14)	-0.84 (1.07)	0.76
	Adjusted[1]	-0.45 (0.91)	-0.45 (0.91)	-0.40 (0.90)	0.70	-0.84 (1.08)	-0.85 (1.08)	-0.80 (1.08)	0.82

[1] values are adjusted for age, height and weight; values are mean (SD); lsBMD: lumbar spine Bone Mineral Density; fnBMD: femoral neck Bone Mineral Density.

and women. The *COMT* genotype was associated with osteoporotic fractures and fragility fractures. In men, the fracture risk for $COMT^{LL}$ (HR=1.6, 95%CI[1.0 –2.5]; P=0.05) and $COMT^{HL}$ (HR=1.6, 95%CI[1.0 –2.4]; P=0.04) was similar, suggesting a dominant effect, and therefore, we analyzed the fractures risks also in carriers *vs.* noncarriers of the $COMT^{L}$ allele. The risk for osteoporotic fractures in male $COMT^{L}$-carriers was 1.6 (HR=1.6, 95%CI[1.0 –2.4]; P=0.02) compared with noncarriers. For fragility fractures, the risk was 2.7 (HR=2.7, 95%CI[1.3–5.9]; P=0.005). Both HRs are adjusted for age, height, and weight. In women, a higher osteoporotic fracture risk was seen for the $COMT^{LL}$ genotype (HR=1.2, 95%CI[1.0 –1.5]; P=0.04). No significant effect of the $COMT^{L}$ allele was seen on fragility fractures in women.

We examined whether estradiol-levels influenced the association between the polymorphism and fracture risk. We did not find any influence, probably due to low

Table 3. Fracture risk according to COMT Met158Val genotype

	Men Nr fractures/total nr (%)		HR / OR [95%CI]	HR / OR [95%CI]	**Women** Nr fractures/total nr (%)		HR / OR [95%CI]	HR / OR [95%CI]
	$COMT^{L}$-carriers	**Non-carriers**	**Crude**	**Adusted[2]**	**$COMT^{L}$-carriers**	**Non-carriers**	**Crude**	**Adusted[1]**
Incident osteoporotic fractures	165/1991 (8.3)	27/524 (5.2)	1.62 [1.1-2.4]	1.56 [1.0-2.4]	568/2822 (20.1)	129/732 (17.6)	1.18 [1.0-1.4]	1.17 [1.0-1.4]
Incident fragility fractures	80/1991 (4.0)	7/524 (1.3)	3.0 [1.4-6.6]	2.72 [1.3-5.9]	234/2822 (8.3)	54/732 (7.4)	1.15 [0.9-1.6]	1.17 [0.9-1.6]
Prevalent vertebral fractures	77/1089 (7.1)	21/300 (7.0)	1.01 [0.6-1.7]	1.02 [0.6-1.7]	109/1401 (7.8)	23/356 (6.5)	1.22 [0.8-2.0]	1.14 [0.7-1.8]

[1]values are adjusted for age, height and weight; OR = odds ratios, OR are presented with 95% confidence intervals; HR: Hazard ratio, HR are presented with 95% confidence intervals

power. The *COMT* polymorphism did not show an association with vertebral fractures in either men or women.

Expression of COMT in osteoblasts

Figure 3 shows mRNA expression levels of *COMT* mRNA relative to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA in three human bone cell lines and in the liver cell line HepG2. We observed a *GAPDH* Ct value of 16 in all three osteoblastic cell lines and the HepG2 cell line. The Ct value for *COMT* was around 20 for all three osteoblast-like cell lines and 19 for the HepG2 cell line. For SV-HFO cells, the relative mRNA expression level was 6.9% (SD=2.7), for MG-63 it was 3.5% (SD=0.6), for SaOS it was 4.3% (SD=0.3), and for HepG2 it was 9.3% (SD=1.4) compared with *GAPDH*. For SV-HFO, there was no significant difference in relative *COMT* mRNA expression compared with the expression in HepG2. For MG-63 and SaOS, the relative *COMT* mRNA expression was two times lower compared with HepG2 *COMT* mRNA expression.

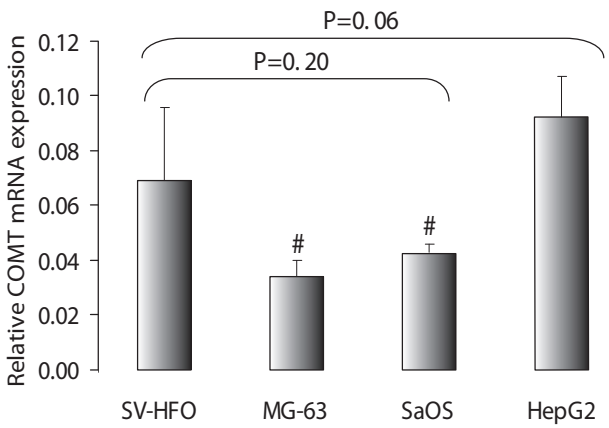


Figure 3. *COMT* mRNA expression relative to GAPDH mRNA expression in three osteoblast-like cell-lines and a HepG2 cell line. Results of two separate experiments are shown. n=4 measurements for all cell lines. Mean relative mRNA expression for SV-HFO was 0.069 (SD=0.027); for MG-63, the relative COMT mRNA expression was 0.034 (SD=0.006); for SaOS, it was 0.043 (SD=0.003); and for HepG2, it was 0.093 (SD=0.014). #, Significant (P<0.05) compared with HepG2 COMT mRNA expression.

DISCUSSION

In this study, we investigated whether the *COMT* Val158Met polymorphism is associated with parameters of bone health in a large prospective population-based cohort study, the Rotterdam Study. We found the Met158 variant to be associated with higher risk for osteoporotic and fragility fractures in elderly men.

In line with a previous study[17], we found no association with lumbar spine BMD and femoral neck BMD in postmenopausal women. Neither was there an association of the polymorphism with BMD measures in elderly men. This is not in concordance with a previous study reporting an association of the $COMT^{L}$-allele with lower femoral neck BMD in young adult Caucasian men (mean age 19 years)[16]. However, in our study population, the mean age for men was 68 years. Because 19-year-old males are not completely full-grown, and their BMD has not yet reached peak levels, the activity of COMT might influence bone mass during accrual, but this influence might stop once peak BMD is reached.

Interestingly, we observed that the $COMT^{L}$-allele increased fracture risk in elderly men, and with a similar trend for women. Male carriers of the $COMT^{L}$-allele had a 60% higher risk for osteoporotic fractures, whereas the risk for fragility fractures was even higher, 170%; however, confidence limits for this HR were wide, so the true effect of this polymorphism on fragility fractures is not clear yet. To be able to establish the true effect of the $COMT^{L}$-allele on fragility fractures, you need a larger study population with data on fragility fractures. The risk for vertebral fractures was not higher in men, suggesting a possible role for COMT in the strength of the appendicular skeleton more than in the axial skeleton. In women, there was a higher osteoporotic fracture risk in $COMT^{LL}$ genotype carriers, but there was no association with fragility fractures in women. It was shown previously that the activity of COMT in liver and red blood cells is 30% higher in men than in women[24]. It could be that women, due to their already lower COMT activity, have developed an alternative pathway to compensate for this. In that case, the effect of the $COMT^{L}$-allele would be less dramatic in women than in men, which could explain the weaker and nonsignificant trend seen in women.

In line with a study by Worda et al. of 159 postmenopausal women[15], we did not observe significant differences in serum estradiol levels between the three genotype groups ($COMT^{HH}$, $COMT^{HL}$, and $COMT^{LL}$) in women. We also found no significant association with estradiol levels in men, which did not correspond with a recently published study by Eriksson et al., who found increased serum estradiol levels in men with the $COMT^{LL}$ genotype[14]. Our results indicate that COMT has no effect on

systemic estradiol levels, which suggests that the COMT variant has more of a local effect than a systemic effect.

The fact that *COMT* mRNA is expressed in three different osteoblastic cell lines indicates that normally COMT is expressed in bone cells *in vivo*. We also found the expression levels in the osteoblastic cell lines to be quite similar to expression levels in HepG2 liver cells, where COMT is biologically active[25]. This raises the possibility that the *COMT* variants influence estrogen levels locally, without affecting systemic levels.

The molecular mechanism by which lower activity of COMT could influence fracture risk is not yet known. A possible explanation might be that the estrogen metabolite 2-hydroxyestrone (2-OHE1), a catechol estrogen, which has already been shown to possess antiestrogenic activity in MCF-7 human breast cancer cells, also plays a role in bone. Binding of 2-OHE1 to the estrogen receptor α might prevent circulating estrogens from binding and interacting with this receptor[26]. Lower activity of the $COMT^L$ variant would lead to higher concentrations of 2-OHE1 in target tissue, thereby inhibiting the function of estradiol in these tissues. In such a situation, one would not expect higher estrogen levels in both men and women carrying the $COMT^L$-allele.

We show an association of the $COMT^L$-allele with an increased risk for fractures in men but not with BMD. Such a genotype effect on fracture risk independent of BMD has already been seen for the *ESR1 Xba*I polymorphism in the GENOMOS Study, a large prospective meta-analysis of almost 20,000 subjects[27] and also for polymorphisms in the vitamin D receptor and collagen 1A1[28] genes. This observation suggests that COMT is involved in bone metabolic pathways other than those reflected in BMD but that still lead to higher fracture risk. It is known that other physiological factors, such as bone size[29], microarchitecture (bone structure)[30], and bone quality[31] can lead to an increased risk for fractures. On the basis of our data, we conclude that COMT does not have an effect directly on BMD but that it can affect one or more of these other factors, thereby leading to a higher fracture risk. Furthermore, in view of the pleiotropic effects of estradiol, we have to consider effects on phenotypes other than bone. For example, the $COMT^L$-allele might affect fall frequency or muscle strength, two predictors of fractures[32].

Our study has some limitations. First, genetic association studies can be influenced by population heterogeneity. However, all subjects in our study population were of Dutch-Caucasian ancestry and have a similar ethnic background. The *COMT* genotypes were also in Hardy-Weinberg equilibrium, and genotype frequencies were similar to those found in studies of other Caucasian populations[16,33]. Our study population can therefore be considered ethnically homogeneous and representative

of the Dutch population. The second possible limitation is that we studied this association in only one Caucasian study sample. However, our study population is large, and we looked at well-defined endpoints. The third limitation might be that we examined a single SNP instead of haplotypes, whereas it has been suggested that studying haplotypes can increase power to detect rare causal alleles[34]. Nevertheless, we studied a polymorphism with well-established functional effects. Another limitation of genetic association studies are false-positive findings caused by multiple testing. Although we tested multiple endpoints, not all are independent. We have three main endpoints: BMD, fracture risk, and estradiol levels. Nevertheless, replication of our findings is needed.

In summary, this population-based study provides evidence that the $COMT^{L}$–allele is associated with an increased risk for osteoporotic and fragility fractures in elderly men. There was no association of this allele with bone parameters in women. The biological mechanism behind this relation remains to be elucidated.

ACKNOWLEDGEMENTS

We are very grateful to the participants of the Rotterdam study and to the DXA and radiograph technicians, L. Buist and H. W. M. Mathot. Furthermore, we acknowledge all participating general practitioners and the many field workers in the research center in Ommoord, Rotterdam, The Netherlands.
This study is supported by The Netherlands Organization of Scientific Research-Research Institute for Diseases in the Elderly (Grant 014-93-015; RIDE2) and the European Commission [Grants QLK6-CT-2002-02629 (GENOMOS) and QLK6–2002-00491 (NEMO)].

WEBRESOURCES

URLs for data presented herein are as follows: dbSNP, http://www.ncbi.nlm.nih.gov/SNP/; HapMap data, http://hapmap.jst.go.jp/cgi-perl/gbrowse/hapmap/.

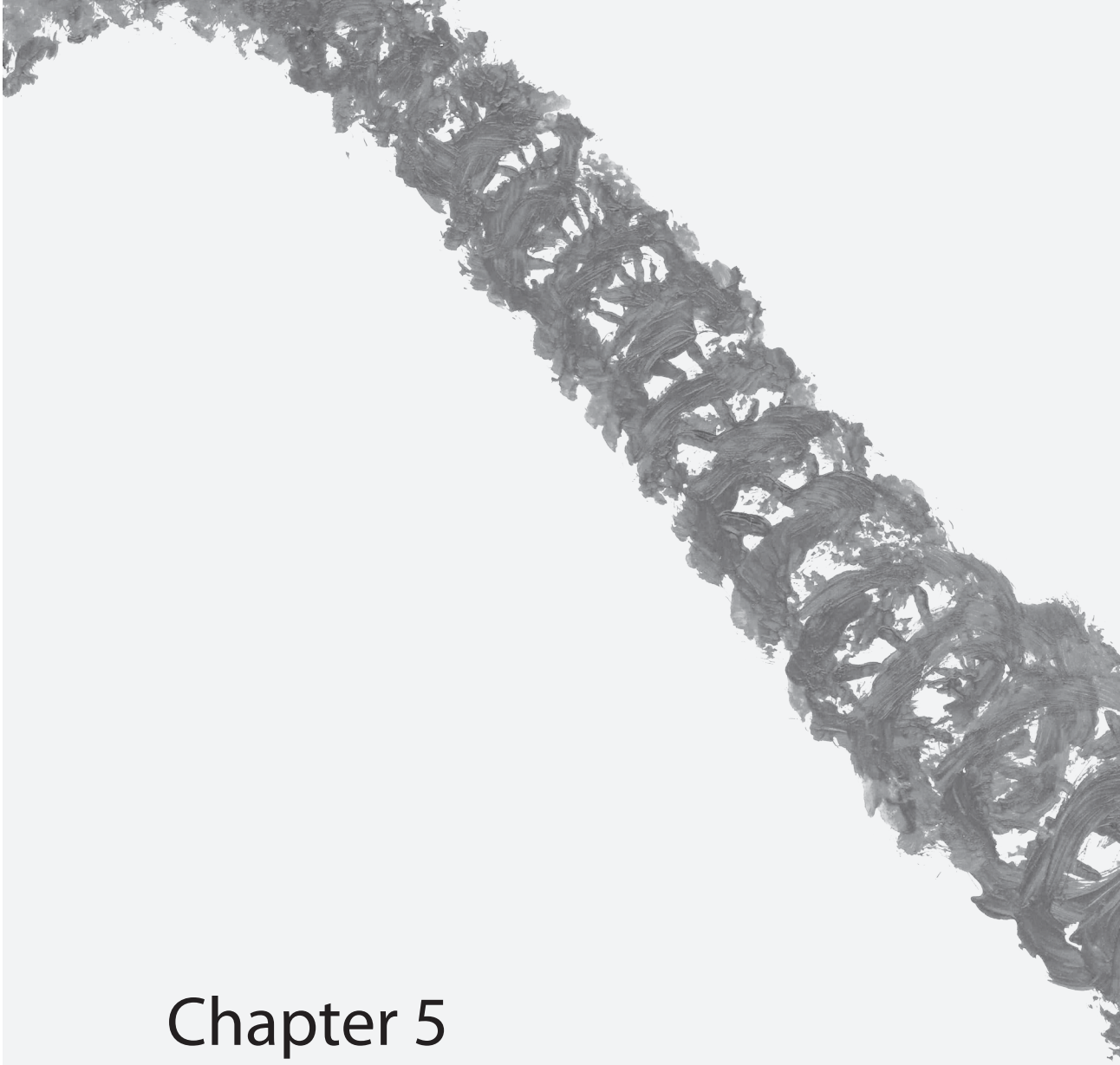
REFERENCES

1. Cooper, C. Epidemiology of osteoporosis. Osteoporos Int 9 Suppl 2, S2-8 (1999).
2. Melton, L.J., 3rd. Hip fractures: a worldwide problem today and tomorrow. Bone 14 Suppl 1, S1-8 (1993).
3. Ralston, S.H. Genetic determinants of susceptibility to osteoporosis. Current Opinion in Pharmacology 3, 286-290 (2003).
4. Yager, J.D. & Liehr, J.G. Molecular mechanisms of estrogen carcinogenesis. Annu Rev Pharmacol Toxicol 36, 203-32 (1996).
5. Zhu, B.T. & Conney, A.H. Functional role of estrogen metabolism in target cells: review and perspectives. Carcinogenesis 19, 1-27 (1998).
6. Riggs, B.L., Khosla, S. & Melton, L.J., 3rd. Sex steroids and the construction and conservation of the adult skeleton. Endocr Rev 23, 279-302 (2002).
7. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms and risk of myocardial infarction. Jama 291, 2969-77 (2004).
8. Schuit, S.C. et al. Height in pre- and postmenopausal women is influenced by estrogen receptor alpha gene polymorphisms. J Clin Endocrinol Metab 89, 303-9 (2004).
9. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. Hum Mol Genet 12, 1745-54 (2003).
10. Gennari, L. et al. A polymorphic CYP19 TTTA repeat influences aromatase activity and estrogen levels in elderly men: effects on bone metabolism. J Clin Endocrinol Metab 89, 2803-10 (2004).
11. Napoli, N. et al. Effect of CYP1A1 Gene Polymorphisms on Estrogen Metabolism and Bone Density. J Bone Miner Res 20, 232-9 (2005).
12. Scanlon, P.D., Raymond, F.A. & Weinshilboum, R.M. Catechol-O-methyltransferase: thermolabile enzyme in erythrocytes of subjects homozygous for allele for low activity. Science 203, 63-5 (1979).
13. Chen, J. et al. Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. Am J Hum Genet 75, 807-21 (2004).
14. Eriksson, A.L. et al. Association between the low activity genotype of catechol-O-methyltransferase and myocardial infarction in a hypertensive population. Eur Heart J 25, 386-91 (2004).
15. Worda, C. et al. Influence of the catechol-O-methyltransferase (COMT) codon 158 polymorphism on estrogen levels in women. Hum Reprod 18, 262-6 (2003).
16. Lorentzon, M., Eriksson, A.L., Mellstrom, D. & Ohlsson, C. The COMT val158met Polymorphism Is Associated With Peak BMD in Men. J Bone Miner Res 19, 2005-11 (2004).
17. Tofteng, C.L. et al. Two single nucleotide polymorphisms in the CYP17 and COMT Genes--relation to bone mass and longitudinal bone changes in postmenopausal women with or without hormone replacement therapy. The Danish Osteoporosis Prevention Study. Calcif Tissue Int 75, 123-32 (2004).
18. 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).
19. Burger, H. et al. The association between age and bone mineral density in men and women aged 55 years and over: the Rotterdam Study. Bone Miner 25, 1-13 (1994).
20. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with

estradiol levels in postmenopausal women. Eur J Endocrinol 153, 327-34 (2005).
21. Van der Klift, M., De Laet, C.E., McCloskey, E.V., Hofman, A. & Pols, H.A. The incidence of vertebral fractures in men and women: the Rotterdam Study. J Bone Miner Res 17, 1051-6 (2002).
22. Eijken, M. et al. 11{beta}-Hydroxysteroid Dehydrogenase Expression and Glucocorticoid Synthesis Are Directed by a Molecular Switch during Osteoblast Differentiation. Mol Endocrinol 19, 621-631 (2005).
23. Shield, A.J., Thomae, B.A., Eckloff, B.W., Wieben, E.D. & Weinshilboum, R.M. Human catechol O-methyltransferase genetic variation: gene resequencing and functional characterization of variant allozymes. Mol Psychiatry 9, 151-60 (2004).
24. Boudikova, B., Szumlanski, C., Maidak, B. & Weinshilboum, R. Human liver catechol-O-methyltransferase pharmacogenetics. Clin Pharmacol Ther 48, 381-9 (1990).
25. Bertocci, B. et al. Human catechol-O-methyltransferase: cloning and expression of the membrane-associated form. Proc Natl Acad Sci U S A 88, 1416-20 (1991).
26. Schneider, J., Huh, M.M., Bradlow, H.L. & Fishman, J. Antiestrogen action of 2-hydroxyestrone on MCF-7 human breast cancer cells. J Biol Chem 259, 4840-5 (1984).
27. Ioannidis, J.P. et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. Jama 292, 2105-14 (2004).
28. Uitterlinden, A.G. et al. Interaction between the vitamin D receptor gene and collagen type I alpha1 gene in susceptibility for fracture. J Bone Miner Res 16, 379-85 (2001).
29. Michelotti, J. & Clark, J. Femoral neck length and hip fracture risk. J Bone Miner Res 14, 1714-20 (1999).
30. Oleksik, A. et al. Bone structure in patients with low bone mineral density with or without vertebral fractures. J Bone Miner Res 15, 1368-75 (2000).
31. Felsenberg, D. & Boonen, S. The bone quality framework: determinants of bone strength and their interrelationships, and implications for osteoporosis management. Clin Ther 27, 1-11 (2005).
32. Tinetti, M.E., Speechley, M. & Ginter, S.F. Risk factors for falls among elderly persons living in the community. N Engl J Med 319, 1701-7 (1988).
33. Palmatier, M.A., Kang, A.M. & Kidd, K.K. Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. Biol Psychiatry 46, 557-67 (1999).
34. de Bakker, P.I. et al. Efficiency and power in genetic association studies. Nat Genet 37, 1217-1223 (2005).

Part B

Candidate Pathway Studies



Chapter 5

Genetic variation of genes in the estrogen pathway and associations with hormone levels, age at natural menopause and bone

Lisette Stolk, Joyce BJ van Meurs, Fernando Rivadeneira, Albert Hofman, Huibert AP Pols, Joop SE Laven, Frank H de Jong, André G Uitterlinden

Manuscript in preparation

ABSTRACT

After menopause women have an increased risk for several diseases, such as osteoporosis and cardiovascular diseases. This increased risk is at least partly thought to be due to the decreased sex steroid levels, like estradiol (E2) and testosterone (T). The factors contributing to the inter-individual variations in plasma sex hormone levels, age at natural menopause and osteoporosis are largely unknown. However, it is known that both environmental and genetic factors play a role. In this study we focused on all common genetic variation in 61 genes of the estrogen pathway. We studied association between these variants and sex steroid levels (E2, and T), sex hormone binding globulin (SHBG), age at natural menopause and two osteoporosis-related end points: lumbar spine bone mineral density (lsBMD) and vertebral fracture (vert fx) risk. We additionally searched for SNPs associated with higher E2 or T levels and higher lsBMD, lower SHBG and higher lsBMD, earlier menopause and lower lsBMD in women and lower lsBMD and higher vert fx risk or vice versa.

In total we analyzed 14,816 SNPs in 61 genes for association with the estrogen-related traits. We applied a multiple testing threshold p-value of $4x10^{-5}$, which was based on the total number of independent SNPs in these genes. For the consistency analysis of two traits we applied a p-value threshold of 0.01 for menopausal age and lsBMD, while we applied a threshold for hormone levels and vert fx risk of 0.05. This was based on the number of samples we had available for the different traits.

We found SNPs in the 5' region of the SHBG gene to be associated with plasma SHBG levels in men (effect: +0.46 nmol/L) and women (effect: +0.29 nmol/L). We did not observe significant associations for plasma estradiol or testosterone levels in this study. The analysis of age at natural menopause showed no polymorphisms with a significant p-value. We observed one SNP in the *HSD17B5* gene region associated with lsBMD in women, but found no associations in men nor did we observe significant associations with vertebral fracture risk.

The analysis for consistently associated SNPs for menopausal age and lsBMD, E2-levels and BMD in women and SHBG-levels and BMD in both men and women, showed no association. In women we found an association for one SNP with T-levels and lsBMD (*NSD1*), and in men we found 4 SNPs in one gene associated with T-levels and lsBMD (*ESRRG*). The consistency analysis of E2-levels and lsBMD in men yielded one significant SNP in the *ESR1* gene. When we searched for SNPs consistently associated with lsBMD and vertebral fracture risk, we observed 7 SNPs located in 3 genes associated in women (*POR, HSD17B5, ESRRA*), and 34 SNPs situated in four different

genes in men (*LHCGR, ESR1, HSD17B3, NCOR1*).

In conclusion, we performed a candidate pathway analysis for several estrogen-related traits, and we observed polymorphisms 5' of the *SHBG* gene to be associated with plasma SHBG levels in men and women. One SNP in *HSD17B5* was associated with lsBMD in women. We did not find significant associations with age at natural menopause or vert fx risk. Our results in the consistency analyses indicated gender-specific associations for genetic variations in the estrogen pathway with these endpoints. Our results require further replication in other cohorts, but also studies to identify the functional variants underlying the associations.

INTRODUCTION

The sex steroid hormones estradiol (E2), and testosterone (T) influence a number of target tissues. The availability of E2 and T to the target tissue is mediated by sex hormone binding globulin (SHBG), which binds the hormones and transports them to the cells. SHBG is produced in the liver, while the sex steroids are produced by the gonads. Levels of sex steroids are influenced by both environmental (e.g., diet, and smoking) and genetic factors[1]. Genetic variants in the genes involved in steroid synthesis, metabolism and signalling are the most likely candidates to explain the inter-individual differences in sex hormone levels. Most of the previous studies identifying genetic factors determining sex steroid levels focused on one candidate gene. We previously showed that polymorphisms in the estrogen receptor alpha gene (*ESR1*) are associated with inter-individual differences in estradiol levels in postmenopausal women[2]. Other studies showed associations of aromatase (*CYP19A1*) polymorphisms with estradiol levels in postmenopausal women[3,4]. Recently, a study examined all genetic variation in 55 sex steroid related genes and found that a polymorphism (rs2470152) in the *CYP19A1* gene was associated with differences up to 13% in estradiol (E2) levels and 15% in estrone (E1) levels in men[5].

During aging sex steroid levels decline, in men there is a gradual fading of these levels with aging. In women there is a marked decline in sex hormone levels after menopause, and the end of estradiol production by the ovaries. The average age at menopause is 50-51 years, with a range between 40-60 years of age[6], and twin studies showed this variability to be genetically determined with estimated heritabilities of 44-65%[7-9]. Such genetic factors might regulate the size of the follicle pool and the rate of its depletion, and their identification could have biological and clinical applications. The exact genetic factors involved in the timing of menopause are not yet known. The substantial drop in estrogens after menopause makes the estrogen synthesis, metabolism and signalling pathways prime candidates to study for their involvement in the timing of menopause. Several genes have already been studied in this respect, with the estrogen receptor alpha (*ESR1*) gene being one of the most studied. We and several others examined polymorphisms in *ESR1*[10-12], but found conflicting results.

After menopause, women have an increased risk for several diseases, like osteoporosis and cardiovascular disease[13]. This increased risk is at least partly thought to be due to the decreased sex steroid levels. In a number of studies it was shown that decreased levels of E2[14-16], and T[15,17] are associated with an increased risk for osteoporosis, while decreased levels of SHBG are associated with a lower risk for

osteoporosis[16]. Osteoporosis is a skeletal disorder characterized by low bone mineral density (BMD) and increased fracture (fx) risk[18]. Twin and family studies have estimated a high heritability for BMD (60%-80%)[19-21]. Because E2, T, and SHBG levels are associated with osteoporosis, genes regulating sex steroid production and metabolism are good candidates to study for involvement in the development of osteoporosis. It is known that estrogens have more influence on trabecular bone compared to cortical bone[22]. The vertebrae are composed of trabecular bone, and therefore, lumbar spine BMD (lsBMD) and vertebral fractures (vert fx) are good phenotypes to study for the association of genetic variation in estrogen-related genes. *ESR1* gene polymorphisms have been analysed extensively for association with BMD, bone loss, turnover markers and/or fractures in women, with conflicting results[23]. A recent meta-analysis of these polymorphisms in approximately 18,000 individuals from eight European centres found evidence of association with fracture risk but not BMD in women[24]. A few studies reported association of *ESR1* genotypes with BMD in men[25], but most studies found no association[26,27].

In the current study we studied the association between all common genetic variation in 61 genes of the estrogen pathway and sex steroid levels (T, E2), SHBG, age at natural menopause and two osteoprosis-related traits: lsBMD and vert fx risk.

METHODS

Study population

Subjects were participants of the Rotterdam Study, a large prospective population-based cohort study of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands[28]. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the elderly. Rationale and design have been described previously[28]. Among the subjects living independently, the overall response rate was 78 percent for home interview and 71 percent for examination in the research centre, where anthropometric characteristics and bone mineral density were measured, and blood samples were taken. The Rotterdam Study was approved by the medical ethics committee of the Erasmus University Medical School, and written informed consent was obtained from each subject. The current study is based on men and women of the baseline cohort of the Rotterdam Study (RSI) and the first extension study, RSII.

Selection of candidate genes

Candidate genes were selected from the NCBI Gene database using ESTROGEN or ANDROGEN as keywords. Only genes involved in synthesis, metabolism or signalling were included in the final list (Table 1). All estrogen or androgen target genes were excluded. A few more genes were added to the list after review by an expert in the field (prof. FH de Jong). The final list consisted of 64 genes located on autosomes and the X-chromosome. The gene location was defined as the chromosomal location of the largest mRNA transcript in the HapMap PhaseII rel 22 database plus 100kb 5' and 3' of this transcript. For this study only genes and SNPs on the autosomes were included, because there were no imputed genotype data available for the X chromosome, giving in total 14,816 common SNPs in 61 genes.

Genotyping and imputation

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. Genotyping using the Illumina HumanHap 550K beadarray was performed as described previously[29]. Imputation of the genome-wide genotyping data on all autosomes was performed using MACH. Of the in total 14,816 SNPs in the 61 genes 436 did not pass the quality thresholds for imputation quality score (>0.5) or minor allele frequency (>0.01) and were excluded from the analysis, giving 14,380 SNPs to be included in this study.

Assessment of sex hormone levels

Non-fasting blood samples were drawn by venapuncture at the baseline examination in the research center. Levels of steroid hormones were measured in plasma, for which blood was collected in 5-ml tubes containing 0.5 ml sodium citrate solution. All tubes were stored on ice before and after blood sampling. Platelet-free plasma was obtained by two-stage centrifugation, first for 10 min at 1600 x g at 4°C and then for 30 min at 7000 x g at 4°C. The samples were immediately frozen in liquid nitrogen and transferred to the laboratory. At the laboratory, plasma samples were stored at -80°C until hormone measurements. For the purpose of the present study, plasma levels of E2, testosterone, and sex hormone binding globulin (SHBG) were estimated in 12 separate batches of samples using coated tube or double antibody radioimmunoassays purchased from Diagnostic Systems Laboratories, Inc. (Webster, TX). Due to the relatively small volumes of plasma available, all values reported are single sample estimations. Intra-assay coefficients of variation, determined on the basis of duplicate results of internal quality control (QC) serum pools with 3 different levels

Table 1. Candidate genes involved in estrogen synthesis, metabolism and signalling by chromosome position.

Gene	Gene name	Chr	Start-100kb	Stop+100kb
HSD3B2	3 beta-hydroxysteroid dehydrogenase type II	1p13.1	119,659,296	119,867,172
DNTTIP2 (ERBP)	Estrogen receptor binding protein	1p22.1	94,007,926	94,217,330
PRDM2 (RIZ1)	Retinoblastoma-interacting sinc finger domain 1	1p36.21	13,803,937	14,124,162
HSD17B7	17 beta-hydroxysteroid dehydrogenase type VII	1q23	160,927,120	161,149,231
ESRRG	Estrogen-related receptor gamma	1q41	214,643,211	215,429,599
INHBB	Inhibin, beta B	2cen-q13	120,719,469	120,925,444
CYP1B1	Cytochrome P450, family 1, subfamily B, polypeptide 1	2p21	38,048,250	38,256,796
LHCGR	Luteinizing hormone/choriogonadotropin receptor	2p21	48,667,425	48,936,367
FSHR	Follicle Stimulating Hormone Receptor	2p21-p16	48,943,157	49,335,134
NCOA1	Nuclear receptor coactivator 1	2p23	24,560,850	24,947,075
SRD5A2	Steroid-5-alpha-reductase, alpha polypeptide 2	2p23	31,503,160	31,759,544
INHA	Inhibin, alpha	2q33-q36	220,045,198	220,248,671
UGT1A1	UDP glucuronosyltransferase 1 family, polypeptide A1	2q37	234,233,658	234,446,684
UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15	4q13	69,094,911	69,318,969
UGT2B7	UDP glucuronosyltransferase 2 family, polypeptide B7	4q13	69,896,814	70,113,293
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	4q13	70,280,473	70,496,205
SULT1E1	Sulfotransferase family 1E, estrogen-preferring, member 1	4q13.1	70,641,519	70,860,458
GNRHR	Gonadotropin-releasing hormone receptor	4q21.2	68,185,701	68,404,399
HSD17B11	17-beta-hydroxysteroid dehydrogenase type XI	4q22.1	88,376,789	88,631,352
SRD5A1	Steroid-5-alpha-reductase, alpha polypeptide 1	5p15	6,586,500	6,822,673
HSD17B4	17beta-estradiol dehydrogenase type IV	5q21	118,716,124	119,005,923
NSD1	Nuclear receptor binding SET domain protein 1	5q35.2-q35.3	176,393,532	176,755,367
HSD17B8	17-beta-hydroxysteroid dehydrogenase type VIII	6p21.3	33,180,397	33,382,585

Table 1. continued.

Gene	Gene name	Chr	Start-100kb	Stop+100kb
CGA	Glycoprotein hormones, alpha polypeptide	6q12-q21	87,751,941	87,961,543
NCOA7	Nuclear receptor coactivator 7	6q22.31-q22.32	126,044,000	126,393,949
ESR1	Estrogen Receptor alpha	6q25.1	151,920,379	152,566,099
INHBA	Inhibin, beta A	7p15-p13	41,595,128	41,809,231
GPR30 (GPER)	G protein-coupled estrogen receptor 1	7p22.3	992,969	1,199,977
POR	Cytochrome P450 reductase	7q11.2	75,282,356	75,554,108
CYP3A5	Cytochrome P450, family 3, subfamily A, polypeptide 5	7q21.1	98,983,754	99,215,542
CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4	7q21.1	99,092,539	99,319,743
CYP3A7	Cytochrome P450, family 3, subfamily A, polypeptide 7	7q21-q22.1	99,040,597	99,270,757
GNRH1	Gonadotropin-releasing hormone 1	8p21-p11.2	25,232,693	25,437,836
HSD17B3	17-beta-hydroxysteroid dehydrogenase type III	9q22	97,937,410	98,204,255
HSD17B5 (AKR1C3)	17-beta-hydroxysteroid dehydrogenase type V	10p15-p14	5,026,568	5,239,876
CYP17A1	Cytochrome P450, family 17, subfamily A, polypeptide 1	10q24.3	104,480,280	104,687,280
HSD17B12	17-beta-hydroxysteroid dehydrogenase type XII	11p11.2	43,558,881	43,934,736
FSHB	Follicle stimulating hormone, beta polypeptide	11p13	30,109,139	30,313,399
ESRRA	Estrogen-related receptor alpha	11q13	63,729,620	63,940,786
FDX1	Ferredoxin 1	11q22	109,705,804	109,940,814
SLCO1B1 (OATP1B1)	Solute carrier organic anion transporter family, member 1B1	12p	21,075,403	21,383,995
HSD17B6	17-beta-hydroxysteroid dehydrogenase type VI	12q13	55,343,375	55,567,841
ESR2	Estrogen Receptor beta	14q23.2	63,663,506	63,975,021
CYP19A1	Cytochrome P450, family 19, subfamily A, polypeptide 1	15q21.1	49,188,964	49,518,086
CYP1A1	Cytochrome P450, family 1, subfamily A, polypeptide 1	15q22-q24	72,698,943	72,904,929
CYP1A2	Cytochrome P450, family 1, subfamily A, polypeptide 2	15q22-q24	72,728,236	72,935,993

Table 1. Continued

Gene	Gene name	Chr	Start-100kb	Stop+100kb
CYP11A1	Cytochrome P450, family 11, subfamily A, polypeptide 1	15q23-q24	72,317,157	72,547,020
HSD17B2	17-beta-hydroxysteroid dehydrogenase type II	16q24.1-q24.2	80,526,364	80,789,638
NCOR1	Nuclear receptor co-repressor 1	17p11.2	15,775,984	16,159,570
PELP1	Proline, glutamate and leucine rich protein 1	17p13.2	4,421,483	4,654,155
SHBG	Sex Hormone Binding Globuline	17p13-p12	7,374,216	7,577,393
HSD17B1	17-beta-hydroxysteroid dehydrogenase type I	17q11-q21	37,857,510	38,060,757
FDXR	Ferredoxin reductase	17q24-q25	70,270,215	70,480,751
CYB5	Cytochrome B 5	18q23	69,971,507	70,210,201
SULT2A1	Sulfotransferase family, cytosolic, 2A, DHEA-preferring, member 1	19q13.3	52,965,682	53,181,404
LHB	Luteinizing hormone beta polypeptide	19q13.32	54,111,050	54,312,159
GNRH2	Gonadotropin-releasing hormone 2	20p13	2,872,268	3,074,391
NCOA3	Nuclear receptor coactivator 3	20q12	45,464,064	45,819,019
NCOA5	Nuclear receptor coactivator 5	20q12-q13.12	44,023,035	44,251,987
NRIP1	Nuclear receptor interacting protein 1	21q11.2	15,155,427	15,362,718
COMT	Catechol-O-methyltransferase	22q11.21	18,209,308	18,436,527

of each analyte, were below 15% for all assays, with the exception of E2 (18%). Since inter-assay variations were relatively large (between 20% and 30%, with the exception of SHBG (14%) and testosterone (19%)) results of all batches were normalized by multiplying all concentrations within a batch with a factor, which made results for the internal QC pools comparable. This reduced inter-assay variations and was considered justified because the patterns of the results of these pools and the mean results for male and female sera in one assay batch were very similar. Assays were performed blind with respect to information on the subject. Albumin was measured using a colorimetric method (KONE Diagnostics, Espoo, Finland). Table 2 shows the number of subjects that have data on plasma hormone levels used in this study, split by gender.

Assessment of menopausal age

Age at menopause was assessed by questionnaire, as described previously[12]. Exclusion criteria were menopause due to hysterectomy, due to irradiation, after

Table 2. Number of subjects with hormone levels per hormone and stratified by gender.

	N_men	N_women
Testosterone	672	760
Estradiol	564	661
SHBG	664	738

stopping taking the pill, or due to drugs. Also women using HRT, or with menopause before the age of 40 or after 60 were excluded.

Lumbar spine BMD and vertebral fracture assessment

BMD (in grams per square centimeter) was determined by dual-energy x-ray absorptiometry (DEXA, DPX-L densitometer, Lunar, Madison, WI) at the lumbar spine (lsBMD) (vertebrae L2, L3, L4), as described elsewhere[30] in both RSI and RSII.

Vertebral fractures were only assessed in RSI. Both at baseline and at follow-up visits, between 1997 and 2001, thoracolumbar radiographs of the spine were obtained. The follow-up radiographs were available for 3,469 individuals (61% of the genotyped population, 1,971 women) who survived an average of 6.4 (SD 0.4) years after the baseline centre visit and who were still able to come to our research centre. All follow-up radiographs were scored for the presence of vertebral fracture by the McCloskey/Kanis method as described earlier[31]. If a vertebral fracture was detected, the baseline radiograph was evaluated as well. If the vertebral fracture was already present at baseline, it was considered a prevalent fracture. If it was not present at baseline, the fracture was defined to be incident.

Statistical analysis – Sex hormone levels

All analyses were carried out for men and women separately. Adjusted Z-scores for all hormone levels in men were analysed using linear regression in MACH2QTL for 14,816 SNPs. In women estradiol levels were analysed using adjusted logistic regression in MACH2DAT. For other hormone levels in women adjusted Z-scores were analysed using linear regression in MACH2QTL. All hormone levels below the detection limit of the assay were set to half of the detection limit and included in the analysis. Because of the non-normal distribution of estradiol in women, the levels were dichotomised on the median, and adjusted for age, BMI, testosterone levels, and years since menopause. In men adjusted Z-scores for E2 were analysed, these levels were adjusted for age, BMI, and testosterone levels. For testosterone and SHBG levels in men and women, the Z-scores were adjusted for age, BMI (and years since menopause in

women). SNP were regarded significant if the $P \leq 4x10^{-5}$ (=0.05 / 1253 (the number of independent SNPs)). The number of independent SNPs was calculated using PLINK, for which we regarded all SNPs with an $r^2>0.2$ as representing one signal. In total 1253 independent signals or SNPs were calculated.

Statistical analysis – Age at natural menopause

The effect size for age at natural menopause was calculated using linear regression in MACH2QTL for 14,816 SNPs in both cohorts separately. Inversed variance meta-analysis of the two RS cohorts combined was calculated using METAL[32]. SNPs were regarded significant if $P \leq 4x10^{-5}$. Power calculations were performed using the genetic Power Calculator of Shaun Purcell (http://pngu.mgh.harvard.edu/~purcell/gpc/).

Statistical analysis – lsBMD and vert fx risk

All analyses were carried out for men and women separately. Correlation between lsBMD and vertebral fracture risk was calculated using simple regression ananlysis in SPSS15.0. Effect sizes for lsBMD were calculated using adjusted linear regression in MACH2QTL for 14,816 SNPs. Vertebral fracture risk was calculated using logistic regression in MACH2DAT, this analysis was adjusted for lumbar spine BMD at baseline visit. Inversed variance meta-analysis for lsBMD in RSI and RSII was carried out using METAL[32]. SNP were regarded significant for lsBMD or vert fx risk if $P \leq 4x10^{-5}$.

Statistical analysis – Consistency analysis

Plasma E2, T, and SHBG-levels are associated with lsBMD, as is the age at menopause. Therefore, we also analysed the consistent overlap of SNPs associated with hormone levels or menopausal age and lsBMD. Low BMD is a risk factor for higher fx risk, so we also analysed the consistent overlap of SNPs associated with lsBMD and vert fx risk. In this analysis significance thresholds used in this analysis were a p-value for age at natural menopause or lsBMD ≤0.01 and ≤0.05 for plasma hormone levels and vert fx. Another criterium was the consistency of the effects, lower E2 or T levels and lower lsBMD, lower SHBG levels and higher lsBMD, earlier menopausal age and lower lsBMD, and lower lsBMD and higher vert fx risk, or vice versa. A more stringent p-value for lsBMD and age at natural menopause was chosen because we have more power compared to vertebral fracture risk and hormone levels. This approach would give us ~1 false positive finding per consistency analysis: 0.01 x 0.05 x 1253 = 0.62.

RESULTS

Single phenotype analysis

The gender-specific association analysis for estradiol and testosterone did not result in significant associations with any of the 14,380 SNPs that were studied. For men significant associations were found for plasma SHBG levels (Table 3).

Twenty-one SNPs were associated with SHBG levels. These SNPs are all located 5' of the start of the first *SHBG* exon, the correlation of these 21 SNPs range from 0.001 to 1.00. The most significant SNP is rs12150660, of which the minor allele is associated with 0.46 nmol/L (SE=0.07) higher SHBG levels per copy of the T-allele (MAF=0.24, P $=5.1x10^{-12}$), which is a difference of approximately 2.5% between the two homozygous genotype groups. In women, the only significant SNP was observed for plasma SHBG levels, where rs12150660 is associated with an 0.29 nmol/L increase (SE=0.07) in SHBG levels per copy of the T-allele (MAF=0.24, P=$1.7x10^{-5}$), which is a difference of approximately 1.3% between the two homozygous genotype groups.

Table 3. Significant associations of E2-pathway SNPs for plasma levels of SHBG in men of the Rotterdam study.

Hormone	N	SNP	Gene	Allele	MAF	Beta (nmol/L) (SE)	p-value
SHBG	664	rs12150660	*SHBG*	T	0.24	0.46 (0.07)	$5.1x10^{-12}$
		rs4227	*SHBG*	G	0.26	0.41 (0.06)	$2.2x10^{-11}$
		rs3933469	*SHBG*	A	0.25	0.42 (0.06)	$3.7x10^{-11}$
		rs4968214	*SHBG*	A	0.28	0.35 (0.06)	$2.8x10^{-9}$
		rs4968212	*SHBG*	T	0.27	0.36 (0.06)	$3.0x10^{-9}$
		rs9899183	*SHBG*	C	0.26	0.37 (0.06)	$7.5x10^{-9}$
		rs4511593	*SHBG*	C	0.33	0.33 (0.06)	$9.7x10^{-9}$
		rs12940684	*SHBG*	C	0.29	0.34 (0.06)	$2.2x10^{-8}$
		rs1050540	*SHBG*	T	0.35	0.33 (0.06)	$3.8x10^{-8}$
		rs1050541	*SHBG*	T	0.44	-0.31 (0.06)	$5.6x10^{-7}$
		rs1642772	*SHBG*	A	0.39	0.26 (0.06)	$3.2x10^{-6}$
		rs727428	*SHBG*	T	0.41	-0.28 (0.06)	$3.5x10^{-6}$
		rs8073498	*SHBG*	C	0.38	0.25 (0.06)	$8.9x10^{-6}$
		rs10438740	*SHBG*	G	0.40	0.24 (0.05)	$1.2x10^{-5}$
		rs2270341	*SHBG*	T	0.41	0.24 (0.06)	$1.4x10^{-5}$
		rs858524	*SHBG*	T	0.40	0.23 (0.06)	$1.8x10^{-5}$
		rs1641523	*SHBG*	C	0.39	-0.26 (0.06)	$1.8x10^{-5}$
		rs8066665	*SHBG*	A	0.40	0.23 (0.06)	$2.5x10^{-5}$
		rs1642762	*SHBG*	C	0.38	-0.26 (0.06)	$2.7x10^{-5}$
		rs2908809	*SHBG*	T	0.40	0.23 (0.06)	$2.8x10^{-5}$
		rs858526	*SHBG*	A	0.40	0.23 (0.06)	$3.0x10^{-5}$

N: number of men with data for SHBG-levels; Gene: Gene region; MAF: Minor Allele Frequency.

In RSI and RSII separately we did not find SNPs significantly associated with age at natural menopause, and none of the SNPs reached the significance level of $4x10^{-5}$ in the meta-analysis of RSI and RSII (Figure 1).

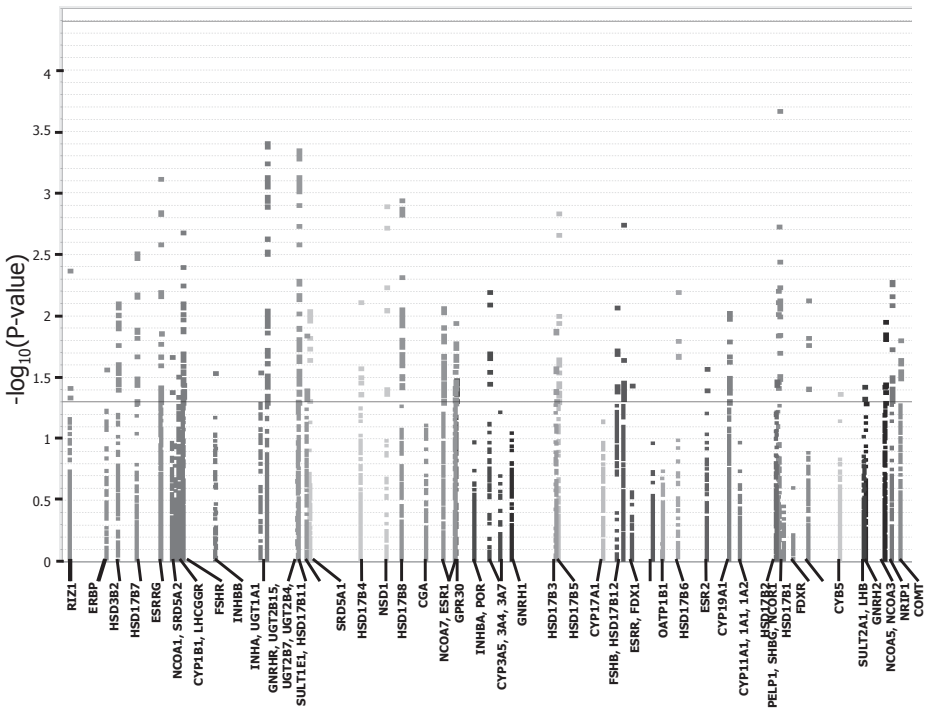


Figure 1. Manhattan-plot for the results of the RSI and RSII meta-analysis of age at natural menopause. The plots shows the negative log P-value for 12,551 SNPs in 61 genes of the sex steroid pathway. The lower horizontal line (at –log10(P-value)=1.3) indicates a p-value of 0.05, the upper horizontal line indicates the significance threshold of $P=4x10^{-5}$.

The combined analysis of RSI and RSII, yielded one SNP significantly associated with lsBMD in women: rs7915365 situated in the 5' region of *HSD17B5*, with an effect of +0.019 g/cm^2 per copy of the minor allele (SE=0.0045, $P=2x10^{-5}$, Table 4). HSD17B5, also known as AKR1C3, converts androstenedione to testosterone. We did not find any significant associations for lsBMD in men, and also no associations with $P\leq4x10^{-5}$ for vertebral fracture risk in men and women of RSI.

Table 4. Top hit for lsBMD in women of the Rotterdam Study I and II meta-analysis.

SNP	chr	position	Gene	MAF	Effect g/cm^2 (SE)	P
rs7915365	10	5047814	*HSD17B5*	0.49	0.019 (0.0045)	$2.1x10^{-5}$

MAF: Minor Allele Frequency; Gene: Gene region; SE: Standard Error.

Consistency analysis

Plasma estradiol, testosterone and SHBG levels and menopausal age in women are associated with BMD. Therefore, we performed a consistency analysis where we searched for SNPs associated with menopausal age or plasma hormone levels and lsBMD. Consistent associations are defined as lower levels of plasma E2 and T, higher levels of SHBG, or earlier menopause together with lower lsBMD, or vice versa. The p-value threshold for hormone levels was set to 0.05, while the p-value threshold was 0.01 for age at natural menopause and lsBMD.

We observed no SNPs associated with menopause and lsBMD in women or with SHBG levels and lsBMD in men and women. For E2-levels and lsBMD we observed no consistently associated SNPs in women, but in men we found one SNP in the *ESR1* gene region associated with lower E2-levels (OR for E2-levels above the median: 0.64 (95%CI[0.47-0.88]; P=0.006) and lower lsBMD (beta=-0.019g/cm^2, SE=0.007; P=0.005), as is shown in Table 5. This SNP is located in *C6orf97*, a gene coding for a hypothetical protein.

Table 5. Consistent associations for E2-levels and lsBMD in men from RSI.

					E2		BMD	
SNP	**chr**	**position**	**Gene**	**MAF**	**OR [95% CI]**	**P**	**Effect g/cm^2 (SE)**	**P**
rs4869738	6	151933844	*ESR1*	0.23	0.64 [0.47-0.88]	0.006	-0.019 (0.007)	0.005

OR: Odds ratio for E2-levels above the median, 95% CI: 95% Confidence Interval, SE: Standard Error, MAF: Minor Allele Frequency.

In women we found two SNPs in the *NSD1* gene region associated with lower T-levels and lower lsBMD (Table 6). Both SNPs had a minor allele frequency of 2%, and are moderately correlated (r^2=0.66). The effects for both SNPs are around -0.4 nmol/L per copy of the minor allele (SE=0.17; P=0.02) for plasma T-levels. The effect of the same allele on lsBMD was approximately -0.045g/cm^2 (SE=0.017; P<0.01).

We also observed significant associations for four SNPs for plasma T-levels and lsBMD in men (Table 7). These four SNPs are all located in the *ESRRG* gene region, representing two signals. The first signal is represented by rs4240918, which has a minor allele frequency of 0.33, and is associated with lower T-levels (effect=-0.119

Table 6. Consistent associations for testosterone levels and lsBMD in women from RSI.

					Testosterone		BMD	
SNP	**chr**	**position**	**Gene**	**MAF**	**Effect nmol/L(SE)**	**P**	**Effect g/cm^2 (SE)**	**P**
rs17077789	5	176549066	*NSD1*	0.02	-0.38 (0.17)	0.02	-0.047 (0.017)	0.005
rs8119	5	176691392	*NSD1*	0.02	-0.41 (0.17)	0.02	-0.043 (0.017)	0.01

MAF: Minor Allele Frequency; Gene: Gene region; SE: Standard Error.

Table 7. Consistent associations for testosterone levels and lsBMD in men from RSI.

					Testosterone		BMD	
SNP	**chr**	**position**	**Gene**	**MAF**	**Effect nmol/L (SE)**	**P**	**Effect g/cm² (SE)**	**P**
rs4240918	1	214697219	*ESRRG*	0.33	-0.119 (0.059)	0.05	-0.018 (0.006)	0.004
rs10735491	1	215179901	*ESRRG*	0.46	-0.118 (0.056)	0.03	-0.016 (0.006)	0.006
rs11572472	1	215192657	*ESRRG*	0.50	0.119 (0.056)	0.03	0.015 (0.006)	0.009
rs1339221	1	215239229	*ESRRG*	0.32	0.114 (0.056)	0.04	0.016 (0.006)	0.009

MAF: Minor Allele Frequency; Gene: Gene region; SE: Standard Error.

nmol/L, SE=0.059; P=0.05) and lower lsBMD (effect=-0.018, SE=0.006; P=0.004). The second signal is represented by the other three SNPs, that are located in intron 1 and 2 of *ESRRG*, with a correlation between the three SNPs of ≥ 0.41. The effects on plasma T-levels and lsBMD are similar to those of rs4240918.

Lower BMD is associated with a higher fracture risk, we therefore, also examined the consistency of SNPs associated with both lsBMD and vert fx risk (p-value threshold 0.01 for lsBMD and 0.05 for vert fx risk, with consistent directions of the effects (i.e., lower lsBMD and higher vert fx risk, or vice versa)). In men 34 SNPs in four gene regions (*LHCGR*, *ESR1*, *HSD17B3*, *NCOR1*) were significantly associated with both lsBMD and vert fx risk (Table 8). These 34 SNPs represent six independent signals, based on the threshold of $r^2<0.2$ for independent signals. In the *LHCGR* gene region seven SNPs are all representing one signal ($r^2>0.37$) the effects of these SNPs on lsBMD are all around 0.02 g/cm² with a change in OR of 35% to 75%. Another associated gene region is located in the *ESR1* gene, from the 5' region of the gene until intron 1. The twelve SNPs are representing two independent signals. The first signal is represented by rs1890010 and rs2504069 ($r^2=1$), showing an effect on lsBMD of -0.02 g/cm² of the minor allele and an OR of 1.33 for vertebral fracture risk. The other signal in this gene is represented by ten SNPs ($r^2=0.75$) which show an effect on lsBMD of 0.02 g/cm², and an decreased odds for vert fx risk of 37%-48%. The SNP associated with lower E2-levels and lower lsBMD (rs4869738) had a low correlation with both signals we observed for the lsBMD and vert fx risk consistency analysis, $r^2=0.021$ for both rs1890010 and rs17081716.

Two other loci associated with lsBMD and vert fx risk are situated in the *HSD17B3* gene region (8 SNPs). The two SNPs with $r^2=1$ are rs11788785 and rs13302476, which both were associated with a 0.023 g/cm² decrease in lsBMD and a 51% increased odds for vert fx risk. The other six SNPs were associated with a 0.02 g/cm² increase in lsBMD and a decreased odds for vert fx risk of 34-40%. The last signal significantly associated with both phenotypes in men is located in the NCOR1 gene region, these SNPs had an $r^2>0.93$, and were associated with an increased lsBMD of 0.016 g/cm², and a decreased

odds for vert fx risk of 0.29% (Table 8).

In women seven SNPs in three gene loci (*POR*, *HSD17B5*, *ESRRA*) are consistently associated with lsBMD and vert fx risk (Table 9). In the *POR* gene region three SNPs were associated with lsBMD and vert fx risk, these three SNPs represent two signals. The first signals comes form rs3757597 and rs3757596 (r^2=1 between the 2 SNPs), and the second from rs12535057. The three SNPs have an effect on lsBMD that is between 0.016 and 0.048 g/cm^2, and an OR for vert fx risk between 29% and 71%. The other two signals that were significant in women are located in the *HSD17B5* and *ESRRA* gene

Table 8. Consistent associations for lsBMD and vert fx risk in men from RSI.

					lsBMD		vert fx risk	
SNP	**chr**	**position**	**Gene**	**MAF**	**Effect g/cm^2 (SE)**	**P**	**OR [95%CI]**	**P**
rs7579411	2	48791860	*LHCGR*	0.41	-0.02 (0.006)	0.0005	1.60 [1.23-2.06]	0.0004
rs6729809	2	48792166	*LHCGR*	0.32	-0.018 (0.006)	0.004	1.62 [1.25-2.10]	0.0003
rs6744967	2	48792297	*LHCGR*	0.48	0.019 (0.006)	0.001	0.65 [0.50-0.85]	0.002
rs4574159	2	48792600	*LHCGR*	0.38	-0.02 (0.006)	0.0006	1.52 [1.17-1.96]	0.001
rs6737881	2	48794184	*LHCGR*	0.27	-0.02 (0.006)	0.002	1.73 [1.32-2.25]	$5.1x10^{-5}$
rs4953616	2	48794932	*LHCGR*	0.28	-0.018 (0.006)	0.004	1.71 [1.32-2.23]	$6.2x10^{-5}$
rs17326446	2	48795068	*LHCGR*	0.49	0.016 (0.006)	0.008	0.74 [0.57-0.97]	0.03
rs1890010	6	152126968	*ESR1*	0.28	-0.02 (0.006)	0.002	1.33 [1.02-1.75]	0.04
rs2504069	6	152127210	*ESR1*	0.28	-0.02 (0.006)	0.002	1.33 [1.02-1.75]	0.04
rs17081716	6	152179341	*ESR1*	0.13	0.024 (0.009)	0.006	0.63 [0.39-1.00]	0.05
rs3853248	6	152181579	*ESR1*	0.12	0.024 (0.009)	0.007	0.59 [0.37-0.95]	0.03
rs3844508	6	152181735	*ESR1*	0.12	0.024 (0.009)	0.007	0.59 [0.37-0.95]	0.03
rs3844509	6	152182157	*ESR1*	0.12	0.024 (0.009)	0.007	0.60 [0.37-0.96]	0.03
rs3866461	6	152188414	*ESR1*	0.12	0.024 (0.009)	0.007	0.61 [0.38-0.97]	0.04
rs7753153	6	152189791	*ESR1*	0.11	0.025 (0.009)	0.007	0.60 [0.37-0.98]	0.04
rs12665044	6	152191565	*ESR1*	0.12	0.024 (0.009)	0.007	0.62 [0.39-0.98]	0.04
rs11155814	6	152192877	*ESR1*	0.14	0.024 (0.009)	0.005	0.63 [0.40-0.98]	0.04
rs7761133	6	152193556	*ESR1*	0.16	0.022 (0.008)	0.008	0.52 [0.33-0.81]	0.004
rs7775047	6	152193886	*ESR1*	0.16	0.022 (0.008)	0.008	0.52 [0.33-0.82]	0.005
rs11788785	9	98071449	*HSD17B3*	0.15	-0.023 (0.008)	0.006	1.51 [1.08-2.11]	0.02
rs9299363	9	98072745	*HSD17B3*	0.37	0.02 (0.006)	0.001	0.60 [0.45-0.80]	0.0006
rs999269	9	98073738	*HSD17B3*	0.37	0.02 (0.006)	0.001	0.60 [0.45-0.80]	0.0006
rs13302476	9	98075007	*HSD17B3*	0.15	-0.023 (0.008)	0.006	1.51 [1.08-2.11]	0.02
rs1983828	9	98080322	*HSD17B3*	0.37	0.02 (0.006)	0.001	0.60 [0.45-0.80]	0.0006
rs1119864	9	98094425	*HSD17B3*	0.34	0.017 (0.006)	0.005	0.68 [0.50-0.91]	0.009
rs8190495	9	98101705	*HSD17B3*	0.37	0.019 (0.006)	0.002	0.66 [0.49-0.88]	0.004
rs2183009	9	98101793	*HSD17B3*	0.37	0.019 (0.006)	0.002	0.66 [0.49-0.88]	0.004
rs2014604	17	16117608	*NCOR1*	0.47	0.016 (0.006)	0.006	0.71 [0.55-0.93]	0.01
rs7219687	17	16118112	*NCOR1*	0.47	0.016 (0.006)	0.006	0.71 [0.55-0.93]	0.01
rs12948244	17	16118825	*NCOR1*	0.47	0.016 (0.006)	0.006	0.71 [0.55-0.93]	0.01
rs4791665	17	16136781	*NCOR1*	0.47	0.017 (0.006)	0.003	0.71 [0.55-0.92]	0.009
rs3112511	17	16148251	*NCOR1*	0.46	0.017 (0.006)	0.003	0.71 [0.54-0.92]	0.009
rs3112513	17	16149537	*NCOR1*	0.46	0.017 (0.006)	0.003	0.71 [0.54-0.92]	0.009
rs6502499	17	16156930	*NCOR1*	0.47	0.017 (0.006)	0.004	0.71 [0.55-0.92]	0.009

chr: chromosome; MAF: Minor Allele Frequency; lsBMD: lumbar spine BMD; vert fx: vertebral fracture risk; SE: Standard Error; OR: Odds Ratio; 95% CI: 95% Confidence Interval.

regions. In both regions, two SNPs are associated with both lsBMD and vert fx risk. The two *HSD17B5* SNPs have moderate correlation of r^2=0.27, and are associated with an increased lsBMD of 0.02 g/cm^2, and a decrease in odds of 26-34% (Table 9). The two SNPs in the *ESRRA* gene region are also moderately correlated (r^2=0.49) and have an effect of 0.06 g/cm^2 per copy of the minor allele for lsBMD and an OR for vert fx risk between 0.07 and 0.19.

Table 9. Consistent associations for BMD and vert fx risk in women from RSI.

					lsBMD		vert fx risk	
SNP	**chr**	**position**	**Gene**	**MAF**	**Effect g/cm^2 (SE)**	**P**	**OR [95%CI]**	**P**
rs3757597	7	75283525	*POR*	0.04	0.048 (0.015)	0.0009	0.29 [0.11-0.75]	0.01
rs3757596	7	75283710	*POR*	0.04	0.048 (0.015)	0.0009	0.29 [0.11-0.75]	0.01
rs12535057	7	75326974	*POR*	0.20	0.016 (0.006)	0.006	0.71 [0.53-0.94]	0.02
rs1937860	10	5042471	*HSD17B5*	0.22	0.016 (0.006)	0.006	0.74 [0.55-0.99]	0.04
rs12261648	10	5054317	*HSD17B5*	0.13	0.02 (0.007)	0.005	0.66 [0.45-0.96]	0.03
rs10128595	11	63841632	*ESRRA*	0.02	0.064 (0.021)	0.002	0.19 [0.04-0.97]	0.05
rs499455	11	63915829	*ESRRA*	0.02	0.065 (0.021)	0.002	0.07 [0.01-0.76]	0.03

chr: chromosome; MAF: Minor Allele Frequency; lsBMD: lumbar spine BMD; vert fx: vertebral fracture risk; SE: Standard Error; OR: Odds Ratio; 95% CI: 95% Confidence Interval.

DISCUSSION

In this study we used a candidate pathway analysis, to study the effects of common genetic variants in genes involved in sex steroid synthesis, metabolism and signalling on plasma sex hormone levels, age at natural menopause, and lsBMD and vert fx risk.

For SHBG levels, we observed SNPs in the 5' region of the *SHBG* gene to be associated with plasma SHBG levels in men and women. These results are in line with previous studies, which already showed association of SNPs near the *SHBG* gene with SHBG levels. Only one of our significant hits was studied before by Thompson et al. who found that the T-allele of the rs727428 SNP had an effect of 4.5 nmol/L per copy of the allele[33]. Although we found a smaller effect size of this polymorphism in our dataset (effect: 0.28 nmol/L), the effect was in the same direction. The rs1799941 polymorphism just in front of the start of the SHBG gene was shown to influence the mRNA expression of the SHBG gene, with a 15-40% difference in SHBG levels between the AA and the GG genotypes[5,34]. This SNP was not analysed in the present study, because it was not described in the HapMap database on which the Illumina HumanHap550K and the imputation are based. Our results together with the previous

results on plasma SHBG levels show that polymorphisms in the 5' region of the gene regulate *SHBG* expression, but it is still unclear which variant is underlying this.

We did not observe significant associations for plasma estradiol or testosterone levels in this study. We also did not replicate the findings by Eriksson et al. for rs2470152 in the aromatase gene (*CYP19A1*) with E1 and E2-levels in men[5]. The reason for this could be the small number of subjects in our study, which is about 10 times smaller compared to the study of Eriksson et al. Due to the non-normal distribution of E2-levels in women we had to analyse these levels dichotomized at the median, which decreases the power of the analysis as compared to a continuous trait. Furthermore, the women in our study are all postmenopausal, which means that all women have low estradiol and testosterone levels and the genetic influences on postmenopausal hormone levels could be of less importance compared to premenopausal and male hormone levels.

We did not observe a significant association for age at natural menopause, although a number of studies showed significant associations of genetic variants in estrogen-related genes with menopausal age[12,35]. We included more samples in our analysis compared to these studies, so we have more power. We also included only women with natural menopause between 40 and 60 years, excluding women with premature ovarian failure (POF), which has a frequency of 1%, or extremely late menopause, as compared to previous studies that include all menopausal ages or all ages above 40 years. We also excluded women using hormone replacement therapy or oral contraceptives at the time of menopause, or women with menopause due to irradiation or surgery. This leads to a better defined phenotype with less heterogeneity, which increases the power of the study, and lower the number of false positive findings. However, menopausal ages were collected retrospectively during interviews, which could cause recall bias. As we think that this recall bias is independent of the genotype of the SNPs we find, it will only decrease the power, but not pose a major problem in this study.

For lsBMD we found one SNP in *HSD17B5* in postmenopausal women associated, while we found no associations for lsBMD in elderly men. HSD17B5 converts androstenedione to testosterone, and is also known as AKR1C3. Polymorphisms in this gene have been associated with bladder cancer[36] and increased gene expression levels have been found in prostate cancer[37]. There is, however, no literature showing polymorphisms in *HSD17B5* influencing bone. We found no SNPs associated with vertebral fracture risk in men or women. This could be due to lack of power, because we only have vertebral X-ray data for 60% of our total dataset and only one in ten had

a vertebral fx.

We next examined SNPs showing consistent association with plasma hormone levels, age at natural menopause and lsBMD and with lsBMD and vertebral fx risk. We found two SNPs in one gene (*NSD1*) for T-levels and lsBMD in women and three genes (*POR*, *HSD17B5*, *ESRRA*) in women for lsBMD and vert fx risk. In men we found one gene for E2-levels and lsBMD (*ESR1*), one gene for T-levels and lsBMD (*ESRRG*) and four genes (*LHCGR*, *ESR1*, *HSD17B3*, *NCOR1*) for lsBMD and vert fx risk.

In postmenopausal women we observed two SNPs in the *NSD1* gene consistently associated with plasma T-levels and lsBMD. NSD1 is a coregulator of the androgen receptor, which is involved in the signalling of testosterone. Mutations in this gene are associated with Sotos syndrome[38,39]. Patient with this syndrome have mental retardation, rapid growth and advanced bone age[40]. This makes *NSD1* a good candidate to study in further research.

In elderly men we found one SNP in *C6orf97*, in the *ESR1* gene region, associated with E2-levels and lsBMD. This SNP had a low correlation with both signals we observed for the lsBMD and vert fx risk consistency analysis in the *ESR1* gene. This could suggest that these three signals represent the same functional variant, and that this variant is tagged by all signals at a very low level of correlation. To be able to draw conclusions on this further study with in depth sequencing of the gene region to detect the causal variant is needed. In men we observed associations of four SNPs in the *ESRRG* gene region with plasma T-levels and lsBMD. ESRRG is an ESR1 co-activator and has recently been shown to affect osteoblast differentiation and bone formation[41]. There are no previous studies on polymorphisms in this gene and the association with plasma T-levels or lsBMD, and therefore these results need replication.

In postmenopausal women we found seven SNPs in three genes consistently associated with lsBMD and vertebral fracture risk. These genes are *HSD17B5*, *ESRRA*, and *POR*, coding for cytochrome P450 oxidoreductase. The two SNPs in *HSD17B5* are correlated with the SNP that was significantly associated with lsBMD with $P<4x10^{-5}$ in the single phenotype analysis ($r^2>0.25$). The SNPs are, however not the same since we set different significance thresholds for the single phenotype analysis ($P<4x10^{-5}$) and the consistency analysis (lsBMD $P<0.01$ and vert fx risk $P<0.05$). At this moment there are no studies on the function of *POR* in the biology of bone, whereas there are two studies on a 23bp variable number of tandem repeats (VNTR) in the *ESRRA* gene. One study shows an association of the VNTR with lsBMD in premenopausal women[42], while the other study could not conform this result in premenopausal women[43]. Bonnelye et al. showed that this orphan receptor plays a role in bone formation in vitro[44], making

ESRRA a good candidate to study in further research.

In men we found 34 SNPs in four other genes consistently associated with lsBMD and vert fx risk, *ESR1*, *LHCGR*, *HSD17B3* and *NCOR1*. For NCOR1 there is no role known in bone biology at this moment, as is the case for HSD17B3. The other consistently associated SNPs are located in the *ESR1* and *LHCGR* genes. For *ESR1*, the two most studied SNPs in intron 1 of this gene, called *Pvu*II and *Xba*I, were found to be associated with BMD in about 5,000 women of two ethnic groups[45], however, this was not shown in a large meta-analysis of European, Caucasian cohorts[24]. Yet, this study reported an association of these same SNPs with vertebral fracture risk[24]. Most of these studies were conducted in women, in men more conflicting results were found. Only one study found an association of the *Pvu*II-*Xba*I polymorphisms with BMD[25], while others did not find this association[26,27]. Although the *Pvu*II and *Xba*I polymorphisms are not the most significantly associated SNPs in our study, the SNPs we find are located in the same LD-block of the gene. This probably indicates that the *Pvu*II and *Xba*I and the polymorphisms we find associated in this study are tagging the same functional SNP. The SNPs in *LHCGR* are located in intron 7 of the gene, in the same LD-block as rs6545061, which was shown by Limer et al.[46] to be associated with heel ultrasound BMD in men. Although they did not examine fracture risk, this study supports our results that *LHCGR* influences bone health in men.

The difference in genes observed in men and women for the bone phenotypes could be due to the difference in incidence of osteoporosis between men and women. In women the incidence is about two times higher compared to men, which could lead to less power in men, and therefore, more false positive findings in men. The difference in incidence is partly due to differences in behavioural factors like smoking and diet, but also to the decrease of estrogens in women after menopause. Estradiol levels in postmenopausal women are three times lower than in men of similar age. Due to the depletion of ovaries there is almost no gonadally produced estrogen and androgen anymore, and therefore the bone is dependent on locally produced hormones, while in men the gonadally produced estrogens and androgens are still present and influencing the bone. It could be that the most important or rate-limiting step in local production of estrogens is different from that in gonadal production of estrogens, giving a difference in SNPs and genes associated with bone endpoints in elderly men and women.

Another explanation for the difference in results in men and women is the physical difference in the size of vertebrae in men and women, and the difference in the strength of the male and female vertebrae. This is due to the higher periosteal

apposition, which also offsets the endosteal bone loss in men, leading to less stress on the vertebrae and less bone loss in men compared to women[47]. Because this might result in less stronger vertebrae in women compared to men, environmental factors may play a bigger role in vertebral fracture risk in women compared to men. Animal studies also revealed differences in male and female bone. Corsi et al.[48] showed in mice that male muscle-derived stem cells have a greater osteoprogenitor potential, leading to differences in bone regeneration between genders. Also, Ishida and Heersche[49] showed that in rat there is a different response to progesterone in cells from male and female rat lumbar vertebrae[49]. Based on this, it can be concluded that male and female lumbar spine bone cells respond differently to sex steroid-related hormones, supporting the results we find in this study, where we identify different genes involved in bone health in elderly men compared to postmenopausal women.

Our study has some strengths and limitations. The major strength is that we conducted a comprehensive hypothesis-based study, where we focused on genes from the estrogen pathway. This in contrast to most of the previously published candidate gene studies, that focused on one or a few genes and only a few polymorphisms per gene. Another strength of our study is the availability of many baseline characteristics that gives us the opportunity to adjust all analyses for many confounding factors, like age, BMI, and precursor hormones.

One of the major limitations of this study is the lack of analysis of the X chromosome genes. This chromosome is the largest genetic difference between men and women and therefore, one of the most interesting to study. We could not study the X chromosome at this point, because there was no imputed data for this chromosome available. The three genes that we missed by this are the androgen receptor (*AR*) and the steroid sulfatase (*STS* or *ARSC*) and arylsulfatase D (*ARSD*) genes. The STS enzyme converts DHEAS and E2-sulfate to DHEA and E2, respectively, whereas ARSD converts E1-sulfate to estrone. Further study is needed to examine the effect of genetic variation in X chromosome genes on bone health.

Another major limitation is the statistical power of our study. We expected *a priori* that the SNPs we studied will explain only a small part of the total variance (approximately 0.5% based on previous GWAS on complex traits). Even for this level of explained variance we would need a sample size of 5,600 subjects for a cut off p-value of $4x10^{-5}$, at 80% power. For none of the phenotypes we studied, we had such a sample size. For the plasma hormone levels the sample size was about 10 times smaller then the calculated sample size. Therefore, replication of our results in larger samples is needed.

The last limitation is that for this study we treated all genes similar and defined a gene region as +/–100kb. We choose this size because we did not want to miss regulatory regions of the genes. But within this 200kb surrounding DNA there could also be other genes located, which in itself could have an effect on the phenotypes we studied. It would probably have been better to define the gene region per gene. Yet, for many genes the exact regulatory region is not known. For some genes, like *ESR1*, *ESR2* and *CYP19A1* the promoter/regulatory region is known, however, to treat all genes equally, we added the 100kb to these known promoter/regulatory region.

In conclusion, we performed a candidate pathway analysis for several plasma hormone levels, age at natural menoapuse in women, lsBMD and vert fx risk in elderly men and postmenopausal women of Northern European ancestry. In this study polymorphisms 5' of the *SHBG* gene were associated with plasma SHBG levels in men and women. Furthermore, we showed in this study no significant association of SNPs in estrogen pathway genes with age at natural menopause. Our results in the lsBMD and vert fx risk analyses indicated gender-specific associations for genetic variations in the sex steroid pathway with lsBMD and vert fx risk. These results require further replication in other cohorts, but also studies to identify the functional variants underlying the associations.

ACKNOWLEDGEMENTS

This study was funded by the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA), Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation 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 Commision (DG XII), and the Municipality of Rotterdam.

REFERENCES

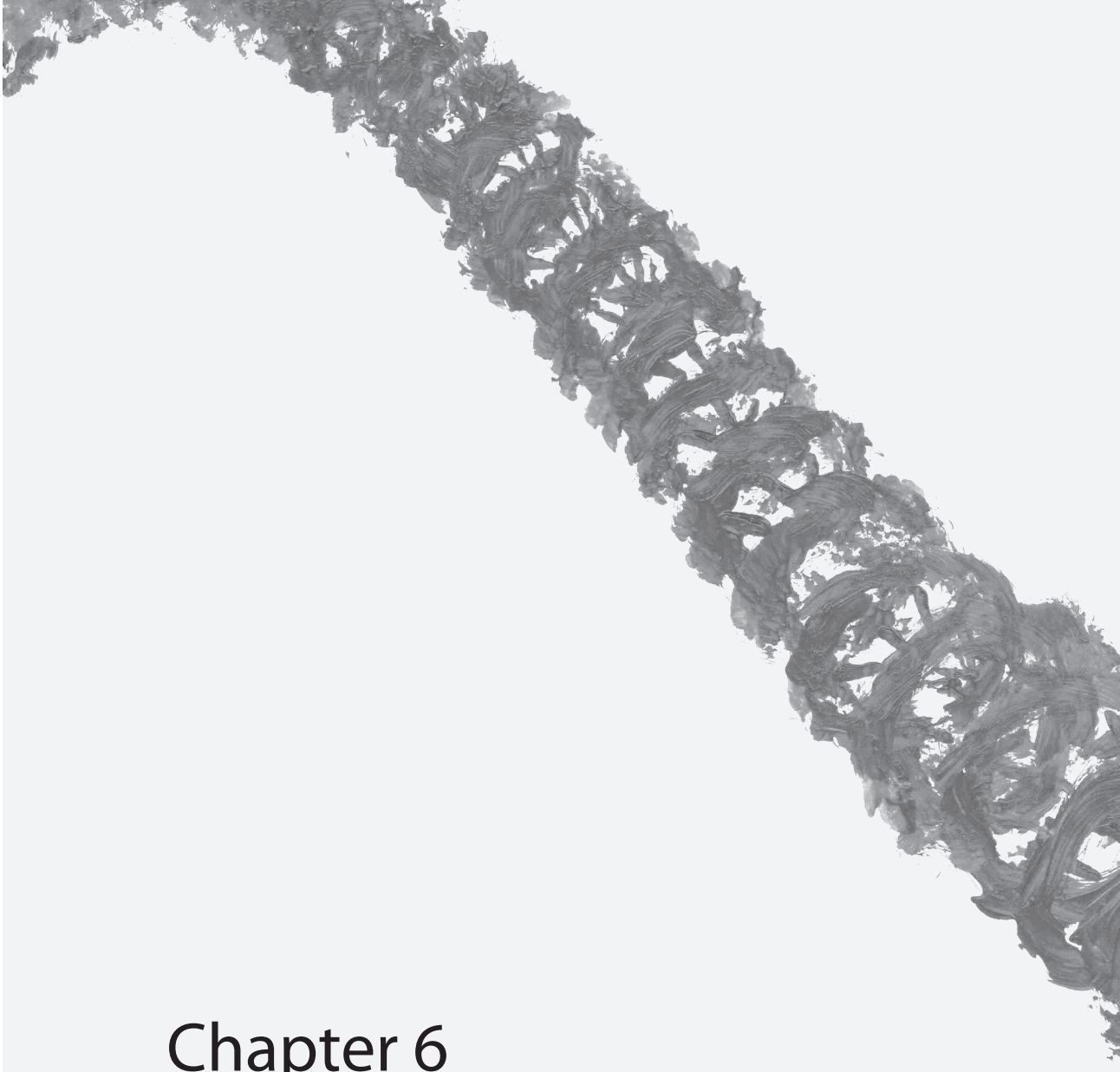
1. Ring, H.Z. et al. Heritability of plasma sex hormones and hormone binding globulin in adult male twins. *J Clin Endocrinol Metab* **90**, 3653-8 (2005).
2. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. *Eur J Endocrinol* **153**, 327-34 (2005).
3. Dunning, A.M. et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *J Natl Cancer Inst* **96**, 936-45 (2004).
4. Haiman, C.A. et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res* **67**, 1893-7 (2007).
5. Eriksson, A.L. et al. Genetic variations in sex steroid-related genes as predictors of serum estrogen levels in men. *J Clin Endocrinol Metab* **94**, 1033-41 (2009).
6. te Velde, E.R., Dorland, M. & Broekmans, F.J. Age at menopause as a marker of reproductive ageing. *Maturitas* **30**, 119-25 (1998).
7. Murabito, J.M., Yang, Q., Fox, C., Wilson, P.W. & Cupples, L.A. Heritability of age at natural menopause in the Framingham Heart Study. *J Clin Endocrinol Metab* **90**, 3427-30 (2005).
8. Snieder, H., MacGregor, A.J. & Spector, T.D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J Clin Endocrinol Metab* **83**, 1875-80 (1998).
9. van Asselt, K.M. et al. Heritability of menopausal age in mothers and daughters. *Fertil Steril* **82**, 1348-51 (2004).
10. Kok, H.S. et al. No association of estrogen receptor alpha and cytochrome P450c17alpha polymorphisms with age at menopause in a Dutch cohort. *Hum Reprod* **20**, 536-42 (2005).
11. Mitchell, E.S. et al. Association of estrogen-related polymorphisms with age at menarche, age at final menstrual period, and stages of the menopausal transition. *Menopause* **15**, 105-11 (2008).
12. Weel, A.E. et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. *J Clin Endocrinol Metab* **84**, 3146-50 (1999).
13. Burger, H.G. The menopausal transition. *Baillieres Clin Obstet Gynaecol* **10**, 347-59 (1996).
14. Guthrie, J.R. et al. The relative effect of endogenous estradiol and androgens on menopausal bone loss: a longitudinal study. *Osteoporos Int* **15**, 881-6 (2004).
15. Kuchuk, N.O. et al. The association of sex hormone levels with quantitative ultrasound, bone mineral density, bone turnover and osteoporotic fractures in older men and women. *Clin Endocrinol (Oxf)* **67**, 295-303 (2007).
16. Mellstrom, D. et al. Older men with low serum estradiol and high serum SHBG have an increased risk of fractures. *J Bone Miner Res* **23**, 1552-60 (2008).
17. Mellstrom, D. et al. Free testosterone is an independent predictor of BMD and prevalent fractures in elderly men: MrOS Sweden. *J Bone Miner Res* **21**, 529-35 (2006).
18. Cooper, C. Epidemiology of osteoporosis. *Osteoporos Int* **9 Suppl 2**, S2-8 (1999).
19. Karasik, D., Cupples, L.A., Hannan, M.T. & Kiel, D.P. Age, gender, and body mass effects on quantitative trait loci for bone mineral density: the Framingham Study. *Bone* **33**, 308-16 (2003).
20. Ralston, S.H. Genetic determinants of osteoporosis. *Curr Opin Rheumatol* **17**, 475-9 (2005).
21. Yang, T.L. et al. Genetic and environmental correlations of bone mineral density at

different skeletal sites in females and males. *Calcif Tissue Int* **78**, 212-7 (2006).
22. Chow, J., Tobias, J.H., Colston, K.W. & Chambers, T.J. Estrogen maintains trabecular bone volume in rats not only by suppression of bone resorption but also by stimulation of bone formation. *J Clin Invest* **89**, 74-8 (1992).
23. Gennari, L. et al. Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review. *Am J Epidemiol* **161**, 307-20 (2005).
24. Ioannidis, J.P. et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *Jama* **292**, 2105-14 (2004).
25. Long, J.R. et al. Association of estrogen receptor alpha and vitamin D receptor gene polymorphisms with bone mineral density in Chinese males. *Calcif Tissue Int* **74**, 270-6 (2004).
26. Langdahl, B.L., Lokke, E., Carstens, M., Stenkjaer, L.L. & Eriksen, E.F. A TA repeat polymorphism in the estrogen receptor gene is associated with osteoporotic fractures but polymorphisms in the first exon and intron are not. *J Bone Miner Res* **15**, 2222-30 (2000).
27. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. *Hum Mol Genet* **12**, 1745-54 (2003).
28. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
29. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
30. Burger, H. et al. The association between age and bone mineral density in men and women aged 55 years and over: the Rotterdam Study. *Bone Miner* **25**, 1-13 (1994).
31. Van der Klift, M., De Laet, C.E., McCloskey, E.V., Hofman, A. & Pols, H.A. The incidence of vertebral fractures in men and women: the Rotterdam Study. *J Bone Miner Res* **17**, 1051-6 (2002).
32. Abecasis, G.R. & Willer, C. METAL. *http://www.sph.umich.edu/csg/abecasis/metal/* (2007).
33. Thompson, D.J. et al. Identification of common variants in the SHBG gene affecting sex hormone-binding globulin levels and breast cancer risk in postmenopausal women. *Cancer Epidemiol Biomarkers Prev* **17**, 3490-8 (2008).
34. Riancho, J.A. et al. Genetic polymorphisms are associated with serum levels of sex hormone binding globulin in postmenopausal women. *BMC Med Genet* **9**, 112 (2008).
35. He, L.N. et al. Association study of the oestrogen signalling pathway genes in relation to age at natural menopause. *J Genet* **86**, 269-76 (2007).
36. Figueroa, J.D. et al. Bladder cancer risk and genetic variation in AKR1C3 and other metabolizing genes. *Carcinogenesis* **29**, 1955-62 (2008).
37. Stanbrough, M. et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* **66**, 2815-25 (2006).
38. Kurotaki, N. et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* **30**, 365-6 (2002).
39. Turkmen, S. et al. Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. *Eur J Hum Genet* **11**, 858-65 (2003).
40. Sotos, J.F., Dodge, P.R., Muirhead, D., Crawford, J.D. & Talbot, N.B. Cerebral Gigantism in Childhood. a Syndrome of Excessively Rapid Growth and Acromegalic Features and a Nonprogressive Neurologic Disorder. *N Engl J Med* **271**, 109-16 (1964).
41. Jeong, B.C. et al. The orphan nuclear receptor estrogen receptor-related receptor gamma negatively regulates BMP2-induced osteoblast differentiation and bone formation. *J Biol Chem* **284**, 14211-8 (2009).
42. Laflamme, N. et al. A frequent regulatory variant of the estrogen-related receptor

alpha gene associated with BMD in French-Canadian premenopausal women. *J Bone Miner Res* **20**, 938-44 (2005).
43. Giroux, S., Elfassihi, L., Cole, D.E. & Rousseau, F. Replication of associations between LRP5 and ESRRA variants and bone density in premenopausal women. *Osteoporos Int* **19**, 1769-75 (2008).
44. Bonnelye, E., Merdad, L., Kung, V. & Aubin, J.E. The orphan nuclear estrogen receptor-related receptor alpha (ERRalpha) is expressed throughout osteoblast differentiation and regulates bone formation in vitro. *J Cell Biol* **153**, 971-84 (2001).
45. Ioannidis, J.P. et al. Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis. *J Bone Miner Res* **17**, 2048-60 (2002).
46. Limer, K.L. et al. Genetic variation in sex hormone genes influences heel ultrasound parameters in middle-aged and elderly men: results from the European Male Aging Study (EMAS). *J Bone Miner Res* **24**, 314-23 (2009).
47. Ruff, C.B. & Hayes, W.C. Subperiosteal expansion and cortical remodeling of the human femur and tibia with aging. *Science* **217**, 945-8 (1982).
48. Corsi, K.A. et al. Osteogenic potential of postnatal skeletal muscle-derived stem cells is influenced by donor sex. *J Bone Miner Res* **22**, 1592-602 (2007).
49. Ishida, Y. & Heersche, J.N. Progesterone stimulates proliferation and differentiation of osteoprogenitor cells in bone cell populations derived from adult female but not from adult male rats. *Bone* **20**, 17-25 (1997).

Part C

Genome-Wide Association Studies



Chapter 6

A Genome-wide Association Study of Sex Hormone Levels; the Rotterdam Study

Lisette Stolk, Joyce BJ van Meurs, Fernando Rivadeneira, Albert Hofman, Huibert AP Pols, Frank de Jong, André G Uitterlinden

Manuscript in preparation

ABSTRACT

Levels of sex steroids are influenced by several environmental and genetic factors and have been associated with several diseases in the elderly, like osteoporosis, breast cancer, osteoarthritis, and cardiovascular disease. In this study we conducted a genome wide association study (GWAS) for plasma levels of estradiol, testosterone, estrone, androstenedione, dehydroepiandrosterone sulfate and sex hormone binding globulin. In approximately 1,500 elderly men and postmenopausal women from the Rotterdam Study we performed gender stratified and combined analysis for these plasma levels. In total 2,414,474 SNPs on all autosomes were included in the analysis for both men and women. We showed highly significant associations of several SNPs in the SHBG gene region with plasma SHBG levels (top hit: rs12150660, effect= +0.37 nmol/L per copy of the minor allele, SE=0.05; $P=3.2x10^{-15}$), and more modest associations for plasma DHEAS levels in women (top hit: rs7656167, odds ratio for DHEAS levels below the median: 1.9 (95%CI[1.5-2.4]); $P=2.6x10^{-8}$) and testosterone in a gender combined analysis (top hit: rs6959028, effect= +0.24 nmol/L per copy of the minor allele, SE=0.04; $P=3.7x10^{-9}$). Because we have little power to detect genome-wide significant associations in our study, we have to be careful with interpretation of our results. The results for SHBG plasma levels are what we expected based on literature, but we cannot be sure that the results for DHEAS and testosterone are true results or false positives due to lack of power, therefore, we need replication of our results.

INTRODUCTION

Sex steroid levels, like estradiol (E2), testosterone (T), dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEAS) have been associated with a number of diseases in the elderly, like osteoporosis[1,2], breast cancer[3], osteoarthritis[4], and cardiovascular disease[5]. Sex hormone binding globuline (SHBG) levels are shown to affect the risk for osteoporosis, breast cancer and cardiovascular disease[5]. SHGB binds testosterone (T) and estradiol (E2), thereby regulating their access to target tissue. Levels of sex steroids are influenced by environmental and genetic factors[6]. Several studies in healthy and diseased men and women have examined associations of genetic variations with levels of estrone (E1)[7-9], E2[7,8,10], SHBG[7,11], DHEAS[12-14], T[7,15] and androstenedione[7,15]. All these studies focused on one or more candidate genes from the estrogen pathway, and all of these studies were based on known biology. We do not know if all genes involved in influencing hormone levels are annotated at this moment. Therefore, in this study we conducted a hypothesis-free genome wide association study (GWAS) for plasma levels of estradiol, testosterone, estrone, androstenedione, dehydroepiandrosterone sulfate and sex hormone binding globulin in approximately 700 healthy elderly men and 800 healthy postmenopausal women from the Rotterdam Study.

MATERIAL and METHODS

Study population

Subjects were participants of the Rotterdam Study, a large prospective population-based cohort studies of Caucasian subjects aged 55 years and over, living in the Ommoord district of Rotterdam, the Netherlands. The study was designed to investigate the incidence and determinants of chronic disabling diseases in the elderly. Rationale and design have been described previously[16]. Among the subjects living independently, the overall response rate was 78 percent for home interview and 71 percent for examination in the research centre, where anthropometric characteristics and bone mineral density were measured, and blood samples were taken. The Rotterdam Study was approved by the medical ethics committee of the Erasmus University Medical School, and written informed consent was obtained from each subject. The current study is based on around 1,500 subjects (800 postmenopausal women) from the baseline cohort of the Rotterdam Study (RSI) for whom genome-

wide genotype data and sex hormone levels were available.

Assessment of hormone levels

Non-fasting blood samples were drawn by venapuncture at the baseline examination in the research center between 08.30h and 16.00h and time of blood draw was noted. Levels of steroid hormones were measured in plasma. For the collection of plasma, blood was collected in 5ml tubes containing 0.5 ml sodium citrate solution. All tubes were stored on ice before and after blood sampling. Platelet-free plasma was obtained by two-stage centrifugation, first for 10 min at 1600 x g at 4°C and then for 30 min at 7000 x g at 4°C. Platelet-free samples were immediately frozen in liquid nitrogen and transferred to the laboratory. At the laboratory, plasma samples were stored at -80°C until hormone measurements. For the purpose of the present study, plasma levels of E2, testosterone, estrone, androstenedione, dehydroepiandrosterone sulfate (DHEAS), and sex hormone binding globulin (SHBG) were estimated in 12 separate batches of samples using coated tube or double antibody radioimmunoassays purchased from Diagnostic Systems Laboratories, Inc. (Webster, TX). The mean minimum detection limit for E2 was 4.8 pmol/l. Non-detectable levels for all hormones were scored as half of the detection limit. Due to the relatively small volumes of plasma available, all values reported are single sample estimations. Intra-assay coefficients of variation, determined on the basis of duplicate results of internal quality control (QC) serum pools with 3 different levels of each analyte, were below 15% for all assays, with the exception of E2 (18%) and E1 (21%). Since inter-assay variations were relatively large (between 20 and 30%, with the exception of SHBG (14%) and testosterone (19%)) results of all batches were normalized by multiplying all concentrations within a batch with a factor, which made results for the internal QC pools comparable. This reduced inter-assay variations and was considered justified because the patterns of the results of these pools and the mean results for male and female sera in one assay batch were very similar. Assays were performed blind with respect to information on the subject. Albumin was measured using a colorimetric method (KONE Diagnostics, Espoo, Finland). Women using hormone replacement therapy (HRT) were excluded from the analysis. Table 1 shows the number of subjects that have data on plasma hormone levels, split by gender.

Table 1. Number of men and women per hormone.

Hormone	Men (N)	Women (N)
Androstenedione	684	740
DHEA	250	274
DHEAS	697	799
Estrone	595	771
Estradiol	469	661
SHBG	664	738
Testosterone	672	760

Genotyping and Imputation

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. Genotyping using the Illumina HumanHap 550K beadarray was performed as described previously[19]. Imputation of the genome-wide genotyping data was performed using MACH. SNPs with imputation quality score below 0.5, Hardy-Weinberg equilibrium $P<1x10^{-6}$, and minor allele frequency (MAF) <0.01 were excluded from further analysis.

Statistical analysis

For all hormones plasma levels below the detection limit of the assay were set to half of the detection limit and included in the analysis. All analyses were carried out for men and women separately. In women estradiol, estrone and DHEAS levels were dichotomised on the median, because of the non-normal distribution of the levels and analysed using adjusted logistic regression in MACH2DAT. To be able to compare the results between men and women, we also analysed these hormone levels in men as a dichotomized trait. This analysis was adjusted for age, BMI, years since menopause (in women), and precursor hormone levels (i.e., estradiol levels were adjusted for testosterone levels, whereas estrone levels were adjusted for androstenedione levels). For all other hormone levels in women and all hormone levels in men adjusted Z-scores were analysed using linear regression in MACH2QTL. In men adjusted Z-scores for all hormones were calculated including age, BMI and if available the precursor as covariates. In women adjusted Z-scores for all hormone levels were calculated including age, BMI, years since menopause and, if available, the precursor hormone. Overlap between men and women was calculated using genomic control adjusted inverse-variance meta-analysis using METAL (Abecasis & Willer). SNPs were regarded significant if $P\leq5x10^{-8}$.

RESULTS

Gender stratified analysis

In total 2,414,474 SNPs on all autosomes were included in the analysis for both men and women. The gender specific analyses showed different top hits in men and women for all plasma hormone level (Table 2). In women we observed two SNPs on chromosome 4 to be significantly associated with dichotomized DHEAS levels (p-value$<5x10^{-8}$). These two SNPs are located in intron 3 of the gene encoding for

hypothetical protein LOC729006 (chromosome 4p15.32) and are associated with a 90% increased odds (95% CI: 50%-140%) for having DHEAS levels below the median. In men no significant associations were found with these SNPs.

In men we observe nine genome-wide significantly associated SNPs for plasma SBHG levels. The most significant SNP was rs12150660, that has an effect of +0.46 nmol/L (SE=0.07; P=5.1x10^{-12}) per copy of the minor allele. This SNP and the eight other genome-wide significant SNPs are all located in the region surrounding the *SHBG* gene on chromosome 17. In women, the effects of these SNPs is 2 times smaller compared to men, with p-values varying between 0.19 and 1.7x10^{-5} (for rs12150660). For the most significant SNP (rs12150660) the effect size in men was +0.46 nmol/L per copy of the T-allele, whereas in women it is +0.29 nmol/L per copy al the T-allele.

Table 2. Genome-wide significant SNPs in the gender specific analyses.

						Men		Women	
Hormone	**SNP**	**All**	**MAF**	**Chr**	**Gene**	**P**	**OR [95%CI]**	**P**	**OR [95%CI]**
DHEAS	rs7656167	T	0.42	4	*LOC729006*	0.14	0.8 [0.7-1.1]	2.6x10^{-8}	1.9 [1.5-2.4]
	rs13109476	A	0.42	4	*LOC729006*	0.14	0.8 [0.7-1.1]	2.8x10^{-8}	1.9 [1.5-2.4]
						P	**Beta (SE)**	**P**	**Beta (SE)**
SHBG	rs12150660	T	0.22	17	*SHBG*	5.1x10^{-12}	0.46 (0.07)	1.7x10^{-5}	0.29 (0.07)
	rs4227	G	0.26	17	*SHBG*	2.2x10^{-11}	0.41 (0.06)	9.3x10^{-4}	0.21 (0.06)
	rs3933469	A	0.25	17	*SHBG*	3.7x10^{-11}	0.42 (0.06)	1.1x10^{-3}	0.21 (0.06)
	rs4968214	A	0.28	17	*SHBG*	2.8x10^{-9}	0.35 (0.06)	2.4x10^{-3}	0.18 (0.06)
	rs4968212	T	0.27	17	*SHBG*	3.0x10^{-9}	0.36 (0.06)	2.4x10^{-3}	0.19 (0.06)
	rs9899183	C	0.26	17	*SHBG*	7.5x10^{-9}	0.37 (0.06)	7.0x10^{-5}	0.25 (0.06)
	rs4511593	C	0.33	17	*SHBG*	9.7x10^{-9}	0.33 (0.06)	0.042	0.11 (0.06)
	rs12940684	C	0.29	17	*SHBG*	2.1x10^{-8}	0.34 (0.06)	2.1x10^{-3}	0.19 (0.06)
	rs1050540	T	0.35	17	*SHBG*	3.8x10^{-8}	0.33 (0.06)	0.19	0.08 (0.06)

All: Allele; Gene: gives the gene region the SNP is located; the beta for SHBG-levels is in nmol/L, MAF: Minor Allele Frequency; P: p-value; OR: Odds Ratio; 95%CI: 95% Confidence Interval; SE: Standard Error.

Combined analysis

To identify SNPs that affect plasma hormone levels in both men and women, we combined both results using inverse variance meta-analysis. This showed genome-wide significant SNPs for both testosterone (three SNPs) and SHBG (nine SNPs), shown in Table 3. For testosterone the most significant SNP was rs6959028 located in an intron of the *HDAC9* gene on chromosome 7p21.1. The overall effect of the minor allele (MAF=0.41) on testosterone levels was +0.24 nmol/L per copy of the allele (SE=0.04, P=3.7x10^{-9}), with similar effects for men and women in the stratified analysis. The two other genome-wide significant SNPs are located on chromosome 6 in an intergenic region and are both associated with a decrease of 0.21 nmol/L testosterone per copy

of the minor allele (SE=0.04, MAF=0.41 and 0.47) with similar p-values for both SNPs. The effect of these two SNPs was also very similar for men and women in the gender stratified analysis.

Table 3 Genome-wide significant results for the men and women meta-analysis.

						Overall		Men		Women	
Hormone	**SNP**	**All**	**MAF**	**Chr**	**Gene**	**P**	**Beta (SE)**	**P**	**Beta (SE)**	**P**	**Beta (SE)**
T	rs6959028	A	0.41	7	*HDAC9*	$3.7x10^{-9}$	0.24 (0.04)	$5.1x10^{-6}$	0.28 (0.06)	$1.3x10^{-4}$	0.21 (0.06)
	rs997732	T	0.41	6	Intergenic	$4.5x10^{-8}$	-0.21 (0.04)	$1.5x10^{-4}$	-0.22 (0.06)	$2.3x10^{-2}$	-0.12 (0.06)
	rs4897448	G	0.47	6	Intergenic	$4.8x10^{-8}$	-0.21 (0.04)	$1.5x10^{-4}$	-0.21 (0.06)	$6.1x10^{-5}$	-0.21 (0.05)
SHBG	rs12150660	T	0.22	17	SHBG	$3.2x10^{-15}$	0.37 (0.05)	$5.1x10^{-12}$	0.46 (0.07)	$1.7x10^{-5}$	0.29 (0.07)
	rs4227	G	0.26	17	*SHBG*	$2.7x10^{-12}$	0.31 (0.04)	$2.2x10^{-11}$	0.41 (0.06)	$9.3x10^{-4}$	0.21 (0.06)
	rs3933469	A	0.25	17	*SHBG*	$4.7x10^{-12}$	0.32 (0.05)	$3.7x10^{-11}$	0.42 (0.06)	$1.1x10^{-3}$	0.21 (0.06)
	rs9899183	C	0.26	17	*SHBG*	$7.6x10^{-12}$	0.31 (0.04)	$7.5x10^{-9}$	0.37 (0.06)	$7.0x10^{-5}$	0.25 (0.06)
	rs4968212	T	0.27	17	*SHBG*	$3.1x10^{-10}$	0.28 (0.04)	$3.0x10^{-9}$	0.36 (0.06)	$2.4x10^{-3}$	0.19 (0.06)
	rs4968214	A	0.28	17	*SHBG*	$3.6x10^{-10}$	0.27 (0.04)	$2.8x10^{-9}$	0.35 (0.06)	$2.4x10^{-3}$	0.18 (0.06)
	rs12940684	C	0.29	17	*SHBG*	$1.3x10^{-9}$	0.26 (0.04)	$2.1x10^{-8}$	0.34 (0.06)	$2.1x10^{-3}$	0.19 (0.06)
	rs727428	T	0.41	17	*SHBG*	$3.4x10^{-8}$	-0.23 (0.04)	$3.5x10^{-6}$	-0.28 (0.06)	$9.5x10^{-4}$	-0.19 (0.06)
	rs4511593	C	0.33	17	*SHBG*	$4.9x10^{-8}$	0.22 (0.04)	$9.7x10^{-9}$	0.33 (0.06)	0.042	0.11 (0.06)

All: Allele; Gene: gives the gene region the SNP is located; the beta for SHBG-levels and T-levels is in nmol/L, MAF: Minor Allele Frequency.

The nine SNPs that are genome-wide significantly associated with plasma SHBG levels are all located in the *SHBG* gene region and all but one are the same as we already reported genome-wide significant for men in the gender stratified analysis. The top hit is again rs12150660, located 11.5kb 5' of the gene, with an overall effect of +0.37 nmol/L (SE=0.05; $P=3.2x10^{-15}$) per copy of the minor allele (MAF=0.22). In Figure 1 the p-values for plasma SHBG levels are plotted for the meta-analysis results, showing a high linkage between all significant SNPs, indicating that they all originate from the same signal.

DISCUSSION

To our knowledge, this study is the first genome-wide association study for a number of plasma sex steroid hormone levels and sex hormone binding globuline in men and women.

We showed highly significant associations of SNPs in the SHBG gene region with plasma SHBG levels, and more modest associations for plasma DHEAS levels in women and testosterone in a gender combined analysis.

The SNPs associated with DHEAS levels are in the gene for the hypothetical protein LOC729006 (chr 4p15.32), which is expressed in the hippocampus[20]. No

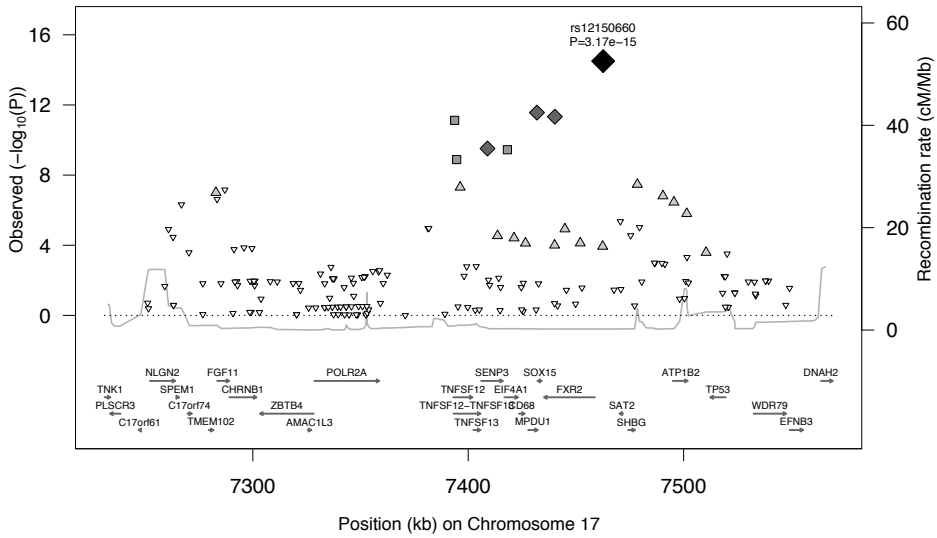


Figure 1. Regional association plot for plasma SHBG levels. The black diamond indicates the most significant SNP after gender combined meta-analysis. The grey scale coding of the shapes indicate the correlation of that SNP with rs12150660 (dark grey diamonds: $0.8 \leq r^2$, light grey squares: $0.5 \leq r^2 < 0.8$, light grey triangles: $0.2 \leq r^2 < 0.5$, white triangles: $r^2 < 0.2$).

function is known for this protein so the mechanism by which genetic variation in this gene could influence DHEAS levels is not known. Plasma testosterone levels were associated in the combined analysis with genetic variation in the histone deacetylase 9 (*HDAC9*) gene and in an intergenic region on chromosome 6. HDAC9 is involved in epigenetic regulation of gene transcription, and also in the immune response[21], however, a function in the regulation of testosterone levels has not been shown before. mRNA expression of this gene is found throughout several tissues in adults, like testis, adrenal glands, bone marrow, and pituitary gland[20]. For the chromosome 6 intergenic region we could not find a possible functionality at this point. Analysis of existing eQTL databases in different tissues (e.g., liver and lymphoblast cell lines) did not show an effect on gene expression of one of these SNPs. So the function of these variants is unknown.

The combined and gender specific analyses further showed significant association of SNPs located close to the *SHBG* gene with plasma SHBG levels. This is in line with previous studies which already showed associations of SNPs nearby the *SHBG* gene with SHBG levels. Only one of our genome-wide significant hits was studied before by Thompson et al.[11] who found rs727428 to have an unadjusted difference of -9.5 nmol/L between the two homozygous genotype groups, with lower

SHBG-levels for carriers of the T-alelle[11]. Although we found a smaller effect size of this polymorphism in our dataset, the effect was in the same direction. However, it is still unclear which of the SNPs is the true functional variant.

We did not find significant associations for E1, E2, androstenedione, and DHEA, where others did find significant associations of genetic variation in the aromatase gene (*CYP19A1*) for E1 and E2 in a young and elderly Swedish male population[8]. This discrepancy might be due to the difference in sample size. The Swedish study examined over 5,000 men in total, which is almost 10 times our sample size for E1 and E2 levels. We therefore have limited power to detect these differences especially because we have to apply a very stringent p-value cut-off. The limited power of our GWAS is an important limitation of our study, therefore, interpretation of our results should be cautious. Furthermore, our women are all postmenopausal, which means that plasma hormone levels for most hormones are very low, and genetic influences on postmenopausal hormone levels could be of less importance compared to premenopausal and male hormone levels. In postmenopausal women, sex steroids are only produced locally and not gonadally anymore; it could be that there is a difference in the most important genes in the production of sex steroids between local and gonadal production. Therefore, in our study the results for the combined elderly men and postmenopausal women analysis could be not informative, because of the different location of the production of sex hormones. Only SHBG levels are not influenced by the postmenopausal state of the women in our study, because it is produced by the liver. One of the strengths of our study is that we do not only have several plasma hormone levels, which gives us the opportunity to adjust the analysis for the precursor hormone. In addition, we also have many baseline characteristics available that gives us the opportunity to adjust for many confounding factors in our own study. To be able to draw firm conclusions about this we need to replicate our findings in other studies.

In conclusion, in this study we showed both gender specific and uniform effects of genetic variation on hormone levels in a genome-wide association study.

ACKNOWLEDGEMENTS

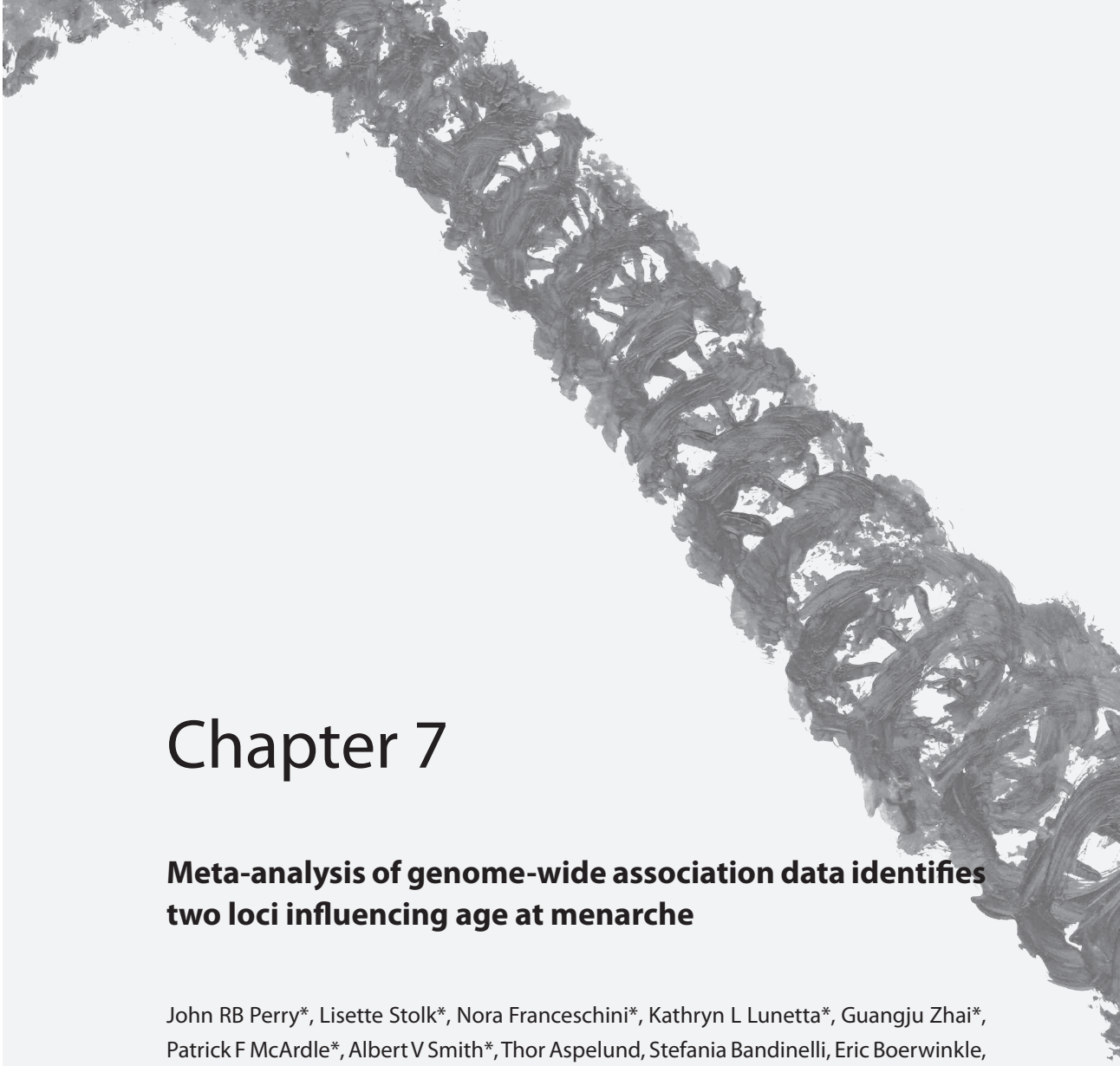
This study was funded by the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA), Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), Research Institute for

Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation 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 Commision (DG XII), and the Municipality of Rotterdam.

REFERENCES

1. Haden, S.T., Glowacki, J., Hurwitz, S., Rosen, C. & LeBoff, M.S. Effects of age on serum dehydroepiandrosterone sulfate, IGF-I, and IL-6 levels in women. *Calcif Tissue Int* **66**, 414-8 (2000).
2. Riggs, B.L., Khosla, S. & Melton, L.J., 3rd. Sex steroids and the construction and conservation of the adult skeleton. *Endocr Rev* **23**, 279-302 (2002).
3. Key, T., Appleby, P., Barnes, I. & Reeves, G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* **94**, 606-16 (2002).
4. Sowers, M.F. et al. Association of bone mineral density and sex hormone levels with osteoarthritis of the hand and knee in premenopausal women. *Am J Epidemiol* **143**, 38-47 (1996).
5. Caldwell, J.D. & Jirikowski, G.F. Sex hormone binding globulin and aging. *Horm Metab Res* **41**, 173-82 (2009).
6. Ring, H.Z. et al. Heritability of plasma sex hormones and hormone binding globulin in adult male twins. *J Clin Endocrinol Metab* **90**, 3653-8 (2005).
7. Dunning, A.M. et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *J Natl Cancer Inst* **96**, 936-45 (2004).
8. Eriksson, A.L. et al. Genetic variations in sex steroid-related genes as predictors of serum estrogen levels in men. *J Clin Endocrinol Metab* **94**, 1033-41 (2009).
9. Haiman, C.A. et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res* **67**, 1893-7 (2007).
10. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. *Eur J Endocrinol* **153**, 327-34 (2005).
11. Thompson, D.J. et al. Identification of common variants in the SHBG gene affecting sex hormone-binding globulin levels and breast cancer risk in postmenopausal women. *Cancer Epidemiol Biomarkers Prev* **17**, 3490-8 (2008).
12. Goodarzi, M.O., Antoine, H.J. & Azziz, R. Genes for enzymes regulating dehydroepiandrosterone sulfonation are associated with levels of dehydroepiandrosterone sulfate in polycystic ovary syndrome. *J Clin Endocrinol Metab* **92**, 2659-64 (2007).
13. Goodarzi, M.O., Xu, N. & Azziz, R. Association of CYP3A7*1C and serum dehydroepiandrosterone sulfate levels in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* **93**, 2909-12 (2008).
14. Smit, P. et al. A common polymorphism in the CYP3A7 gene is associated with a nearly 50% reduction in serum dehydroepiandrosterone sulfate levels. *J Clin Endocrinol Metab* **90**, 5313-6 (2005).
15. Marioli, D.J. et al. Association of the 17-hydroxysteroid dehydrogenase type 5 gene polymorphism (-71A/G HSD17B5 SNP) with hyperandrogenemia in polycystic ovary syndrome (PCOS). *Fertil Steril* (2008).
16. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
17. Sodergard, R., Backstrom, T., Shanbhag, V. & Carstensen, H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. *J Steroid Biochem* **16**, 801-10 (1982).
18. van den Beld, A.W., de Jong, F.H., Grobbee, D.E., Pols, H.A. & Lamberts, S.W. Measures of bioavailable serum testosterone and estradiol and their relationships with muscle

strength, bone density, and body composition in elderly men. *J Clin Endocrinol Metab* **85**, 3276-82 (2000).
19. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
20. NCBI_UniGene. http://www.ncbi.nlm.nih.gov/unigene. (2009).
21. Verdin, E., Dequiedt, F. & Kasler, H.G. Class II histone deacetylases: versatile regulators. *Trends Genet* **19**, 286-93 (2003).



Chapter 7

Meta-analysis of genome-wide association data identifies two loci influencing age at menarche

John RB Perry*, Lisette Stolk*, Nora Franceschini*, Kathryn L Lunetta*, Guangju Zhai*, Patrick F McArdle*, Albert V Smith*, Thor Aspelund, Stefania Bandinelli, Eric Boerwinkle, Lynn Cherkas, Gudny Eiriksdottir, Karol Estrada, Luigi Ferrucci, Aaron R Folsom, Melissa Garcia, Vilmundur Gudnason, Albert Hofman, David Karasik, Douglas P Kiel, Lenore J Launer, Joyce BJ van Meurs, Michael A Nalls, Fernando Rivadeneira, Alan R Shuldiner, Andrew Singleton, Nicole Soranzo, Toshiko Tanaka, Jenny A Visser, Michael N Weedon, Scott G Wilson, Vivian Zhuang, Elizabeth A Streeten*, Tamara B Harris*, Anna Murray*, Tim D Spector*, Ellen W Demerath*, André G Uitterlinden*, Joanne M Murabito*

* These authors contributed equally to this work

Nature Genetics ***41****(6)2009;648-650*

ABSTRACT

We conducted a meta-analysis of genome-wide association data to detect genes influencing age at menarche in 17,510 women. The strongest signal was at 9q31.2 (rs2090409; $P=1.7x10^{-9}$), where the nearest genes include *TMEM38B*, *FKTN*, *FSD1L*, *TAL2* and *ZNF462*. The next best signal was near the *LIN28B* gene (rs7759938; $P=7.0x10^{-9}$), which also influences adult height. We provide the first evidence for common genetic variants influencing female sexual maturation.

INTRODUCTION

Menarche is the start of menstruation and occurs at a mean age of approximately 13 years, normally about 2 years after the onset of puberty[1]. Twin and family studies suggest a significant genetic component to menarcheal age, with at least 50% heritability[2-4]. Linkage and candidate gene studies have not confirmed any loci that influence normal variation in age at menarche[2,5]. Genome-wide association (GWA) studies have been successful in identifying many variants associated with complex disease and quantitative traits and we therefore used this approach to identify genes involved in determining age at menarche. As earlier age at menarche is associated with shorter stature and obesity, the identified variants may not only clarify the genetic control of female sexual maturation but may also point to regulatory mechanisms involved in normal human growth and obesity.

METHODS

Study populations

AGES-Reykjavik Study: The Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967[6]. A total of 19,381 people participated in the Reykjavik Study examination, a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was invited to participate in all subsequent examinations, while one group was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific examinations of the study. Between 2002 and 2006, the AGES-Reykjavik Study re-examined 5,764 survivors of the original Reykjavik Study. Successful genotyping was available for 1,849 AGES women participants who were eligible for this study. The AGES-Reykjavik Study GWAS was approved by the National Bioethics Committee and the Data Protection Authority and also was covered under the MedStar Institutional Review Board. All subjects provided written informed consent.

ARIC: The ARIC study is a multi-center prospective investigation of atherosclerotic disease in a bi-racial population[7]. White and African American men and women aged 45-64 years at baseline were recruited from 4 communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals participated in the

baseline examination in 1987-1989, with four follow-up examinations in approximate 3-year intervals, during 1990-1992, 1993-1995, and 1996-1998. Only White women with genotype data and age at menarche between 9 and 17 years of age were included in this analysis (N=4,247). This study was approved by the institutional review board at each field center, and this analysis was approved by the University of North Carolina at Chapel Hill School of Public Health Institutional Review Board on research involving human subjects. All subjects provided written informed consent.

FHS: The Original Cohort of the Framingham Heart Study was enrolled in 1948 to study determinants of cardiovascular disease and other major illnesses[8,9]. In 1971, Offspring of the Original Cohort participants and Offspring spouses including 2,641 women (mean age 36 years) were enrolled into the Framingham Offspring Study. Offspring participants have been examined approximately every 4 years[10,11]. From 2002 to 2005, 4,095 adults including 2,641 women (mean age 40 years) with at least one parent in the Offspring cohort were enrolled in the Framingham Third Generation cohort[12]. The Framingham Heart Study examinations were approved by the institutional review board at Boston University Medical Center. All participants provided written informed consent.

HAPI Heart Study: The Heredity and Phenotype Intervention (HAPI) Heart Study was initiated in 2002. Participants of the HAPI Heart Study comprised adults from the Old Order Amish community of Lancaster County, PA, who were recruited over a three-year period. Study participants were included if they were aged 20 years and older and considered to be relatively healthy based on exclusion criteria of severe hypertension (blood pressure > 180/105 mm Hg), malignancy, and kidney, liver or untreated thyroid disease. The study aims and recruitment details, including ascertainment criteria, have been described previously[13]. Physical examinations were conducted at the Amish Research Clinic in Strasburg, PA and a reproductive health questionnaire was completed by female participants. Women presenting pregnant or within 6 months postpartum were excluded from the study.

InCHIANTI: The InCHIANTI study is a population-based epidemiological study aimed at evaluating factors that influence mobility in the older population living in the Chianti region of Tuscany, Italy. Details of the study have been previously reported[14]. Briefly, 1,616 residents were selected from the population registry of Greve in Chianti (a rural area: 11,709 residents with 19.3% of the population greater than 65 years of age) and Bagno a Ripoli (Antella village near Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n= 1,453) and participants ranged between 21–102 years of age. The study protocol was approved by the Italian

National Institute of Research and Care of Aging Institutional Review. There were 85 parent-offspring pairs, 6 sib-pairs and 2 halfsibling pairs documented. We investigated any further familial relationships using IBD of 10,000 random SNPs using RELPAIR and uncovered 1 parent-offspring, 79 siblings and 13 half-sibling[15]. We utilized the correct family structure inferred from genetic data for all analyses.

RSI and RSII: Rotterdam Study I and II, ongoing prospective population-based cohort studies, focus on chronic disabling conditions of the elderly in the Netherlands. In summary, men and women aged 55 years or older, living in Ommoord, a suburb of Rotterdam, the Netherlands, were invited to participate[16].

TwinsUK: The TwinsUK cohort consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics[17,18].

Phenotype Definition

AGES: A reproductive history questionnaire was administered to all women at the entry into the AGES Reykjavik study (2002-2006). The question asked to determine age at menarche was: "*At what age did your menstrual periods begin?*" Women answering between 9 and 17 years were included in the analyses. Height was measured at their first visit to the Reykjavik Study (1967–1997) when subjects were at a mean age of approximately 50 years.

ARIC: A reproductive history questionnaire was administered to all women at the baseline (1987-1989) visit. The following question was asked to determine age at menarche: "*At approximately what age were you when your menstrual periods started?*" Responses were given to the closest whole year value (11, 12, 13, etc). Mean age at menarche among ARIC white women was 12.88 (SD 1.54). Height was obtained during the baseline clinic visit by trained technicians.

FHS: At the second Offspring examination (1979 to 1982) and at the first Third Generation examination (2002 to 2005), women were asked, "*Age at start of menses*" and "*How old were you when you had your first menstrual period (menses)?"* respectively. The self-reported age at first period was recorded. Offspring women were asked again about menarche at the time of participation in the Framingham Osteoporosis Study (1996 – 2001: *"About how old were you when you had your first menstrual period?"*). If menarche data were missing from Offspring examination two, the self-reported data from the Osteoporosis examination was used (n=214). There were 1,777 Offspring

Cohort and 2,024 Third Generation women who reported an age at menarche between 9 and 17 years with genotyping available. The mean age at menarche was 12.8 years (SD=1.5) in the combined Offspring and Third Generation women in the sample. Height was measured by trained technicians at the first Offspring and Third Generation examinations.

HAPI Heart Study: During the baseline visit, women were given a reproductive history questionnaire. The self reported age at first period was recorded from the question "*How old were you when you had your first menstrual period?*"

InCHIANTI: During baseline visit women were asked "*How old were you when you had your first menstrual period?*"

RSI and RSII: At the first RSI interview (1989-1993) and at the first RSII interview (2000-2001) women were asked "*How old were you when you had your first menstrual period?*". Self-reported age at menarche was available for 3,175 (RSI) and 1,000 (RSII) women. Mean age at menarche for the RSI cohort was 13.5 years (SD=1.6) and 13.3 years (SD=1.5) for RSII.

TwinsUK: Data on age at menarche was obtained by self-administered questionnaire. All females from the TwinsUK cohort were asked a question *"How old were you when you had your first menstrual period?"* There were 5523 female twins reporting their age at menarche. Of them, 2276 females of European descent (458 MZ pairs, 548 DZ pairs, and 264 singletons) had genotyping data available and were included in the analysis. The mean age at menarche was 12.99 years with SD=1.55 (range 9-17).

Genotyping and Imputation

There were four different genotyping platform used by the eight cohort studies: Illumina Human CNV 370 (AGES)/HumanHap 317K (TwinsUK), the Affymetrix Genome-Wide Human SNP Array 6.0 (ARIC), the Affymetrix 500K mapping array (HAPI) and the Affymetrix 500K in combination with the 50K supplemental array (FHS) and the Illumina Infinium II Human Hap 550 SNP chip array (InCHIANTI, RSI, RSII). Each study performed genotyping quality control checks based on duplicate sample genotyping, SNP call rate, Hardy-Weinberg equilibrium, Mendelian inconsistencies, and sex mismatch, and principle components methods were used to evaluate the presence of population stratification (details provided in Table 1).

Because there were only about 55,000 overlapping SNPs from the four genotyping platforms, each study imputed 2.5 million HapMap SNPs for each participant using currently available imputation methods. InCHIANTI and TwinsUK

used IMPUTE (http://www.stats.ox.ac.uk/~marchini/software/gwas/impute) and all other cohorts used the MACH algorithm (http://www.sph.umich.edu/csg/abecasis/MaCH/). All studies imputed the genotype "dosage" (0, 1, 2) for the expected number of minor alleles. Imputation quality was determined by either the r^2 value produced by MACH or calculated empirical variance divided by the expected variance (oevar) and for SNPTEST the 'proper info' output variable was used to determine imputation quality. SNP imputation methods and quality control procedures for each cohort are included in Table 2.

Table 1. Cohort genotype and phenotype descriptions.

Study	N	Mean (SD)	Array	Call-rate cut-off	MAF cut-off	HWE cut-off
ARIC	4,247	12.9 (1.5)	Affymetrix6.0	0.95	0.01	$1x10^{-5}$
FHS	3,801	12.8 (1.5)	Affymetrix500K + Affymetrix50K	0.97	0.01	$1x10^{-6}$
RSI	3,175	13.5 (1.6)	Illumina550K	0.98	0.01	$1x10^{-6}$
TwinsUK	2,276	13.0 (1.6)	Illumina317K	0.95[#]; 0.99[$]	0.01	$5.7x10^{-7}$
AGES	1,849	13.6 (1.3)	Illumina370CNV	0.97	0.01	$1x10^{-6}$
RSII	1,000	13.3 (1.6)	Illumina550K	0.98	0.01	$1x10^{-6}$
InChianti	597	13.3 (1.5)	Illumina550K	0.98	0.01	$1x10^{-4}$
HAPI Heart Study	565	13.1 (1.3)	Affymetrix 500K	0.95	0.01	$1x10^{-6}$

MAF: Minor Allele Frequency; HWE: Hardy-Weinberg Equilibrium; [#]: 0.95, if MAF>0.05; [$]: 0.99, if 0.01<MAF<0.05.

Statistical Analysis

AGES: Analysis was performed using linear regression against the imputed genotype dosage with the ProbABEL package. Birthyear was included as a covariate.

ARIC: Population stratification was estimated using principal component methods (EIGENSTRAT)[19], after removing few related individuals. Two principal components were significantly associated with age at menarche in linear regression models (alpha=0.05) and so they were included, along with year of birth and study center, as covariates in the genetic analyses. We used linear regression models and assumed additive genetic effects to study the association of imputed and genotyped SNPs (dosage data) and age of menarche. The analyses were implemented in the ProbABEL package from the ABEL set of programs (http://mga.bionet.nsc.ru/yurii/ABEL/)[20].

FHS: SNP weights for 10 principal components (PCs) were inferred using a maximal set of independent individuals; the PCs for the remaining individuals were

computed using the SNP weights obtained from the unrelated set of individuals. The first PC (PC1) was significantly associated with age at menarche (P<0.01), and therefore was included as a covariate in all SNP association analyses. In addition, we adjusted for birth cohort by decade. Linear mixed effects models were used to account for familial correlations. Each SNP was tested for association with age at menarche using an additive genetic model.

HAPI Heart Study: Analysis was performed using in house developed software. In brief, we performed a measured genotype approach utilizing a t-test of the beta coefficient for the SNP variable. We included birth year as a fixed covariates in the model and a polygenic component modeled as a random effect to account for the full 13-14 generation pedigree of the Amish. A total of 338,598 autosomal SNPs were used for imputation after applying four filters: (1) not in HapMap, (2) frequency < 0.01, (3) Hardy-Weinberg $P<1x10^{-6}$, and (4) missingness > 0.05.

InCHIANTI: Analysis performed using linear regression allele dosage in SNPTEST (http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest). Birthyear was included as a covariate.

RSI and RSII: Adjusted linear regression analysis was done using MACH2QTL (http://www.sph.umich.edu/csg/abecasis/MaCH/), birthyear was included as a covariate.

TwinsUK: Because of the relatedness in the TwinsUK cohort, we utilized the GenABEL software package[20] which is designed for GWAS analysis of family-based data by incorporating pair-wise kinship matrix calculated using genotyping data in the polygenic model to correct relatedness and hidden population stratification. The score test implemented in the software was used to test the association between a given SNP and the age at menarche with adjustment for birthyear as a covariate.

Table 2. Cohort imputation descriptions.

Study	number of SNPs	Imputation Program	Analysis program	Lambda
ARIC	2,423,704	MACH v1.0.16	ProABEL	1.03
FHS	2,529,104	MACH v1.0.15	R packages Kinship, LME	1.01
RSI	2,542,887	MACH	MACH2QTL	1.03
TwinsUK	2,544,233	IMPUTE	GenABEL	1.01
AGES	2,532,729	MACH	ProbABEL	1.03
RSII	2,542,887	MACH	MACH2QTL	1.01
InChianti	2,565,135	IMPUTE	SNPTEST	1.04
HAPI Heart Study	2,543,014	MACH	ITSNBN (in-house)	1.04

Imputation Backbone (NCBI build): phased CEU haplotypes, HapMap release 21 (build 35).

Meta-analysis: Menarche GWA

Inverse variance meta-analysis of the 8 studies was performed using the latest version of METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). A SNP within a study was not included if the minor allele frequency (MAF) was < 1% or imputation quality score was < 0.4 for SNPTEST or < 0.3 for MACH in that study. Genomic control was applied to the meta-analysis in METAL to correct for relatedness and population stratification (http://www.sph.umich.edu/csg/abecasis/metal) (Table 2). The meta-analysis included 2,551,160 autosomal, QC-ed, SNPs and 17,510 samples.

Height SNPs

Association statistics for measured adult height were calculated for rs2090409, rs314277 and rs7759938 in the same study samples used for the age at menarche genome wide analysis. The total number with measured height of the 17,510 menarche samples was 16,371. The fixed-effects weighted meta-analysis was performed in METAL.

RESULTS and DISCUSSION

Twenty-eight SNPs passed the conventional genome-wide significance threshold of $P<5x10^{-8}$ and were at either 9q31.2 or 6q21. The 18 SNPs on chromosome 9 were in linkage disequilibrium (LD), with $r^2>0.31$ (Figure 1A), as were the 10 SNPs on chromosome 6, with $r^2>0.96$ (Figure 1B).

To identify more than one signal that could account for the association findings, we carried out conditional analysis adjusting for the SNP with the lowest p-value in the region (rs7759938 for chromosome 6 and rs2090409 for chromosome 9). Within 1Mb flanking each SNP, the lowest adjusted p-values for association with age at menarche were P=0.0017 and P=0.0077 for chromosomes 9 (1,030 SNPs) and 6 (775 SNPs), respectively. These findings suggest a single signal accounting for the associations at each locus. The quantile-quantile plot (Figure 2) showed modest deviation away from the null when these top two signals were removed, suggesting the presence of additional loci for this trait.

The strongest signal at 9q31.2 was observed with rs2090409, where each A allele was associated with approximately a 5-week reduction in menarcheal age ($P=1.7x10^{-9}$). All studies showed consistent evidence of association with the same direction of effect in all but one study, similar effect sizes and p-values between 0.8

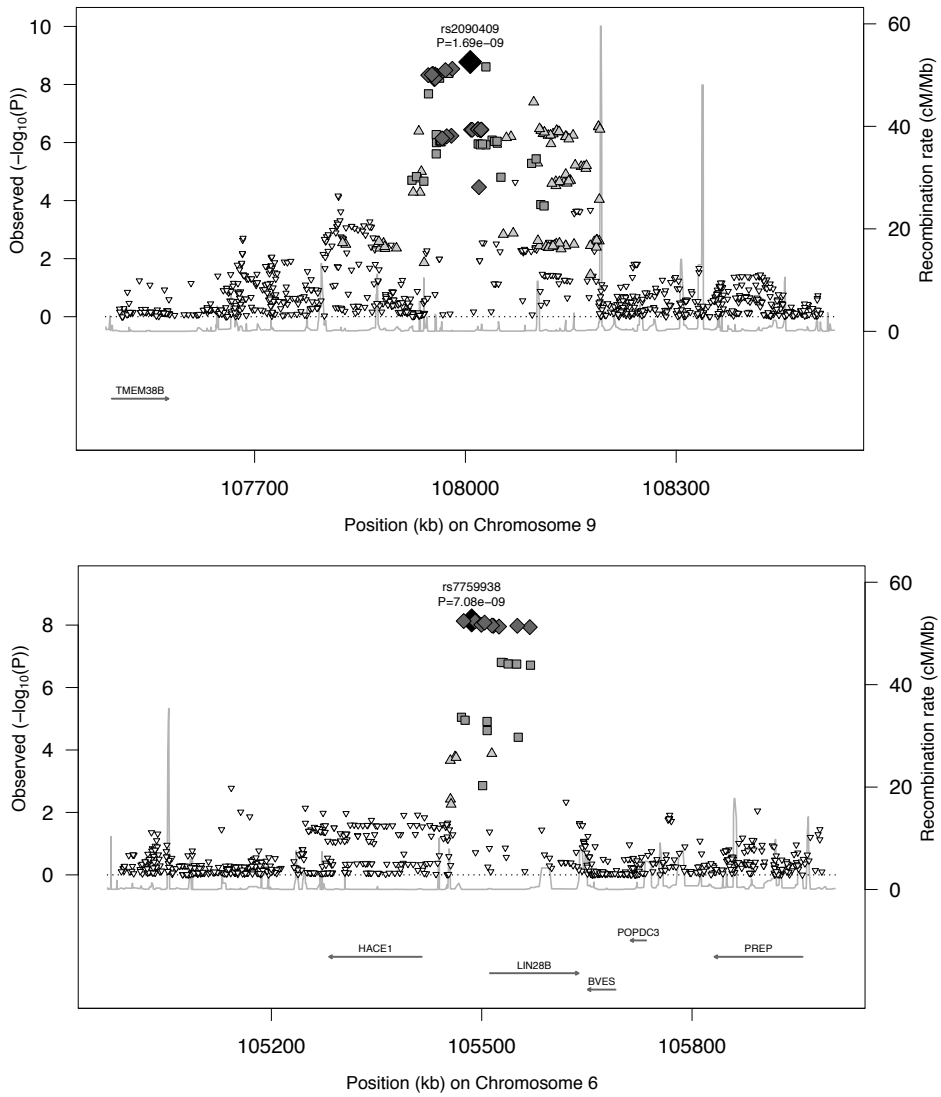


Figure 1. Genomic context (based on NCBI B36) of the top two independent signals at 9q31.2 (1A) and 6q21 (1B) plotted against association –log10 p-values. Only UCSC Refseq genes are shown. r^2 between each SNP and the black diamond indicates the most significant SNP. The grey scale coding of the shapes indicate the correlation of that SNP with the top hit (dark grey diamonds: $0.8 \leq r^2$, light grey squares: $0.5 \leq r^2 < 0.8$, light grey triangles: $0.2 \leq r^2 < 0.5$, white triangles: $r^2 < 0.2$). Chromosome positions are based on build hg18.

and 0.0003 (Table 3).

The recombination region containing rs2090409 includes only a hypothetical gene (*BC039487*). Outside of this, the only RefSeq gene within a 1Mb window is a

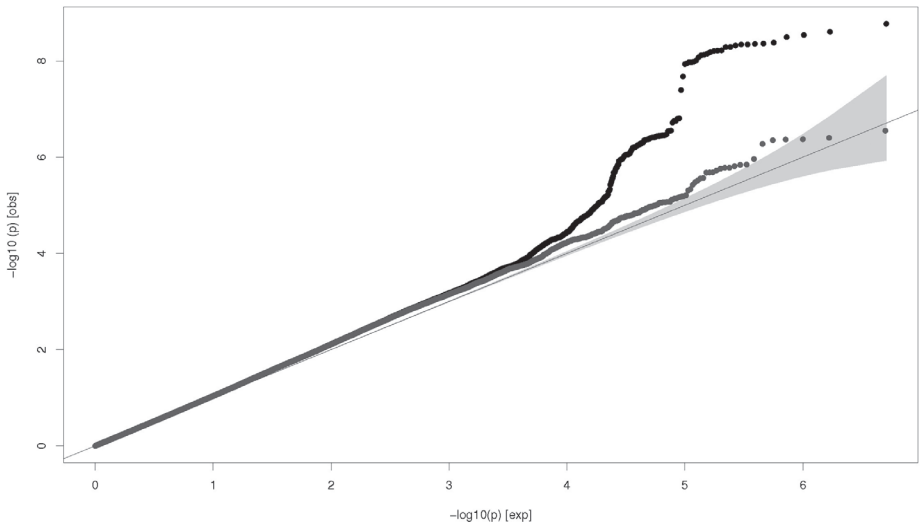


Figure 2. A quantile-quantile plot of the 2.55 million imputed SNPs on age at menarche (black dots) and excluding the 2 SNPs which reached genome-wide significance and all SNPs within 500kb upstream and downstream of those top signals (grey dots). 95% confidence interval is shown in light grey.

transmembrane protein gene, *TMEM38B*, which is approximately 400kb proximal to the GWAS signal. In mice, *Tmem38b* is expressed strongly in brain and the null mutation is neonatal lethal.

Within 2Mb of the signal, genes include *SLC44A1*, *FKTN*, *FSD1L*, *TAL2* and *ZNF462*, none of which is an obvious candidate gene for involvement in menarche. However, a SNP in *ZNF462*, 650kb from our signal but not in LD (r^2=0.086), has been previously associated with variation in height[21]. The 6q21 signal was within a recombination interval that included only one gene, *LIN28B* (Figure 1B) and was also associated with approximately a 5-week reduction in menarcheal age per T allele (rs7759938; P=7.0x10^{-9}). The effect was consistent across all studies, with p-values between 0.27 and 0.001 (Table 3). A common variant in the *LIN28B* gene has previously been associated with normal variation in adult height[22]. The most significant menarche-associated variant (rs7759938) and the previously reported height variant (rs314277) lie within 28.7kb of each other and are likely to represent the same signal, as r^2=0.26 and D'=1 in HapMap. The allele associated with earlier age at menarche is associated with decreased height, which is consistent with epidemiological data. Early menarche has been correlated with reduced stature, and the mechanism is probably mediated

Table 3. Genome-wide significant associations with age at menarche.

SNP	Study	Allele	Frequency	Imputation quality	Effect (yr)	SE	P
rs2090409	ARIC	A	0.31	1.00	-0.10	0.04	0.004
	FHS	A	0.31	0.99	-0.07	0.04	0.07
	RSI	A	0.34	1.01	-0.08	0.04	0.06
	TwinsUK	A	0.32	0.99	-0.11	0.03	0.0003
	AGES	A	0.27	0.99	-0.08	0.05	0.09
	RSII	A	0.35	1.06	-0.15	0.07	0.04
	InChianti	A	0.33	0.98	-0.26	0.09	0.005
	HAPI Heart Study	A	0.28	0.95	0.03	0.10	0.78
	Meta-analysis	**A**	**0.31**	**1.00**	**-0.10**	**0.02**	**$1.7x10^{-9}$**
rs7759938	ARIC	C	0.33	0.94	0.12	0.04	0.001
	FHS	C	0.33	0.90	0.10	0.04	0.009
	RSI	C	0.31	1.00	0.06	0.04	0.16
	TwinsUK	C	0.32	0.98	0.06	0.03	0.04
	AGES	C	0.34	1.00	0.08	0.04	0.07
	RSII	C	0.30	0.94	0.09	0.08	0.27
	InChianti	C	0.29	0.98	0.24	0.09	0.008
	HAPI Heart Study	C	0.23	0.91	0.34	0.11	0.002
	Meta-analysis	**C**	**0.33**	**0.96**	**0.09**	**0.02**	**$7.0x10^{-9}$**

Meta-analysis p-values are corrected by individual-study genomic control inflation factors. Alleles are based on forward strand and positions on NCBI build 36. Meta-analysis frequency is calculated as weighted average across all studies. Imputation quality refers to the imputation quality score generated by MACH (oevar) / SNPTEST (proper_info).

through earlier exposure to estrogens resulting in earlier closure of the epiphyseal plates[23]. We therefore tested all published common variants influencing height - 44 independent loci - for association with age at menarche in our dataset[24].

Six of the alleles were also associated with menarcheal age (P<0.05), with the strongest associations at *LIN28B* (P=0.0001) and *PXMP3* (P=0.003) (Table 4). We also tested the association of the newly identified menarche-associated variant, rs7759938, with measured height in our study population and found that it was associated with height, P_{meta}=0.0001, in the same direction in all but one study; that is, the C-allele was associated with reduction in age at menarche and also reduced stature. The published height SNP (rs314277) did not reach nominal significance with height in our study (P_{meta}=0.26). These data suggest that some of the previously identified loci that influence adult height may also have a general role in adolescent growth.

At a given chronologic age, girls with earlier age at menarche tend to have greater body mass index (BMI) and adiposity than girls with a later age at menarche[25-27]. A marked secular decline in age at menarche occurred in Europe in the nineteenth and

Table 4. Effects of height loci on age at menarche.

Gene	SNP	Chr	Position	Height Increasing Allele	Menarche Increasing Allele	Effect (Years)	SE	P	Direction
LIN28B	rs314277	6	105514355	A	A	0.096	0.03	0.0001	++++++++
PXMP3	rs7846385	8	78322734	C	T	0.049	0.02	0.003	+-++++++
C6orf173	rs4549631	6	127008001	C	T	0.030	0.01	0.02	+-++++++
GNA12	rs798544	7	2729628	C	C	-0.035	0.02	0.03	--------
DYM	rs8099594	18	45245158	A	A	0.032	0.02	0.04	+-++++++
SCMH1	rs6686842	1	41303458	T	C	-0.028	0.01	0.04	------++

Effect sizes in years, P-values corrected by individual study genomic control inflation factors. Alleles based on forward strand, positions on NCBI build 36. Direction of effect is given for each study in following order: RSI, RSII, ARIC, FHS, InChianti, TwinsUK, AGES, Hapi Heart study.

early twentieth centuries, which has been attributed to improved nutrition and health[1]. This trend may be continuing as a consequence of the obesity epidemic[28] and may involve a common metabolic response to the current nutritional environment[29] or be attributable, at least in part, to shared genetic influences or pleiotropy[30]. We therefore investigated the effect on menarcheal age of the ten currently known common gene variants associated with variation in BMI. Of these ten loci, eight showed an association in the direction consistent with epidemiological data ($P=1.6x10^{-6}$, based on Fisher's combined probability test: $-2x\Sigma(\ln \text{p-value})$ against χ^2 on (10 x 2) df), and five were nominally significant ($P<0.05$) (Table 5). The two loci with the largest observed effects on BMI (*FTO* and *TMEM18*) also had the strongest evidence for association with menarcheal age ($P=0.0008$ and $7.0x10^{-5}$, respectively).

Table 5. Effects of 10 BMI genes on age at menarche.

SNP	Chr	Position	Nearby Gene	BMI Increasing allele	Menarche P	N	Effect (Years)	SE	Directions
rs6548238	2	624905	*TMEM18*	C	$7x10^{-5}$	13,263	-0.1	0.02	++?+++++
rs9939609	16	52378028	*FTO*	A	0.0008	17,510	-0.05	0.01	-+--+---
rs4074134	11	27603861	*BDNF*	C	0.006	17,510	-0.05	0.02	++++++++
rs2815752	1	72524461	*NEGR1*	A	0.02	17,510	-0.03	0.01	-------+
rs10938397	4	45023455	*GNPDA2*	G	0.04	17,510	-0.03	0.02	-+++-++-
rs11084753	19	39013977	*KCTD15*	G	0.21	17,510	-0.02	0.02	++++-+--
rs7498665	16	28790742	*SH2B1*	G	0.41	17,510	-0.01	0.02	++++--++
rs7647305	3	187316984	*SFRS10*	C	0.48	17,510	-0.01	0.02	++++---+
rs17782313	18	56002077	*MC4R*	C	0.82	17,510	0.004	0.02	++--+-+-
rs10838738	11	47619625	*MTCH2*	G	0.96	17,510	0.0009	0.02	--+-++-+

Effect sizes in years, based on BMI increasing allele. P-values corrected by individual study genomic control inflation factors. Alleles based on forward strand, positions on NCBI build 36. Direction of effect is given for each study in following order: RSI, RSII, ARIC, FHS, InChianti, TwinsUK, AGES, HAPI Heart study; ? indicates where data were not available.

This study provides the first evidence for common genetic variants influencing normal variation in the timing of female sexual maturation. Our findings also indicate a genetic basis for the phenotypic associations between age at menarche and both height and BMI.

ACKNOWLEDGEMENTS

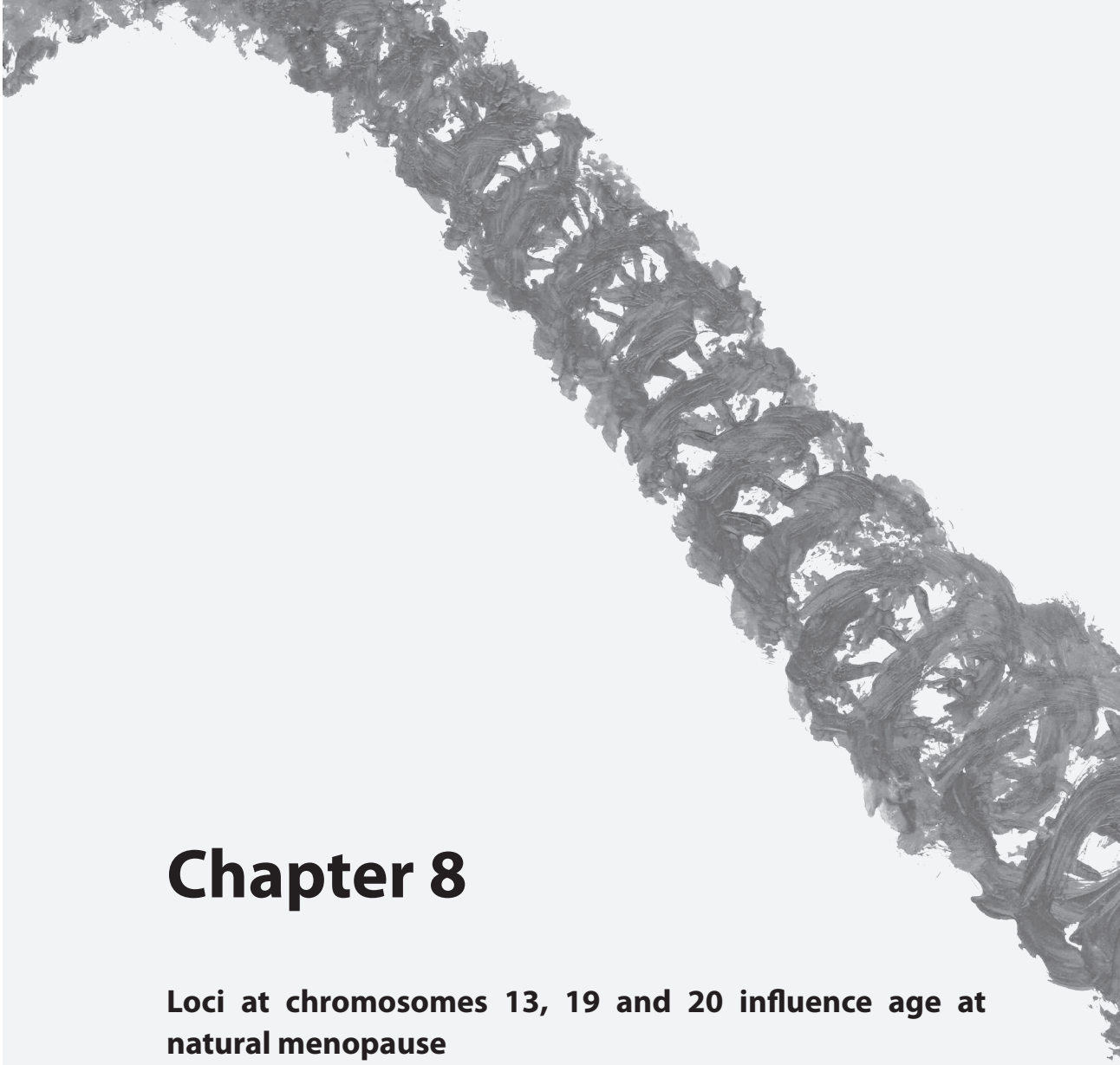
We are indebted to the participants and staff of the studies for their important contributions. **AGES:** The Age, Gene/Environment Susceptibility Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The authors thank the participants and the staff at the IHA for their contribution. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. **ARIC**: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions. **FHS:** This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially 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). A portion of this research utilized 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. The Framingham Heart Study phenotype-genotype analyses were supported by the National Institute of Aging (Genetics of Reproductive Life Period and Health Outcomes, R21AG032598). R01 AR/AG 41398 funded some of the menarche data collection. **HAPI Heart Study**: This work was supported by the National Institute of Health [grant number U01 HL72515]; the University of Maryland General Clinical Research Center [grant number M01 RR 16500); the Clinical Nutrition Research Unit of Maryland [grant number P30 DK072488]. **InCHIANTI:** The InCHIANTI

study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001-2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004-2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002);supported in part by the Intramural research program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland. **RSI and RSII**: The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810, and funding from the European Commision (HEALTH-F2-2008-201865 (GEFOS); HEALTH-F2-2008-35627 (TREAT-OA)). 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. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. **TwinsUK**: TwinsUK is supported by the Wellcome Trust from the Department of Health via 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 and King's College Hospital NHS Foundation Trust. 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. **Administrative:** The Longevity Consortium, funded by the National Institute of Aging, grant number U19 AG023122, provided administrative resources for this work.

REFERENCES

1. Marshall, W.A. & Tanner, J.M. Variations in pattern of pubertal changes in girls. *Arch Dis Child* **44**, 291-303 (1969).
2. Anderson, C.A. et al. A genome-wide linkage scan for age at menarche in three populations of European descent. *J Clin Endocrinol Metab* **93**, 3965-70 (2008).
3. Snieder, H., MacGregor, A.J. & Spector, T.D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J Clin Endocrinol Metab* **83**, 1875-80 (1998).
4. Towne, B. et al. Heritability of age at menarche in girls from the Fels Longitudinal Study. *Am J Phys Anthropol* **128**, 210-9 (2005).
5. Gajdos, Z.K. et al. Association studies of common variants in 10 hypogonadotropic hypogonadism genes with age at menarche. *J Clin Endocrinol Metab* **93**, 4290-8 (2008).
6. Harris, T.B. et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* **165**, 1076-87 (2007).
7. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
8. 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).
9. Dawber, T.R., Meadors, G.F. & Moore, F.E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41**, 279-81 (1951).
10. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).
11. 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).
12. 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).
13. Mitchell, B.D. et al. The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am Heart J* **155**, 823-8 (2008).
14. Ferrucci, L. et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc* **48**, 1618-25 (2000).
15. Epstein, M.P., Duren, W.L. & Boehnke, M. Improved inference of relationship for pairs of individuals. *Am J Hum Genet* **67**, 1219-31 (2000).
16. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
17. Andrew, T. et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-77 (2001).
18. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899-906 (2006).
19. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
20. 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).
21. Bosetti, C., Tavani, A., Negri, E., Trichopoulos, D. & La Vecchia, C. Reliability of data on medical conditions, menstrual and reproductive history provided by hospital controls.

J Clin Epidemiol **54**, 902-6 (2001).
22. Lettre, G. et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* **40**, 584-91 (2008).
23. Onland-Moret, N.C. et al. Age at menarche in relation to adult height: the EPIC study. *Am J Epidemiol* **162**, 623-32 (2005).
24. Weedon, M.N. & Frayling, T.M. Reaching new heights: insights into the genetics of human stature. *Trends Genet* **24**, 595-603 (2008).
25. Anderson, S.E., Dallal, G.E. & Must, A. Relative weight and race influence average age at menarche: results from two nationally representative surveys of US girls studied 25 years apart. *Pediatrics* **111**, 844-50 (2003).
26. Garn, S.M., LaVelle, M., Rosenberg, K.R. & Hawthorne, V.M. Maturational timing as a factor in female fatness and obesity. *Am J Clin Nutr* **43**, 879-83 (1986).
27. van Lenthe, F.J., Kemper, C.G. & van Mechelen, W. Rapid maturation in adolescence results in greater obesity in adulthood: the Amsterdam Growth and Health Study. *Am J Clin Nutr* **64**, 18-24 (1996).
28. Euling, S.Y. et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics* **121 Suppl 3**, S172-91 (2008).
29. Harris, M.A., Prior, J.C. & Koehoorn, M. Age at menarche in the Canadian population: secular trends and relationship to adulthood BMI. *J Adolesc Health* **43**, 548-54 (2008).
30. Wang, W., Zhao, L.J., Liu, Y.Z., Recker, R.R. & Deng, H.W. Genetic and environmental correlations between obesity phenotypes and age at menarche. *Int J Obes (Lond)* **30**, 1595-600 (2006).



Chapter 8

Loci at chromosomes 13, 19 and 20 influence age at natural menopause

Lisette Stolk*, Guangju Zhai*, Joyce BJ van Meurs, Michael MPJ Verbiest, Jenny A Visser, Karol Estrada, Fernando Rivadeneira, Frances M Williams, Lynn Cherkas, Panos Deloukas, Nicole Soranzo, Jules J de Keyzer, Victor JM Pop, Paul Lips, Corinne El Lebrun, Yvonne T van der Schouw, Diederick E Grobbee, Jacqueline Witteman, Albert Hofman, Huibert AP Pols, Joop SE Laven, Tim D Spector*, André G Uitterlinden*

* These authors contributed equally to this work

Nature Genetics ***41****(6)2009;645-647*

ABSTRACT

We conducted a genome-wide association study for age at natural menopause in 2,979 European women and identified six SNPs in three loci associated with age at natural menopause: chromosome 19q13.4 (rs1172822; –0.4 year per T-allele (39%); P = $6.3x10^{-11}$), chromosome 20p12.3 (rs236114; +0.5 year per A-allele (21%); P=$9.7x10^{-11}$) and chromosome 13q34 (rs7333181; +0.5 year per A-allele (12%); P=$2.5x10^{-8}$). These common genetic variants regulate timing of ovarian aging, an important risk factor for breast cancer, osteoporosis and cardiovascular disease.

INTRODUCTION

Menopause, the time of a woman's life when menstrual cycle ceases owing to depletion of the follicle pool, is a key event in reproductive aging. It influences a woman's well-being and is an important risk factor for several major age-related diseases including cardiovascular disease, breast cancer and osteoporosis[1]. Age at menopause averages around 50–51 years and ranges between 40 and 60 years of age[2]; twin studies have shown this variability to be genetically determined with heritabilities of 44%–65%[3-5]. Such genetic factors might regulate the size of the follicle pool and the rate of its depletion, and their identification could have biological and clinical applications. Typical for complex quantitative traits, genome-wide linkage studies of menopause have been unsuccessful, and candidate gene studies have mainly focused on the estrogen pathway and have had conflicting results[6]. This suggests that the apparent effect sizes for genetic variants are small and that the major causative loci have not been identified. Genome-wide association studies (GWAS) have proven successful in identifying common susceptibility genes with small effect sizes for many complex diseases and traits[7] and might be suitable to identify genetic factors involved in determining age at menopause.

METHODS

Description of cohorts

Stage 1 (genome-wide association) samples.

The discovery cohort (stage 1) for age at natural menopause was a sample of 2,419 postmenopausal women from the Rotterdam Study baseline, an ongoing prospective population-based cohort study, focused on chronic disabling conditions of the elderly in the Netherlands. In summary, 7,983 men and women aged 55 years or older, living in Ommoord, a suburb of Rotterdam, the Netherlands, were invited to participate[8]. Self-reported age at menopause was assessed by questionnaire[9] and was defined as 12 months after periods ceased. Self-reported age at natural menopause between 40 and 60 years (collected retrospectively) and genome-wide genotype data were available for n= 2,368 women. The second GWA analysis was run in a cohort of 611 women from the TwinsUK study, a population-based twin cohort study[10,11]. Age at menopause was defined as 12 months after periods ceased between the age of 40 and 60 (collected retrospectively), and women with hysterectomy, ovariectomy and those

on longterm HRT were excluded from the analysis.

Stage 2 (replication) samples.

The EPOS study is a cross-sectional study[12] and includes 5,896 pre-, peri- and post-menopausal women all living in the city of Eindhoven, the Netherlands. For 1,500 women DNA was available. Women who had hysterectomy, ovariectomy or were using hormones at time of menopause were excluded for this study. After follow-up, 1,080 women with natural menopause between 40 and 60 were included in this study, based on self-reported age at natural menopause. The menopausal age for this study was collected both retrospectively and prospectively.

The Rotterdam Study extension is an extension of the Rotterdam Study baseline with the same inclusion criteria that started in 1999[8]. In short, 3,011 subjects who had become 55 years of age or moved into the study district since the start of the original study in 1990, became participant of the extension cohort. Of the 3,000 participants (men and women) 785 were women who had experienced a natural menopause between 40 and 60 at the time of inclusion in this study and had DNA available. Participants were included based on self-reported age at natural menopause, retrospectively.

The subset of women from the PROSPECT-Frailty study was described previously[13]. In brief, 402 women who had experienced a natural menopause, had an intact uterus, had at least one intact ovary and had not used sex steroids after the reported date of last menstruation, were included in this subset and of whom 375 with age at natural menopause between 40 and 60 were used in this study.

The LASA (Longitudinal Aging Study Amsterdam) study[14] is a population-based cohort study, including 440 women of whom 320 experienced (self-reported, collected retrospectively) natural menopause between 40 and 60 years and were included in this study.

Sequenom iPLEX and Taqman Allelic Discrimination genotyping

Genomic DNA was extracted from samples of peripheral venous blood according to standard procedures. 1-2 ng genomic DNA was dispensed into 384-wells plates using a Caliper Sciclone ALH3000 pipetting robot (Caliper LS, Mountain View, CA, USA). Genotyping was done using Sequenom iPLEX genotyping and Taqman Allelic Discrimination.

Multiplex PCR assays were designed for the Sequenom iPLEX genotying using Assay Designer on the website (https://mysequenom.com/tools/genotyping/default.

aspx). For this, sequences containing the SNP site and at least 100bp of flanking sequence on either side of the SNP were used. Briefly, 2 ng genomic DNA was amplified in a 5 ul reaction containing 1 × Taq PCR buffer (Sequenom), 2 mM $MgCl_2$, 500 uM each dNTP, 100 nM each PCR primer, 0.5 U Taq (Sequenom). The reaction was incubated at 94°C for 4 minutes followed by 45 cycles of 94°C for 20 seconds, 56°C for 30 seconds, 72°C for 1 minute, followed by 3 minutes at 72°C. Excess dNTPs were then removed from the reaction by incubation with 0.3 U shrimp alkaline phosphatase (Sequenom) at 37°C for 40 minutes followed by 5 minutes at 85°C to deactivate the enzyme. Single primer extension over the SNP was carried out in a final concentration of between 0.731 uM and 1.462 uM for each extension primer (depending on the mass of the probe), iPLEX termination mix (Sequenom), 10x iPLEX Buffer Plus and iPLEX enzyme (Sequenom) and cycled using the following program; 94°C for 30 seconds followed by 94°C for 5 seconds, 5 cycles of 52°C for 5 seconds, and 80°C for 5 seconds, the last three steps were repeated 40 times, then 72°C for 3 minutes. The reaction was then desalted by addition of 6 mg clear resin (Sequenom) followed by mixing (15 minutes) and centrifugation (5 min, 3,000 rpm) to settle the contents of the tube. The extension product was then spotted onto a 384 well spectroCHIP using the SEQUENOM MassARRAY Nanodispenser RS1000 before analysis on the MassARRAY Compact System (Sequenom). Data collection was performed using SpectroACQUIRE 3.3.1.13 and clustering was called using TYPER Analyzer 4.0.3.18 (Sequenom). Additionally, to ensure data quality, genotypes for each subject were also checked manually.

Genotypes for rs4955755, rs494620, rs17153527, rs1433892, rs236114, and rs2326679 were generated using Taqman Allelic Discrimination (Applied Biosystems Inc., Foster City, CA, USA). All assays were available at www.appliedbiosystems.com as pre-designed assays. The PCR reaction mixture included 1-2 ng of genomic DNA in a 2 μl volume and the following reagents: FAM and VIC probes (200 nM), primers (0.9 uM), 2x Taqman PCR master mix (Applied Biosystems Inc., Foster City, CA, USA). Reagents were dispensed in the 384-well plates using the Deerac Equator NS808 (Deerac Fluidics, Dublin, Ireland). PCR cycling reactions were performed in 384 wells PCR plates in an ABI 9700 PCR system (Applied Biosystems Inc., Foster City, CA, USA) and consisted of initial denaturation for 15 minutes at 95°C, and 40 cycles with denaturation of 15 seconds at 95°C and annealing and extension for 60 seconds at 60°C. Results were analysed by the ABI Taqman 7900HT using the sequence detection system 2.22 software (Applied Biosystems Inc., Foster City, CA, USA).

Gene expression

We analysed top associated SNPs for relationships with expression levels of genes using the Sanger GENEVAR database. This provides gene expression data for EBV-transformed lymphoblastoid cell lines of all HapMap individuals[15] for whom SNP genotypes are known.

Statistical methods

Tests for normality and transformations were carried out using Ladder test in STATA. This test showed non-normal distribution of the Rotterdam Study baseline data, cubic-transformation of age at natural menopause was performed in SPSS15.0 as were linear regression tests to test for the association of six SNPs with cubic-transformed age at natural menopause.

In the Rotterdam Study baseline cohort SNPs with a minor allele frequency ≥ 5%, HWE $P>1x10^{-6}$, and a genotyping call-rate of at least 90% were selected (n=535,354 SNPs) and association analysis was performed using linear regression in PLINKv1.04[16]. The TwinsUK samples were genotyped using the Illumina HumanHap 300K array as described previously[17], and after quality control 317,818 SNPs were left for analysis. Because of the relatedness in the TwinsUK sample, we used a score test which takes this into account[18] in our analysis. The analysis was performed using GenABEL software[19].

Inversed variance fixed effects meta-analysis of summary data of the initial analysis of the Rotterdam Study baseline and the TwinsUK study was performed using METAL[20]. Only SNP information present in both studies was used, leading to a meta-analysis of 315,418 SNPs. For each study in stage 2 we calculated the association with age at natural menopause using linear regression in PLINK. Inverse variance fixed effects meta-analysis on summary statistics for the stage 2 studies was performed using METAL. The overall p-values, effect sizes, and standard errors for the combined stage 1 and 2 were also calculated using inverse variance fixed effects meta-analysis in METAL. Comprehensive meta-analysis 2.0 (Biostat, Englewood, NJ, USA) was used to calculate heterogeneity statistics for the genome-wide significant hits.

Odds ratios for menopause before the age of 50 were calculated using logistic regression in PLINKv1.04, meta-analysis was performed using the program comprehensive meta-analysis (Biostat, Englewood, NJ, USA). If heterogeneity existed ($I^2>25\%$) a random-effects model[21] was used for the analysis (rs236114), otherwise a fixed effects model (inverse variance method) was applied (rs7333181, rs1551562, rs1172822, rs2384687, and rs897798).

The effect of covariates on the association of the SNPs with age at natural

menopause was calculated using linear regression in SPSS15.0. Association of the SNPs with gene expression levels from the Sanger GENEVAR database were calculated using simple ANOVA or linear regression when a trend was present. A SNP with a p-value < $5x10^{-8}$ was considered genome-wide significant.

RESULTS

In this GWAS we used a two-stage design to identify previously unknown loci influencing age at menopause. We included women with self-reported natural age at menopause (defined as 12 months without regular periods) between 40 and 60 years, excluding those with hysterectomy, uni- or bilateral ovariectomy, menopause induced by irradiation or occurring after stopping the contraceptive pill, or those currently using hormone replacement therapy.

In stage 1 we genotyped 2,368 women of the Rotterdam Study baseline with the Illumina HumanHap 550v3 Beadarray. After quality control, 535,354 SNPs were left for analysis. The genomic inflation factor (λ) was 1.01669 for this analysis, indicating no population stratification, so we based our results on the uncorrected p-values. The strongest association signals were found for rs2151145 ($P=5.3x10^{-6}$) on chromosome 9, rs236114 ($P=5.6x10^{-6}$) on chromosome 20 and rs1172822 ($P=6.3x10^{-6}$) on chromosome 19 (Figure 1, upper panel).

We combined the results from the Rotterdam Study baseline with GWA data from the TwinsUK study. A total of 611 women with natural menopause using the same definitions and exclusions as above were genotyped with the Illumina HumanHap 300K beadarray, and after quality control 317,818 SNPs were left for analysis. After adjusting for relatedness and genomic control, we did not observe any genome-wide significant signals in this study (Figure 1, middle panel). Because of the different study designs we conducted meta-analysis on summary statistics of the two studies using METAL on 315,418 SNPs common to both cohorts (2,979 women), but we did not observe any genome-wide significant SNPs (Figure 1, lower panel). From this meta-analysis, all SNPs with $P \leq 1x10^{-4}$, corresponding to 32 SNPs from 24 loci (with five loci having multiple significant SNPs), were followed up in stage 2.

Twentyfour SNPs were genotyped using Sequenom iPLEX genotyping and seven SNPs using Taqman allelic discrimination (Applied Biosystems) in 2,560 samples of four additional cohorts of postmenopausal females of European ancestry (Table 1); one of the SNPs (rs11786333) failed genotyping. For the remaining 31 SNPs, we

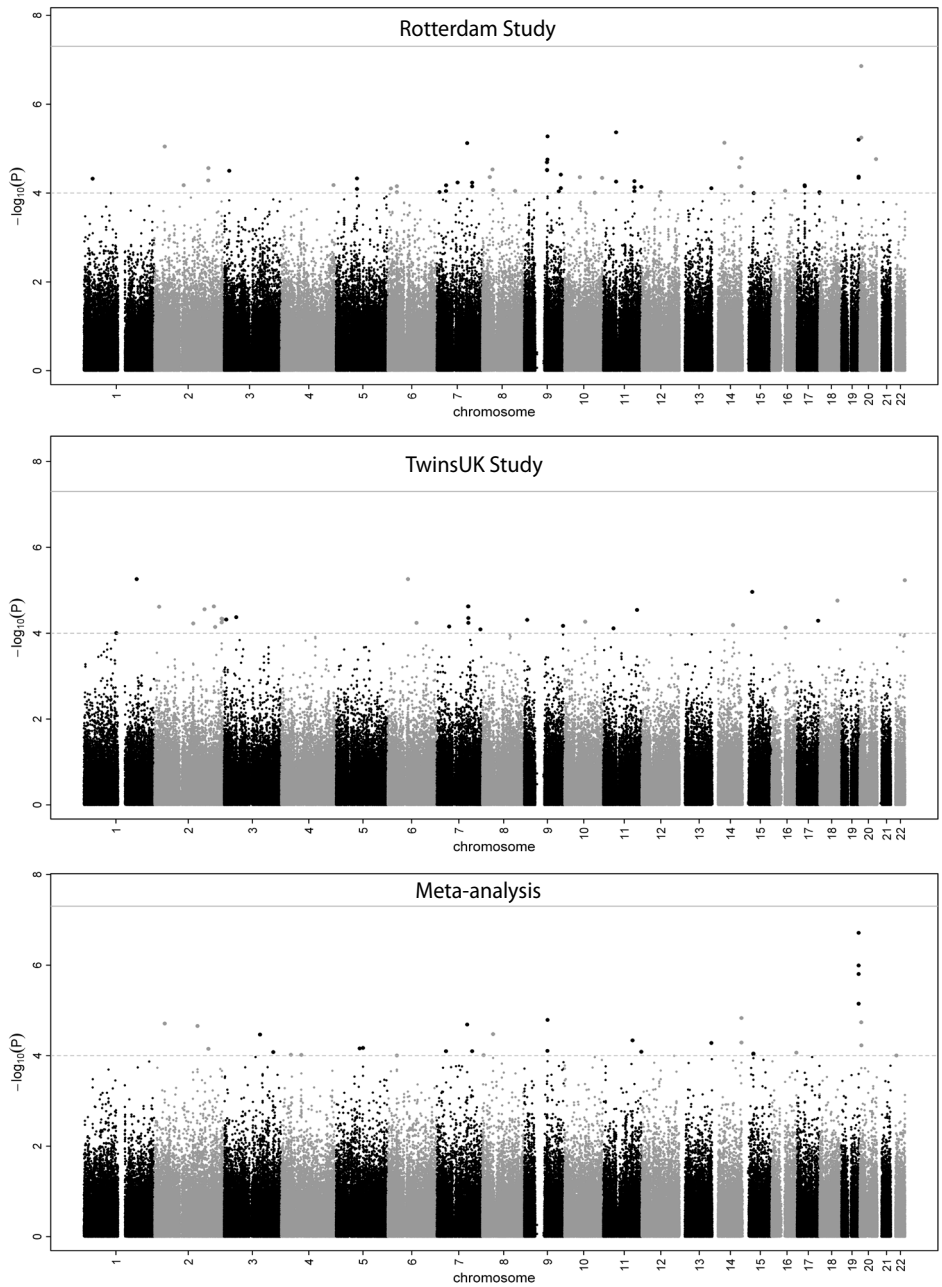


Figure 1. P-values for age at natural menopause in the two GWAS and the stage 1 meta-analysis. Manhattan plot for p-values for the Rotterdam Study baseline (upper panel), the TwinsUK study (middle panel), and the meta-analysis of stage 1 (lower panel). The lid grey line indicates the genome-wide Bonferroni adjusted threshold of $P=5x10^{-8}$, and the dashed grey line indicates $P=1x10^{-4}$.

Table 1. Sample size and mean age at natural menopause for each study.

Study	Number of subjects	Age (years)	Age at natural menopause (yrs)	Stage	Genotyping method
Rotterdam Study baseline	2,368	70.3 (9.6)	49.9 (4.0)	1	Illumina HumanHap550Kv3.0
TwinsUK	611	55.6 (6.7)	48.5 (3.8)	1	Illumina HumanHap300K
EPOS	1,006	51.4 (2.4)	50.9 (2.4)	2	Sequenom / Taqman
Rotterdam Study extension	785	65.1 (8.4)	50.2 (4.5)	2	Sequenom / Taqman
PROSPECT-Frailty	375	66.5 (3.9)	49.5 (4.5)	2	Sequenom / Taqman
LASA	320	75.7 (6.5)	49.1 (5.0)	2	Sequenom / Taqman
Total	5,465				

Data are mean (SD).

calculated combined P-values, betas and standard errors using inverse variance fixed-effects meta-analysis and identified six common SNPs that were genome-wide significant in the combined stage 1 and 2 analysis (Table 2). Four SNPs on chromosome 19 were significant: rs1172822 (MAF=0.39), $P=6.3x10^{-11}$, beta=–0.39 year per T-allele (SE=0.060); rs2384687 (MAF=0.40), $P=1.4x10^{-10}$, beta=0.38 year per C-allele (SE=0.059); rs1551562 (MAF=0.25), $P=1.0x10^{-9}$, beta=0.428 year per G-allele (SE=0.070); and rs897798 (MAF=0.48), $P=3.9x10^{-8}$, beta=0.31 year per G-allele (SE=0.056). These four SNPs are likely to report the same signal because the linkage is high ($D'>0.92$, $r^2>0.5$, Figure 2). On chromosome 20, rs236114 (MAF=0.21) was genome-wide significantly associated with age at natural menopause ($P=9.7x10^{-11}$, beta=0.495 year per A-allele (SE=0.077)). Furthermore, on chromosome 13 rs7333181 (MAF = 0.12) was genome-wide significant: $P=2.50x10^{-8}$, beta=0.520 (SE=0.093). The six genome-wide significant hits showed no heterogeneity ($I^2<25\%$), so fixed effects models were used.

In addition, we estimated the risk for menopause before the age of 50 by allele of the six genome-wide significant SNPs (Figure 2). We conducted fixed-effects meta-analysis for SNPs not showing heterogeneity (rs7333181, rs1551562, rs1172822, rs2384687, rs897798), and random effects meta–analysis for rs236114, for which I^2 was 31%. This meta-analysis showed that the A allele of rs1172822 is associated with a

Table 2. The six genome-wide significant SNPs and association with age at menopause.

					Stage 1	Overall meta-analysis		
SNP	Chr	Position	Minor allele	MAF	P	P	β	SE
rs1172822	19q13.4	60511657	T	0.39	$1.94x10^{-7}$	$6.28x10^{-11}$	0.391	0.060
rs236114	20p12.3	5883385	A	0.21	$1.83x10^{-5}$	$9.71x10^{-11}$	0.495	0.077
rs2384687	19q13.4	60523000	C	0.40	$1.57x10^{-6}$	$1.39x10^{-10}$	-0.381	0.059
rs1551562	19q13.4	60506693	G	0.25	$1.02x10^{-6}$	$1.04x10^{-9}$	-0.428	0.070
rs7333181	13q34	111019298	A	0.12	$5.26x10^{-5}$	$2.50x10^{-8}$	0.520	0.093
rs897798	19q13.4	60525566	G	0.48	$7.13x10^{-6}$	$3.91x10^{-8}$	-0.308	0.056

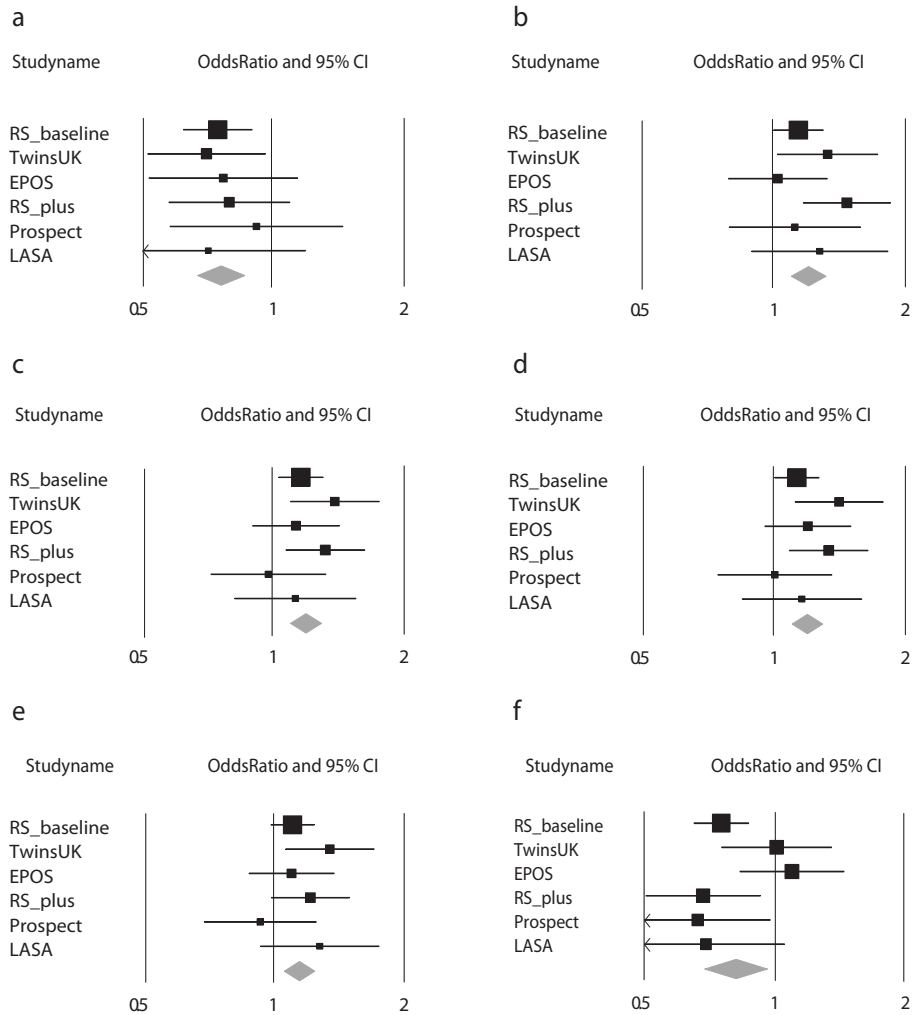


Figure 2. Meta-analysis of risk for early menopause (<50 years) by genotype for the six genome-wide-significant hits. (a) rs7333181 on chromosome 13, (b) rs1551562 on chromosome 19, (c) rs1172822 on chromosome 19, (d) rs2384687 on chromosome 19, (e) rs897798 on chromosome 19, (f) rs236114 on chromosome 20.

19% increased risk for natural menopause before 50 years (OR=1.19, 95%CI[1.09–1.29]; P=6.2x10^{-5}). The other SNPs on chromosome 13, 19 and 20 showed a similar increase or decrease in risk.

The initial analysis was not adjusted for covariates such as age, body mass index, smoking, age at menarche, parity and use of oral contraceptives and female hormones. To rule out an effect of these covariates on the association of the genome-

wide–significant hits, we carried out adjusted linear regression of these SNPs in the Rotterdam Study baseline cohort (Table 3). None of the previously found associations was affected by the adjustment for these covariates, indicating that the effect of the SNP occurs directly on age at natural menopause and not via one of the covariates. We calculated the total explained variance in age at natural menopause for these SNPs in the combined replication studies to be 1.1% (range 0.1–0.5% per SNP).

Table 3. Adjustment for covariates. Model 1 is the uncorrected model, model 2 is corrected for age, bmi, smoking, age at menarche, parity, use of oral contraceptives, and hormone replacement therapy.

	Model 1		Model 2	
SNP	**β**	**P**	**β**	**P**
rs7333181	0.442	$1.75x10^{-2}$	0.448	$1.52x10^{-2}$
rs1551562	-0.478	$6.09x10^{-4}$	-0.47	$6.86x10^{-4}$
rs1172822	-0.474	$1.28x10^{-4}$	-0.479	$9.76x10^{-5}$
rs2384687	-0.385	$1.63x10^{-3}$	-0.398	$1.05x10^{-3}$
rs897798	-0.396	$1.03x10^{-3}$	-0.407	$6.73x10^{-4}$
rs236114	0.667	$6.61x10^{-6}$	0.684	$3.22x10^{-6}$

We then conducted fine mapping of these signals using meta-analysis of imputed data of the stage 1 studies, and found three SNPs on chromosome 13 with more or equal significance as rs7333181, two SNPs on chromosome 19 and one on chromosome 20 with higher significance compared to the previously reported SNPs (Table 4). For all three loci, the imputed SNPs are located in the same linkage disequilibrium (LD) block as the genome-wide-significant SNPs.

Table 4. P-values, r2 and D′ results of fine-mapping of the three genome-wide loci.

Locus	SNP	Fine-mapping SNPs	MAF	D'	r^2	P	β	SE	Location
chr 13	rs7333181	rs7333181	0.12	-	-	$5.26x10^{-5}$	0.623	0.176	220kb 3' of C13orf16; 260kb 3' of ARHGEF7
		rs1361542	0.12	1.00	0.95	$4.00x10^{-4}$	0.625	0.177	220kb 3' of C13orf16; 260kb 3' of ARHGEF7
		rs1163623	0.12	1.00	1.00	$4.00x10^{-4}$	-0.619	0.176	220kb 3' of C13orf16; 260kb 3' of ARHGEF7
		rs1756091	0.12	1.00	1.00	$4.00x10^{-4}$	0.619	0.176	220kb 3' of C13orf16; 260kb 3' of ARHGEF7
chr 19	rs1172822	rs1172822	0.39	-	-	$1.94x10^{-7}$	-	0.117	Intron 17 of BRSK1
		rs4806660	0.39	1.00	0.97	$1.00x10^{-7}$	0.479	0.090	Intron 7 LOC284471
		rs11668344	0.37	0.97	0.89	$1.00x10^{-7}$	0.484	0.092	Intron 2 LOC284471
chr 20	rs236114	rs236114	0.21	-	-	$1.83x10^{-5}$	0.642	0.141	Intron 4 of MCM8
		rs16991615	0.07	1.00	0.36	$3.00x10^{-7}$	1.091	0.214	Exon 9 Glu-Lys

MAF: Minor Allele Frequency, based on Rotterdam Study I; D′ and r^2 based on CEU HapMap population; P: p-value; β: beta.

DISCUSSION

The linkage block in which the four genome-wide significant SNPs on chromosome 19 are present covers parts of two genes. One of these genes is *BRSK1*, which codes for an AMP-activated protein kinase (AMPK)-related kinase. It is highly expressed in human forebrain where it is required for neuronal polarization[22], but also moderately expressed in mammalian ovaries according to gene atlas[23]. While *Brsk1* knockout mice do not show any obvious phenotype[22], particular phenotypes related to reproductive aging (such as reduced ovarian follicle counts, uterine hypertrophy and elevated FSH-levels) have not been investigated as far as we know.

Interestingly, the downstream targets of BRSK1 are several members of the family of AMPK-related kinases phosphorylate tau, a microtubule-associated protein regulating stability of the microtubule network[22]. This includes the maternal embryonic leucine zipper kinase (MELK; chr 9) that is highly expressed in spermatogonia (testis) and oocytes (ovaries)[24,25]. BRSK1 is specifically activated by phosphorylation, together with 12 other AMPKs including MELK, through the serine/threonine protein kinase 11, named STK11 or LKB1[26]. LKB1 (19p13.3) is viewed as a master regulator of cell polarity (by regulating cytoskeletal dynamics) being the only protein with this activity, and is expressed in mouse oocytes[27]. Mutating *LKB1* affects epithelial, neuronal and oocytes polarity thereby influencing cell growth.

The other gene located in the linkage block is *TMEM224*, this gene encodes a transmembrane protein of which the function is not yet clear. The rs7246479 SNP is located in exon 8 of *TMEM224* and encodes a hypothetical amino acid change (Leu199Phe). Therefore we can not exclude involvement of this protein in the onset of menopause, but much additional research will be needed to establish such a role.

The four genome-wide significant genes are also located in the 5' region of several genes, of which *SUV420H2* is the most close to the LD-block. SUV420H2 is a lysine-N-methyltransferase responsible for trimethylation of H4-K20, which is involved in gene silencing[28]. Cells deficient for Suv420h2 or both Suv420h2 and its homologue Suv420h1, show elongation of telomeres and increased telomere recombination[29]. Elongation of telomeres is associated with reduced fecundity and fertility in Drosophila[30], whereas in a small study in humans (n=30) an opposite effect was shown in that longer telomeres were associated with longer reproductive lifespan[31]. These findings point towards a possible effect of SUV420H2 in the onset of menopause. However, these are small studies and the disconcordant results in Drosophila and humans need to be elucidated before any conclusions can be drawn.

rs7333181 on chromosome 13 is located in a copy-number variation region (CNV), detected by Redon et al. in one individual of 270 HapMap subjects[32]. As shown by Stranger et al. CNVs can affect gene expression of genes located several kb away from this CNV[33], however, if this CNV is also affecting gene expression still has to be elucidated. The gene located most closely to this SNP is *ARHGEF7* (260kb 3' of the gene). In mice *Arhgef7* is expressed in ovaries, with increasing expression levels during maturation of the follicles[34]. A role for ARHGEF7 in ovarian ageing has not been described, however it plays a role in cell proliferation through phosphorylation of FOXO3a[35]. Interestingly, *Foxo3a* knockout mice are infertile due to early depletion of the follicle pool[36-38]. Furthermore, mutations in *FOXO3a* were identified in premature ovarian failure (POF) patients[39]. These findings may suggest that ARHGEF7 may affect ovarian ageing through regulation of, amongst others, FOXO3a.

The chromosome 20 SNP is located in an intron of the *MCM8* (minichromosome maintenance complex component 8) gene. The more significantly associated SNP from the imputed data is a non-synonymous SNP in exon 9 of this gene (E341K), and could influence the protein structure or function of MCM8. It is involved in DNA replication as are the other MCM proteins[40]. In mice *Mcm8* expression in the ovaries is highest in the primordial follicles and decreases with maturation of the follicles[34]. Lower levels of MCM8 lead to lower levels of DNA replication[41]. Whether MCM8 plays a role in ovarian ageing, either at the hypothalamic-pituitary level or at the ovarian level, remains to be studied.

ACKNOWLEDGEMENTS

The authors are very grateful to the study participants and staff from the Rotterdam Study and the TwinsUK, EPOS, PROSPECT-Frailty and LASA studies.

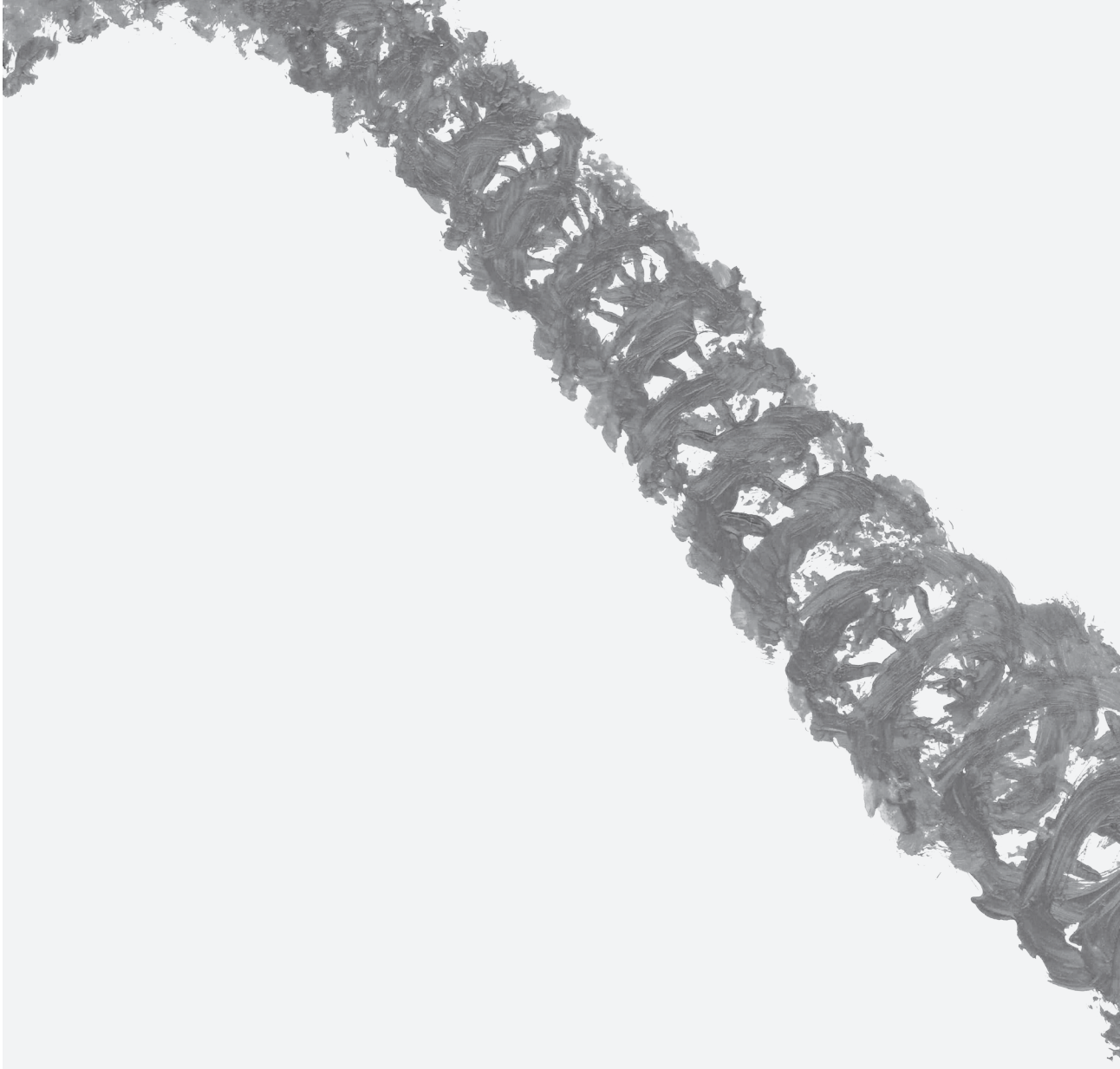
This study was funded by the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA), Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), Research Institute for Diseases in the Elderly (014-93-015; RIDE2) and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank P. Arp, M. Jhamai, M. Moorhouse, M. Verkerk and S. 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. TwinsUK is supported by the Wellcome Trust from the Department of Health via 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 and King's College Hospital NHS Foundation Trust. The LASA study is largely funded by the Ministry of Health, Welfare and Sports of The Netherlands. The PROSPECT-Frailty study was funded by The Netherlands Organization for Health Research and Development (ZON) no. 2100.0011. We thank K. Lunetta for helpful discussion.

REFERENCES

1. Burger, H.G. The menopausal transition. *Baillieres Clin Obstet Gynaecol* **10**, 347-59 (1996).
2. te Velde, E.R., Dorland, M. & Broekmans, F.J. Age at menopause as a marker of reproductive ageing. *Maturitas* **30**, 119-25 (1998).
3. Murabito, J.M., Yang, Q., Fox, C., Wilson, P.W. & Cupples, L.A. Heritability of age at natural menopause in the Framingham Heart Study. *J Clin Endocrinol Metab* **90**, 3427-30 (2005).
4. Snieder, H., MacGregor, A.J. & Spector, T.D. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. *J Clin Endocrinol Metab* **83**, 1875-80 (1998).
5. van Asselt, K.M. et al. Heritability of menopausal age in mothers and daughters. *Fertil Steril* **82**, 1348-51 (2004).
6. Kok, H.S. et al. No association of estrogen receptor alpha and cytochrome P450c17alpha polymorphisms with age at menopause in a Dutch cohort. *Hum Reprod* **20**, 536-42 (2005).
7. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).
8. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
9. Weel, A.E. et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. *J Clin Endocrinol Metab* **84**, 3146-50 (1999).
10. Andrew, T. et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-77 (2001).
11. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899-906 (2006).
12. Smeets-Goevaers, C.G. et al. The prevalence of low bone mineral density in Dutch perimenopausal women: the Eindhoven perimenopausal osteoporosis study. *Osteoporos Int* **8**, 404-9 (1998).
13. Lebrun, C.E. et al. Arterial stiffness in postmenopausal women: determinants of pulse wave velocity. *J Hypertens* **20**, 2165-72 (2002).
14. Knipscheer, C.P., Dykstra, P.A., van Tilburg, T.G. & de Jong-Gierveld, J. [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**, 110-9 (1998).
15. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet* **1**, e78 (2005).
16. 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).
17. Richards, J.B. et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
18. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *Am J Hum Genet* **81**, 913-26 (2007).
19. 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).
20. Abecasis, G.R. & Willer, C. METAL. *http://www.sph.umich.edu/csg/abecasis/metal/* (2007).
21. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177-88 (1986).

22. Kishi, M., Pan, Y.A., Crump, J.G. & Sanes, J.R. Mammalian SAD kinases are required for neuronal polarization. *Science* **307**, 929-32 (2005).
23. GeneAtlas. http://genatlas.medecine.univ-paris5.fr/.
24. Lizcano, J.M. et al. LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *Embo J* **23**, 833-43 (2004).
25. Thelie, A. et al. Differential regulation of abundance and deadenylation of maternal transcripts during bovine oocyte maturation in vitro and in vivo. *BMC Dev Biol* **7**, 125 (2007).
26. Bright, N.J., Carling, D. & Thornton, C. Investigating the regulation of brain-specific kinases 1 and 2 by phosphorylation. *J Biol Chem* **283**, 14946-54 (2008).
27. Szczepanska, K. & Maleszewski, M. LKB1/PAR4 protein is asymmetrically localized in mouse oocytes and associates with meiotic spindle. *Gene Expr Patterns* **6**, 86-93 (2005).
28. Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-62 (2004).
29. Benetti, R. et al. Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J Cell Biol* **178**, 925-36 (2007).
30. Walter, M.F. et al. Effects of telomere length in Drosophila melanogaster on life span, fecundity, and fertility. *Chromosoma* **116**, 41-51 (2007).
31. Aydos, S.E., Elhan, A.H. & Tukun, A. Is telomere length one of the determinants of reproductive life span? *Arch Gynecol Obstet* **272**, 113-6 (2005).
32. Redon, R. et al. Global variation in copy number in the human genome. *Nature* **444**, 444-54 (2006).
33. Stranger, B.E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-53 (2007).
34. Pan, H., O'Brien M, J., Wigglesworth, K., Eppig, J.J. & Schultz, R.M. Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development in vitro. *Dev Biol* **286**, 493-506 (2005).
35. Chahdi, A. & Sorokin, A. Endothelin-1 couples betaPix to p66Shc: role of betaPix in cell proliferation through FOXO3a phosphorylation and p27kip1 down-regulation independently of Akt. *Mol Biol Cell* **19**, 2609-19 (2008).
36. Brenkman, A.B. & Burgering, B.M. FoxO3a eggs on fertility and aging. *Trends Mol Med* **9**, 464-7 (2003).
37. Castrillon, D.H., Miao, L., Kollipara, R., Horner, J.W. & DePinho, R.A. Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. *Science* **301**, 215-8 (2003).
38. Liu, L. et al. Infertility caused by retardation of follicular development in mice with oocyte-specific expression of Foxo3a. *Development* **134**, 199-209 (2007).
39. Watkins, W.J. et al. Mutational screening of FOXO3A and FOXO1A in women with premature ovarian failure. *Fertil Steril* **86**, 1518-21 (2006).
40. Johnson, E.M., Kinoshita, Y. & Daniel, D.C. A new member of the MCM protein family encoded by the human MCM8 gene, located contrapodal to GCD10 at chromosome band 20p12.3-13. *Nucleic Acids Res* **31**, 2915-25 (2003).
41. Maiorano, D., Lutzmann, M. & Mechali, M. MCM proteins and DNA replication. *Curr Opin Cell Biol* **18**, 130-6 (2006).



Chapter 9

General Discussion

Estrogens, the female sex hormones, are of vital importance to bone, the heart and other organs. This can be seen in women who, after menopause, have an increased risk for osteoporosis, cardio-vascular disease, and cancer increase, most likely due to the depletion of estrogens after menopause. In this thesis genetic variants were studied to identify the role of estrogens in a number of so-called estrogen-related traits, like sex hormone levels, exposure time to estrogens, based on menarcheal and menopausal age, and two bone phenotypes: bone mineral density (BMD) and fracture (fx) risk. The first objective of this thesis was to study genetic variations in genes of the estrogen synthesis, metabolism and signalling for their association with sex steroid hormone-levels, age at menopause, BMD and vertebral fracture (vert fx) risk. The second objective was to examine the effect of common genetic variants on sex steroid levels, age at menarche and age at menopause in a hypothesis-free setting using a novel technology; the Genome-Wide Association Study (GWAS). In the current chapter the main findings of this thesis are brought together and placed in a broader context. The findings of candidate gene studies and GWAS are compared, and both analysis approaches are discussed. Furthermore, recent findings for genetic variants in the estrogen receptor alpha gene are discussed, and at the end several recommendations for future research on the GWAS top hits are presented.

Genetics of complex traits

In this thesis several genetic association approaches were used to study complex traits. In chapter 2 to 5 a hypothesis-based candidate gene/pathway approach was used. For this approach candidate genes are chosen based on prior knowledge of the function of the gene and the biology of the trait that is studied. Due to technological progress in genotyping techniques over the course of this thesis it became possible to conduct hypothesis-free GWAS for the same complex traits. This approach is not based on *a priori* knowledge about a gene or trait, but screens all common genetic variation for an association with the trait.

Both the candidate gene and GWAS approach are based on the common disease, common variant (CD/CV) hypothesis[1], which predicts that a set of common genetic variants with small effects is responsible for diseases occurring commonly in a population. This hypothesis originates from the idea that the continuous range of phenotypic variation in populations cannot be explained by rare variants with large effects. This idea was already posted over a century ago. Recent GWAS results are clear

evidence for this CD/CV hypothesis. To date over 400 GWAS for over 100 phenotypes have been performed, most of them showing reproducible significant associations[2,3]. However, for most of the phenotypes these GWAS results only explain a small part of the total genetic variance that was expected based on earlier heritability studies[4,5]. This leaves room for an alternative hypothesis: the common disease, rare variant (CD/RV) hypothesis[6]. The rationale behind this is that evolutionary very recently arising rare variants influencing disease risk are more likely to have a functional effect on the disease compared to older common variants[6]. A number of studies on candidate genes that showed multiple rare variants in one gene to contribute to the risk for several diseases[7], support this CD/RV hypothesis. The genetic etiology of common complex diseases or traits is likely to be a combination of both common and rare variants and therefore, both need to be studied further. However, at this moment technological and statistical methods to identify rare variants contributing to a trait or disease in a hypothesis-free manner cannot be applied on a large scale, yet. Therefore, in this thesis all studies were based only on the CD/CV hypothesis.

Genetic variation of the estrogen pathway and estrogen related traits

As shown in chapter 2 to 5 of this thesis, plasma levels of sex steroids, lumbar spine BMD (lsBMD) and fracture risk are all traits influenced by genetic variations in estrogen pathway genes. In this thesis several genes have been studied in a single candidate gene analysis, while we also performed a candidate pathway analysis of 61 genes. We showed that age at natural menopause is not influenced by estrogen-related genes in our study although power to detect significant associations was low.

In chapter 2, a candidate gene study on fine-mapping of previously studied SNPs (*Pvu*II and *Xba*I) in the the estrogen receptor alpha (*ESR1*) gene was presented. This study showed a significant association with vertebral fracture risk and plasma estradiol levels in postmenopausal women for a number of SNPs in the *Pvu*II-*Xba*I LD-block. It was also examined whether haplotype analysis adds additional genetic information. We were able to select a certain haplotype and in this way we could narrow down the number of possible causal SNPs. This haplotype analysis could contribute to finding the causal variant in 2 ways; either the causal allele is genotyped and present on one of the haplotypes, or the causal variant is highly linked to one of the haplotype alleles.

In chapter 3 a proline insertion-deletion polymorphism in the ERα co-activator retinoblastoma interacting zinc finger 1 (*RIZ1*) gene is examined for its association with

BMD and fx risk, and we did not observe an association. In chapter 4 a candidate gene study on catechol-O-methyltransferase (*COMT*) was presented showing a significant association of the Met158Val polymorphism with fragility fracture risk in elderly men, but not with BMD or vert fx risk. The polymorphisms that were studied for *RIZ1* and *COMT* were non-synonymous (amino-acid changing) variants, affecting the function of the protein *in vitro*[8-10]. Both variants were also associated with bone endpoints in men and women before[9,11,12].

In the candidate pathway analysis examining 61 gene involved in estrogen synthesis, metabolism and signalling in chapter 5 we showed that *ESR1* is one of the genes significantly influencing both lsBMD and vert fx risk in elderly men, but not plasma estradiol levels or levels of other sex hormones. The *RIZ1* and *COMT* genes, however, did not show significant associations in the candidate pathway analysis. This indicates that probably genetic variation in other genes from the estrogen pathway are more important for the studied phenotypes as *RIZ1* and *COMT*.

GWAS of estrogen related traits

Sex hormone levels, such as estradiol, testosterone and levels of sex hormone binding globuline are the the underlying biological traits (so-called endophenotypes)[13] of several estrogen-related diseases. The estrogen exposure time in women is defined as the time between the start (age at menarche) and end (age at menopause) of menstrual cycles, and this is also an important determinant for estrogen-related traits. Therefore, knowing the genetic background of sex hormone levels, age at menarche, and age at natural menopause is useful in the study of estrogen-related traits and diseases. Using GWAS in the Rotterdam Study I (RSI) we tried to identify genome-wide genetic variants influencing sex steroid levels (chapter 6), but this did not result in reliable significant results, due to low power. These power issues were also evident in the GWAS of menarcheal age or menopausal in the Rotterdam Study. Only combining menarche or menopause GWAS results from several studies, yielded genome-wide significant (gws) associations. This shows that for GWAS large sample sizes are needed to find true significant signals.

Age at menarche and age at menopause reflect the opposite limits of reproductive life span and intuitively, could share the same genetic background. However, when looking more closely to the two traits, the physiology behind these phenotypes is completely different. Menarche is a puberty measure reflecting the

maturation of female reproduction, while menopause is a measure for ovarian ageing. This is, indeed, supported by our GWAS results, showing two chromosomal loci associated with menarche (at chromosome 6q21 and 9q31.2), and three completely different chromosomal loci associated with age at menopause (at chromosome 13q34, 19q13.4 and 20p12.3). Of the two menarche loci, 6q21 is also associated with height, which is in line with the fact that later menarche results in a larger height at adult age[14]. The mechanism is probably mediated through later exposure to estrogens resulting in later closure of the growth plates leading to longer longitudinal growth. The three menopause hits are located in three gene regions that previously have not been associated with ovarian ageing, yet, the exact genes underlying the effect on menopause are not known. For one region on chromosome 13 there was no gene in the region, on chromosome 19 there are several genes located in the regions of which none was suggested in ovarian ageing before, and in the chromosome 20 locus only one gene was located, but with an unknown function at this moment. When comparing the top hits ($P<5x10^{-8}$) of age at menarche and age at menopause no overlap in the signals was found. However, there is overlap for one SNP when comparing all signals with $P<1x10^{-4}$ in meta-analyses for age at menarche of ~60,000 women and age at menopause of ~30,000 women in a large consortium of studies (K. Lunetta, L. Stolk, P. Sulem, J. Perry, C. Elks, manuscripts in preparation). It is not clear whether this overlap is due to chance or if it is a real functional overlap of the signals, and this is under further investigation.

Candidate genes versus GWAS

The results of the candidate gene studies and GWAS in this thesis are difficult to compare. One reason is the difference in significance threshold due to multiple testing adjustments. For single SNP associations the p-value threshold for nominal significance is 0.05, while for GWAS it is $5x10^{-8}$. For example, the association of the *ESR1* SNP *Xba*I (rs9340799) with estradiol levels had a p-value of 0.005 in the candidate gene study and will not pass the significance threshold in a GWAS with the 3,574 postmenopausal women that were included in chapter 2 of this thesis. For a SNP explaining 0.2% of all genetic variance of a trait, approximately 3,900 subjects are required for a significant association in a candidate gene study (P=0.05). While around 19,000 subjects are required for a significant association in a GWAS ($P=5x10^{-8}$). This holds for the majority of the complex traits and diseases, not only those studied in

this thesis, because for common variants we expect small effect sizes explaining only a very small part of the variance[5].

The significance threshold for GWAS is based on the estimated total number of independent common variants, which is 1,000,000. This threshold is used in almost all GWAS independent of the number of SNPs that is studied. The p-value threshold of a candidate gene study could also be increased to $5x10^{-8}$, because the studied SNPs are included in these 1,000,000 independent common variants. As a result, the sample size in a candidate gene study has to be increased to the level of GWAS to have enough statistical power. In this way the comparison of the candidate gene studies and GWAS results will be less complicated, and it will be more clear what the true effect of the candidate gene, or candidate SNP is on the explained variance.

Although GWAS needs much more samples to find reliable significant associations for quantitative traits, this method has been more successful in identifying genetic variants influencing complex traits and diseases in the past three years compared to candidate gene studies in the twenty years before[15]. Until now, hundreds of associations of common variants with over 100 traits and diseases have been reported [16]. The median frequency of the risk alleles was over 30% and effect sizes are very modest (OR around 1.3). Of the reported associations, only 13% was located in exons, or untranslated regions (UTRs), 45% was intronic and the rest was located intergenic[2]. Many GWAS associations, therefore, only identified chromosomal loci rather than genes involved in the trait or disease. For this reason GWAS is the first step in the discovery of new biological pathways underlying complex traits and diseases. Understanding the underlying pathways could eventually lead to new drug targets, although this could take a long time. GWAS hits for quantitative traits will lead to better understanding of the biology underlying these traits. Depending on the relevance of the trait for a disease this could also contribute to better understanding of the etiology of diseases.

While GWAS are highly successful in the identification of new genetic loci involved in traits and diseases, the explained variance of these loci together is low for many traits (between 2-5%)[5]. This could have several reasons. First: the lack of power for many GWAS, due to low sample size, which could be too low to detect more common variants contributing to the explained genetic variance. This is supported by the recent findings for height in the GIANT consortium, where in 135,000 subjects 210 loci explaining 14% of the variance were found[17].

Second: to achieve enough samples to have enough power to detect small effects in GWAS, results for many studies are combined in a meta-analysis. These

different studies all have a different environmental background, and the environment influencing traits could differ a lot between the studies. For example for BMI one of the major environmental influences is that of diet. Even when studying only populations from Northern European descent (from the USA and northern Europe) this could influence the genetic explained variance that is found, because the diet in the USA differs from the diet in Europe, but even within Europe there are differences in diet between countries.

Third: it could be that for many traits and diseases there are no functional common variants present in the most important genes. One could imagine that genetic variants that are deleterious for life or reproduction are subject to negative selection and therefore, never become common in a population[6,18]. Shorter height is not likely to have direct deleterious effects on reproduction or life, so it is more likely to find a large explained variance of the common variants for height. For menopausal age, which effects reproductive lifespan, probably most of the common variants have very small effects since, mutations with large effects are not tolerated and they are subject to negative selection.

Where is the "missing heritability"?

For most complex traits and diseases not more then 5% of the heritability is explained by the common variants discovered by GWAS, the remaining is termed the "missing heritability" or "genetic dark matter"[4]. But where to look for this? It is known that with the current SNP arrays around 90% of all common genetic variation is covered, so part of the "missing heritability" will be in the 10% that is not studied yet. This will probably be a smaller part than that is already identified in the current GWAS. Another part of the missing heritability is to be found in rare variants, which are genetic variants that are only present in a very small number of subjects in a population. In studies examining rare variants the frequency of multiple variants in one gene in the cases is compared to the controls. This approach was used for genes known to be involved in blood pressure[19], HDL-cholesterol[20,21], and type 1 diabetes[22]. These rare variants may have a higher penetrance compared to the previously identified common variants for these traits (penetrance is a measure for the proportion of carriers of a genetic variant that develop a disease). The explained variance, however, is difficult to calculate, because the single rare variants are very infrequent in the population. Therefore, the explained variance for each single rare variant is expected to be small. A way to solve

this is to cluster the rare variants per gene and calculated the explained variance per gene instead of per variant. Another question for association studies examining rare variants is which significance threshold to use. For common variants the estimated total number of independent common variants (based on HapMap data) was used to calculate the significance threshold. But the question is what the estimated number of independent rare variants will be. This question can partly be answered after the release of the 1,000 genomes project. This project involves sequencing the genomes of approximately 1,200 people from around the world[23], so a more clear picture of the rare variants could arise from this effort. The 1,000 genomes project withholds a number of different ethnicities, and therefore the number of subjects for each ethnic group is limited. Therefore, the number of rare variants that is discovered is probably low, and limited by the number of samples. Furthermore, only a subset of samples is sequenced in depth (with 20x coverage) and the remaining is only sequenced with on average 4x coverage, whereas 20x coverage is regarded as the minimum/golden standard. We have to wait for the data-release to revise this issue and draw firm conclusions.

A second class of genetic variants contributing to the heritability are the structural variants, the copy-number variations (CNVs). These duplications or deletions of chromosomal regions can occur in a rare or common fashion like SNPs[24]. Several rare and common CNVs (also known as CNPs – copy number polymorphisms) have been associated with diseases, like schizophrenia and Crohn's disease[25,26]. Furthermore, structural variants were associated with gene expression within the HapMap population[27], suggesting a mechanism for CNV-function in traits or diseases. At this moment SNP arrays are used to detect structural variants. It is known that a part of the CNVs present in the genome is missed using this technique[28]. Several algorithms to identify CNVs in the SNP-array data are developed until now, and these algorithms are being improved constantly to be able to get as much information from the SNP-arrays as possible.

Structural variants are more common than previously expected. Itsara et al.[29] showed that 65%-80% of individuals harbor a structural variant of at least 100kb in size, 5%-10% of the individuals studied carry a CNV of at least 500kb in length, while at least 1% of all individuals harbor a CNV of >1Mb. Comparisons of sequence maps of CNVs with those identified by SNP-arrays showed that the SNPs present on the arrays tag about 50% of the total number of known CNVs[28]. To increase the coverage of the SNP-arrays for CNVs, SNP-arrays including probes for known CNVs are developed. This will result in the ability to screen for structural variants in large sample sizes within the

next few years. On these new arrays only probes for known CNVs are present, while several studies showed that there are probably more structural variants than that are identified until now[30,31]. The 1,000 genomes project will probably yield many more CNVs, and the SNP-arrays with CNV probes could be updated again. Another approach to detect structural variants is by sequencing of the genome. The sequenced genomes have to be mapped back onto a reference genome to detect the structural variants. The 1,000 genomes project could be used to create such a reference genome. At the moment it is hard to estimate the explained variance for the structural variants, but for CNVs it will probably be similar to that of rare SNP variants, and for CNPs similar to the explained variance of common SNPs.

A third class of heritable genetic variants are epigenetic changes both of the DNA, and the histones. Epigenetic changes are differences in the DNA that do not involve changes in the underlying DNA sequence. One of the examples of epigenetics is DNA-methylation. Most human DNA-methylation is present at cytosines in so-called CpG islands. CpG islands are regions of at least 200bp and with a GC percentage that is greater than 50% and with an observed/expected CpG ratio that is greater than 60%[32]. 1% of all dinucleotides are CpGs[33], and 70%-80% of those are methylated[34]. Most DNA-methylation is present in promoter regions of genes, where it results in the inhibition of gene expression. It is known that DNA-methylation is heritable[35,36], and that methylation changes are associated with disease[36]. To measure epigenetic changes at the DNA level several arrays are available measuring methylation of ~28,000 CpGs in approximately 14,000 genes, which is genome-wide, but not comprehensive. The best option to identify the genome-wide methylation-status of all CpGs would be sequencing of the complete genome. Meissner et al.[37] described an approach to identify genome-wide DNA methylation, involving restriction of genomic DNA, and sequencing of the restriction fragments.

At this moment it is not yet possible to create a genome-wide dataset for rare variants, structural variants or DNA-methylation, because it is not possible to reliably sequence the whole genome in 10,000 individuals within a reasonable time and cost frame. To be able to do this in a high throughput manner, we have to wait for the so-called next-next generation sequencing based on nano-technology, which will probably not be available in the coming two years.

Even when the technology evolves and all of the above is possible, still a part of the heritability will be unexplained, which could be due to gene-gene and gene-environment interactions. At this point it is computational impossible to detect these interactions at a genome-wide scale. Furthermore, the statistical power that is

needed to calculated such interactions will be extremely low. For gene-environment interactions an additional problem is the quantification and standardization of many environmental factors. Due to limited power interactions are probably difficult to study genome-wide. Therefore, a hypothesis-based approach is the way forward and there is a clear need for functional knowledge of the newly discovered loci from GWAS.

To be able to identify most of the heritability of a trait or disease (preferably) genome-wide datasets of common, rare and structural variants should be combined together with DNA-methylation data and gene-gene and gene-environment interactions. At this moment this is not possible in a hypothesis-free, genome-wide manner, while candidate gene studies based on an a priori hypothesis are feasible. So candidate gene studies might form the next step in the identification of the "missing heritability". Only when we have the technological and computational resources and the statistical power we can do this in a genome-wide hypothesis-free manner.

The estrogen receptor alpha revised

In the estradiol-signalling pathway the estrogen receptor alpha (*ESR1*) plays a central role and is therefore the most studied gene. This has resulted in many genetic association studies examining this candidate gene in relation to several phenotypes in which estradiol is playing a role. In the Rotterdam Study we found associations of the *ESR1* SNPs *Pvu*II and *Xba*I with estradiol levels[38], BMD[39,40], age at natural menopause[41], height[42], osteoarthritis (OA)[43], myocardial infarction (MI)[44] and age-related macula degeneration (AMD)[45]. In addition, mouse *Esr1* knockout studies showed effects on several organ systems, emphasizing the pleiotropy of the gene[46,47]. For these previously studied estradiol-related phenotypes GWAS have been performed by now, and we can ask: how pleiotropic is *ESR1* in the GWAS era?

GWAS for E2-levels (n=~1,300)[48], menopausal age (n=~3,000)[49], OA (n=~1,000 cases)[50], MI (n=~3,000 cases)[51], and AMD (n=96 cases)[52] showed no genome-wide significant (gws) ($P \leq 5x10^{-8}$) associations with SNPs in the *ESR1* gene region until now. The reason for this could be that in these studies the sample size is low and that the estrogen receptor only has a small effect on these phenotypes, so more statistical power (i.e., larger sample sizes) is needed to detect these SNPs at genome-wide significance level. It could also be that previous candidate gene studies for *ESR1* polymorphisms showed false positive results. If there is an effect, it is relatively small and using the current sample size of these studies there is no indication of an association for *ESR1*

SNPs with these phenotypes.

For BMD[53,54] and height[17] several SNPs in the *ESR1* gene region were gws associated. Interestingly, both the BMD and the height GWAS identified several independent gws SNPs (Figure 1). The two height top hits indicated in light grey in the figure are both from the same study on 135,000 samples ($P=6x10^{-13}$ and $6x10^{-12}$)[17]. The four BMD top hits indicated in the figure originate from two studies, one by Styrkarsdottir et al.[54], and one from Rivadeneira et al.[53], on data of the GEFOS consortium, which included also the samples used in the first BMD GWAS paper. The paper by Styrkarsdottir et al.[54] was based on in total 13,700 samples and the two SNPs had a $P<5x10^{-11}$ for lumbar spine BMD. Both SNPs were located 5' of the *ESR1* gene, in *C6orf97*, coding for a hypothetical protein. The two top hits in the *ESR1* gene region from the GEFOS paper are located in the promoter region of the *ESR1* gene (Figure 1). A GWAS for breast cancer in 10,000 Chinese cases and controls showed that a SNP just 3' of *C6orf97* was associated with breast cancer ($P=2x10^{-15}$)[55].

In the lower part of Figure 1 the correlation between the SNPs is shown. One of the height top hits is moderately correlated to *Pvu*II and *Xba*I ($r^2>0.42$), suggesting that the previously described association of *Pvu*II and *Xba*I originate from the same signal as is this GWAS top hit. Furthermore, it is clear that the breast cancer top hit in Chinese is correlated to one of the BMD top hits of Styrkarsdottir et al., and could therefore, also point to the same signal. If the mechanism underlying the association is via estradiol, you expect a higher risk for breast cancer to coincide with a higher BMD for the minor allele. The minor allele of SNP C is present in 38% of the Chinese population, and is associated with a higher risk for breast cancer. SNP B had a minor allele frequency of 47% in a European population and was associated with a lower BMD. The effects of the two SNPs are not in the expected direction, this could be due to the difference in population. Chinese and European populations differ in their genetic background, which could affect the association of these SNPs with the phenotypes that were studied. Because none of the top hits for height, BMD, and breast cancer is identical, it is not possible to exactly compare the results for the three phenotypes. To be able to answer the question how pleiotropic *ESR1* is in the GWAS era, results for all phenotypes should be available for the same SNP. One of the top hits from Rivadeneira et al. is located close to promoter exon F, which is known to be the only active promoter in bone[56,57], but the correlation with *Pvu*II and *Xba*I is very low for these SNPs. This suggests that the *Pvu*II and *Xba*I SNPs are probably not the functional SNPs for BMD, but that the functional or causal variant could be located in the promoter region close to exon F.

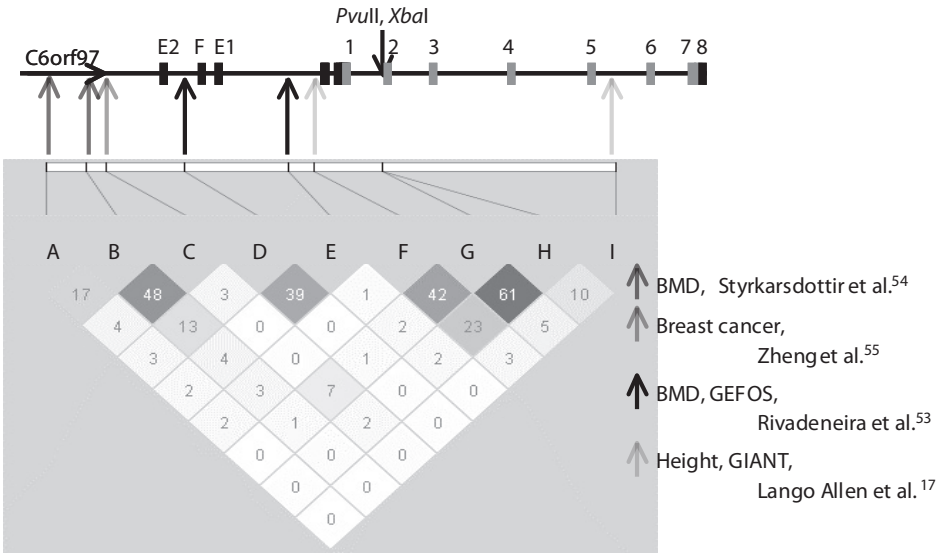


Figure 1. Association signals of *ESR1* SNPs in candidate gene studies and GWAS. The black arrow indicates the *Pvu*II and *Xba*II polymorphisms from the candidate gene studies (SNP G and H). The arrows indicate the top hits from the GWASs, with in dark grey the study of Styrkarsdottir et al.[54] (SNP A and B), and in black the hits for the GWAS by Rivadeneira et al.[53] (SNP D and E). In lighter grey the breast cancer top hit[55] (SNP C) is shown, and in the lightest grey the two top hits for the height GWAS by Langoo Allen et al.[17] (SNP F and I) are shown. The graph below the gene structure indicates the correlation of the different top hits.

The candidate gene studies for *ESR1* until now only focussed on the gene itself, but the receptor has to bind to specific sites in the DNA to be able to affect the expression of target genes, so-called EREs (estrogen response elements). Thus, the effects of *ESR1* could also be modified by SNPs in these EREs. Therefore it would be worthwhile to also study the genetic variation in these EREs together with the *ESR1* gene in a candidate gene study for all estradiol-related phenotypes.

Future directions

In this thesis we showed that the candidate gene approach could yield reliable results, when the right candidate gene is chosen. Furthermore, we showed that GWAS discovers new genetic variation influencing age at menarche and age at natural menopause. However, most GWAS results do not identify genes, but genetic loci, which makes the interpretation of these results more challenging compared

to candidate gene studies. So, the results from the GWAS are the beginning of the exploration of new biological pathways involved in estrogen-related traits. For this exploration candidate gene studies are the next step forward, after the identification of the gene underlying the mechanism of association.

To be able to identify the genes and true causal variants of GWAS more in depth analysis of the top hit regions is needed. One step could be the identification of the true functional common variant by deep-sequencing of the LD-block of the top hits. With the common variants identified by this, association studies could be performed. This could lead to the identification of a functional variant associated with the trait or disease. Subsequently, this variant should be tested in *in vitro* experiments to identify the effects of the variant on the function of the gene, and eventually the biological mechanism underlying the association.

Another feature that can be studied relatively easily is the possible effect of the SNPs on the expression of genes. For this purpose several databases exist in which expression levels and SNPs are measured and correlation between the two is studied. These datasets are now available for a couple of tissues and cell lines, namely (immortalized) lymphoblastoid cells[58,59], liver[60] and brain[61]. For menopause associated SNPs these tissues are not likely to be the target tissues and possibly, mRNAs influenced by our top hits are not expressed in these tissues. This holds for most of the traits and diseases that are studied. To be able to detect the effect of the top SNPs on mRNA expression it is necessary to study the expression of the genes in for example ovaries or specific brain areas. Recently, the National Institutes of Health (USA) sponsored a pilot to create the so-called GTEx (Genotype-Tissue Expression) database, which will contain the correlation of genetic variation (SNPs) and gene expression in multiple human tissues[62]. This will eventually provide a tissue-specific list of variants associated with specific transcripts, and can be used to detect these gene expression effects of GWAS top hits in a tissue-specific manner. With this database it will be possible to examine *in silico* the correlation of your GWAS top hits with candidate genes in the tissues of interest for your phenotype.

Next to searching databases for possible effects on gene expression, another step in identifying the mechanisms underlying the GWAS top hits is in depth functional analysis of the candidate genes. The first step could involve the expression of the candidate genes in a GWAS top hit in the tissue of interest. This will give an indication of which of the genes is a good candidate for further study. Subsequently, studies could be performed in which the gene is knocked out in cells of the target tissue, to examine the effect on the cellular level. Furthermore, knock out animal models for

the gene could be studied to examine the overall effects of the absence of the gene. These knockout experiments could also confirm the role of the candidate gene in the mechanism underlying the phenotype.

Concluding Remark

The underlying genetic background for traits like sex homone levels, age at menarche or age at natural menopause and for a diasease as osteoporosis are far from being all known, and still require extensive research. Nevertheless, the findings presented in this thesis provide more insight into this genetic background. Hopefully these results will contribute to a better understanding of the inter-individual differences in the population for these traits and disease, and may be translated in future interventions for infertility treatment and osteoporosis treatment or prevention.

REFERENCES

1. Reich, D.E. & Lander, E.S. On the allelic spectrum of human disease. *Trends Genet* **17**, 502-10 (2001).
2. 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).
3. Hirschhorn, J.N. Genomewide association studies--illuminating biologic pathways. *N Engl J Med* **360**, 1699-701 (2009).
4. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21 (2008).
5. McCarthy, M.I. Exploring the unknown: assumptions about allelic architecture and strategies for susceptibility variant discovery. *Genome Med* **1**, 66 (2009).
6. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69**, 124-37 (2001).
7. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**, 212-9 (2009).
8. Chen, J. et al. Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am J Hum Genet* **75**, 807-21 (2004).
9. Grundberg, E. et al. A deletion polymorphism in the RIZ gene, a female sex steroid hormone receptor coactivator, exhibits decreased response to estrogen in vitro and associates with low bone mineral density in young Swedish women. *J Clin Endocrinol Metab* **89**, 6173-8 (2004).
10. Scanlon, P.D., Raymond, F.A. & Weinshilboum, R.M. Catechol-O-methyltransferase: thermolabile enzyme in erythrocytes of subjects homozygous for allele for low activity. *Science* **203**, 63-5 (1979).
11. Grundberg, E. et al. The impact of estradiol on bone mineral density is modulated by the specific estrogen receptor-alpha cofactor retinoblastoma-interacting zinc finger protein-1 insertion/deletion polymorphism. *J Clin Endocrinol Metab* **92**, 2300-6 (2007).
12. Lorentzon, M., Eriksson, A.L., Mellstrom, D. & Ohlsson, C. The COMT val158met Polymorphism Is Associated With Peak BMD in Men. *J Bone Miner Res* **19**, 2005-11 (2004).
13. Rice, J.P., Saccone, N.L. & Rasmussen, E. Definition of the phenotype. *Adv Genet* **42**, 69-76 (2001).
14. Onland-Moret, N.C. et al. Age at menarche in relation to adult height: the EPIC study. *Am J Epidemiol* **162**, 623-32 (2005).
15. Khoury, M.J. et al. Genome-Wide Association Studies, Field Synopses, and the Development of the Knowledge Base on Genetic Variation and Human Diseases. *Am J Epidemiol* (2009).
16. http://www.genome.org/gwastudies.
17. Lango Allen, H. et al. The identification of over 210 loci involved in adult height variation provides important insights into the contribution of common variation to a model complex trait. in *ASHG* (2009).
18. Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R. & Amos, C.I. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* **82**, 100-12 (2008).
19. Ji, W. et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* **40**, 592-9 (2008).

20. Cohen, J.C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72 (2004).
21. Romeo, S. et al. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* **39**, 513-6 (2007).
22. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J.A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387-9 (2009).
23. Spencer, G. International Consortium Announces the 1000 Genomes Project. (2008).
24. McCarroll, S.A. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* **17**, R135-42 (2008).
25. International Schizophrenia, C. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-41 (2008).
26. McCarroll, S.A. et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* **40**, 1107-12 (2008).
27. Stranger, B.E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-53 (2007).
28. Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E. & Nickerson, D.A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* **40**, 1199-203 (2008).
29. Itsara, A. et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**, 148-61 (2009).
30. Kidd, J.M. et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56-64 (2008).
31. Korbel, J.O. et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420-6 (2007).
32. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-82 (1987).
33. Swartz, M.N., Trautner, T.A. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem* **237**, 1961-7 (1962).
34. Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333**, 143-9 (2004).
35. Roemer, I., Reik, W., Dean, W. & Klose, J. Epigenetic inheritance in the mouse. *Curr Biol* **7**, 277-80 (1997).
36. Bjornsson, H.T., Fallin, M.D. & Feinberg, A.P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet* **20**, 350-8 (2004).
37. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-70 (2008).
38. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. *Eur J Endocrinol* **153**, 327-34 (2005).
39. Ioannidis, J.P. et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *Jama* **292**, 2105-14 (2004).
40. van Meurs, J.B. et al. Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk. *Hum Mol Genet* **12**, 1745-54 (2003).
41. Weel, A.E. et al. Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. *J Clin Endocrinol Metab* **84**, 3146-50 (1999).
42. Schuit, S.C. et al. Height in pre- and postmenopausal women is influenced by estrogen receptor alpha gene polymorphisms. *J Clin Endocrinol Metab* **89**, 303-9 (2004).
43. Bergink, A.P. et al. Estrogen receptor alpha gene haplotype is associated with

radiographic osteoarthritis of the knee in elderly men and women. *Arthritis Rheum* **48**, 1913-22 (2003).
44. Schuit, S.C. et al. Estrogen receptor alpha gene polymorphisms and risk of myocardial infarction. *Jama* **291**, 2969-77 (2004).
45. Boekhoorn, S.S. et al. Estrogen receptor alpha gene polymorphisms associated with incident aging macula disorder. *Invest Ophthalmol Vis Sci* **48**, 1012-7 (2007).
46. Couse, J.F. & Korach, K.S. Estrogen receptor null mice: what have we learned and where will they lead us? *Endocr Rev* **20**, 358-417 (1999).
47. Mueller, S.O. & Korach, K.S. Estrogen receptors and endocrine diseases: lessons from estrogen receptor knockout mice. *Curr Opin Pharmacol* **1**, 613-9 (2001).
48. Stolk, L. et al. A Genome-wide Association Study of Sex Hormone Levels; the Rotterdam Study. in *Genetics of Estrogen-Related Traits: From Candidate Genes to GWAS* (2009).
49. Stolk, L. et al. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat Genet* (2009).
50. Kerkhof, J.M. et al. A Genome-Wide Association Study identifies a locus on chromosome 7q22 to influence susceptibility for osteoarthritis. *Arthritis & Rheumatism* **accepted for publication**(2009).
51. Myocardial Infarction Genetics, C. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
52. Klein, R.J. et al. Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-9 (2005).
53. Rivadeneira, F. & GEFOS. Twenty bone mineral density loci identified by large-scale meta-analysis of genome-wide association studies. *Nature Genetics* **in press**(2009).
54. Styrkarsdottir, U. et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* **358**, 2355-65 (2008).
55. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-8 (2009).
56. Denger, S. et al. ERalpha gene expression in human primary osteoblasts: evidence for the expression of two receptor proteins. *Mol Endocrinol* **15**, 2064-77 (2001).
57. Lambertini, E., Penolazzi, L., Giordano, S., Del Senno, L. & Piva, R. Expression of the human oestrogen receptor-alpha gene is regulated by promoter F in MG-63 osteoblastic cells. *Biochem J* **372**, 831-9 (2003).
58. Stranger, B.E. et al. Population genomics of human gene expression. *Nat Genet* **39**, 1217-24 (2007).
59. Dixon, A.L. et al. A genome-wide association study of global gene expression. *Nat Genet* **39**, 1202-7 (2007).
60. Schadt, E.E. et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6**, e107 (2008).
61. Myers, A.J. et al. A survey of genetic human cortical gene expression. *Nat Genet* **39**, 1494-9 (2007).
62. Hardy, J. & Singleton, A. Genomewide association studies and human disease. *N Engl J Med* **360**, 1759-68 (2009).

Chapter 10

Summary

Samenvatting

SUMMARY

Estradiol is one of the reproductive hormones, and has effects not only on reproductive tissue affecting female development and fertility, but also on non-reproductive tissues in both men and women. The inter-individual differences in estradiol levels are to a large extend affected by both genetic and environmental factors. During a woman's lifetime the increase in systemically available estradiol is marked by menarche while later in life the decline in these levels is marked by menopause. Both these traits could therefore influence the level of estradiol available after menopause, while both also have a genetic background. Osteoporosis is a disease characterized by low bone mineral density (BMD) and increased fracture risk. It is known that both environmental and genetic factors play a role in the etiology of this disease. One of the risk factors that increase the risk for osteoporosis are low estradiol levels, but also early menopause is one of those factors. The aims of this thesis were therefore 1) to identify genetic variations in genes of the estrogen pathway that are associated with height, E2-levels, bone mineral density, and risk for vertebral fractures, and 2) to identify new genes and pathways involved in inter-individual variations in hormone levels, menarche and menopause.

In **part A**, the association of polymorphisms in three candidate genes with estradiol levels, age at natural menopause, BMD and fracture risk is presented. In **chapter 2** fine-mapping of an area in the Estrogen receptor alpha (*ESR1*) gene (the *Pvu*II and *Xba*I LD-block) is presented. In women of a subset of the Rotterdam Study I and in the complete Rotterdam Study I the associations for several SNPs in one LD-block with BMD, fracture risk, E2-levels, and menopausal age were studied. We also performed haplotype analysis based on 14 SNPs showing that one additional SNP in the promoter region of the gene could distinguish a third haplotype in the previously defined *Pvu*II–*Xba*I haplotype 1 and explain the associations we found before.

Chapter 3 shows that, contrary to previous observations in other studies the Pro704 insertion-deletion polymorphism in the ERα co-activator *RIZ1* (Retinoblastoma interacting zinc finger domain) was not associated with BMD or fracture risk in men and women of the Rotterdam Study I. The association of the Met158Val polymorphism in the estradiol degrading enzyme COMT (Catechol-O-Methyltransferase) with fracture risk in men and women, independent of BMD is described in **chapter 4**. Our study is the first to describe the association with fracture risk, although, a previous study in a Swedish cohort showed an association with BMD of the same allele.

In **part B** of this thesis a candidate pathway analysis approach is presented.

Using imputed genotype data for over 14,000 polymorphisms in 61 genes of the estrogen pathway, the association with sex hormone levels, age at natural menopause and BMD and fracture risk is analysed in the Rotterdam Study (**chapter 5**). Several SNPs in the 5' region of the *SHBG* gene are associated with a 2.5% difference in SHBG levels between the homozygous (extreme) genotype groups in both men and women. For plasma estradiol or testosterone significant associations in both men and women were found, and also not for age at natural menopause. One SNP in the *HSD17B5* gene was associated with lumbar spine (ls) BMD in postmenopausal women, while no associations for lsBMD in elderly men were observed. HSD17B5 is an enzyme involved in the conversion of androstendione to testosterone. For non of the >14,000 studied SNPs an association with vertebral fracture risk was found.

Next we examined in this dataset SNPs that showed consistent associations with plasma hormone levels or age at natural menopause and with lsBMD. In addition we sought to identify SNPs consistently associated with lsBMD and vertebral fx risk. In women we found an association for a SNP in NSD1 (an androgen receptor co-activator) with testosterone-levels and lsBMD, and in men we found 4 SNPs in the *ESRRG* (a co-activator of ESR1) gene associated with testosterone-levels and lsBMD. The consistency analysis of E2-levels and lsBMD in men yielded a significant SNP in the *ESR1* gene. 7 SNPs located in 3 genes associated in women (*POR, HSD17B5, ESRRA*), and 34 SNPs situated in four different genes in men (*LHCGR, ESR1, HSD17B3, NCOR1*) were consistently associated with lsBMD and vertebral fracture risk. In conclusion, this study showed marked differences between men and women in genes that have a consistent association with hormone levels and lsBMD, or lsBMD and fracture risk. However, the power of our study was low, so all results should be replicated in other studies to be able to draw firm conclusions.

The previous studies are all based on known biology in choosing relevant candidate genes. However, maybe not all genes of the estrogen pathway are already annotated, therefore we also conducted hypothesis-free genome-wide association studies (GWAS) (**Part C**). In these GWAS >300,000 of the common genetic variants in the human genome are tested for an association with the phenotype, without *a priori* knowledge about the effect of the variant or gene on these estrogen-related traits. A GWAS on sex hormone levels and sex hormone binding globuline levels in a subset of men and women from the Rotterdam Study I is presented in **chapter 6**. Highly significant associations of several SNPs in the SHBG gene region with plasma SHBG levels in men and women were found. These signals were the same as those that were found previously in the estrogen pathway analysis. Furthermore, a more

modest association for two SNPs with plasma DHEAS levels in women was observed. These SNPs were located in a hypothetical protein, which makes it hard to explain the biological relevance of this association. In a combined analysis of men and women significant associations for three SNPs with testosterone-levels were found. Two of these SNPs are located in an intergenic region on chromosome 6 and one in the *HDAC9* gene, a histone deacetylase. For all three SNPs no functional mechanism underlying the association is known. In this study the power was also low, so again further study is needed to validate our results.

A study on age at menarche in 17,510 women of seven studies from the CHARGE consortium is discussed in **chapter 7**. This study identified two loci that survived the multiple testing adjustment for 1,000,000 common independent variants of $P=5x10^{-8}$. The first hit is located on chromosome 9 in an intergenic region, with the closest gene located at approximately 400kb distance. This makes it more challenging to define the gene underlying this association. The other signal was found in the *LIN28B* gene region on chromosome 6. The top SNP of this locus was also associated with adult height, which is in line with the observation that later menarche results in longer adult height. **Chapter 8** shows the results of a genome-wide association study for age at natural menopause in 2,979 European women. This study identified six SNPs in three loci located on chromosome 13, 19 and 20 associated with age at natural menopause. The largest effect was seen for the SNPs on chromosome 13, which showed a difference in menopausal age of 1 year between the two homozygous genotype groups. The three loci have not been indicated to play a role in natural menopause before, and therefore the functional mechanism underlying these associations remains unknown. The total explained variance in age at natural menopause for these three loci was 1.1%, leaving much of the genetic variation unexplained. We expect to explain more of this genetic variance when the sample size is increased, however, we do not expect that all genetic variance will be explained by common variants. The additional genetic variance can be explained by several classes of genetic variation. One is the rare variants, but also copy number variants could play a role. Copy number variants are deletions or duplications of large stretches of a chromosome. Another part of the genetic variance could be explained by epigenetic changes (differences in DNA-methylation patterns). Next to these variants, the interaction between variants or variants and the environment will explain some of the missing heritability.

In **chapter 9** the main findings of this thesis are brought together and placed in a broader context. The findings of candidate gene studies and GWAS are compared, and both analysis approaches are discussed. Furthermore, recent findings for genetic

variants in the *ESR1* gene are discussed, and at the end several recommendations for future research on the GWAS top hits are presented.

SAMENVATTING

Oestradiol, ook wel het vrouwelijk geslachtshormon genoemd, heeft naast effecten op de ontwikkeling van voortplantingsorganen en vruchtbaarheid in vrouwen, ook een effect op andere weefsels en organen in zowel mannen als vrouwen. Tijdens het leven van een vrouw is er een sterke stijging van de oestradiol spiegels die wordt gemarkeerd door de menarche (de eerste menstruatie) en later een sterke daling in deze spiegels, die gemarkeerd wordt door de menopauze (het einde van de mentruatie). De verschillen tussen individuen in oestradiol spiegels worden voor een groot deel veroorzaakt door zowel genetische als omgevingfactoren. Daarnaast hebben ook menarche en menopauze een genetische achtergrond en kunnen beide de beschikbare hoeveelheid oestradiol na de menopauze beïnvloeden. Na de menopauze hebben vrouwen een verhoogd risico op het krijgen van osteoporose, dit is een ziekte die gekenmerkt wordt door een lage bot mineraal dichtheid (BMD) en een verhoogd risico op botbreuken (fracturen). Het is bekend dat zowel omgevingsfactoren als genetische factoren een rol spelen bij het ontstaan van deze ziekte. Een van de risico factoren voor osteoporose is lage oestradiol spiegels, net als een vroege menopauze. De doelstellingen van dit proefschrift zijn dan ook: 1) het identificeren van genetische variaties (polymorfismen) in genen betrokken bij oestrogeen metabolisme die geassocieerd zijn met lengte, oestradiol-spiegels, BMD en het risico op wervelkolom fracturen, en 2) het identificeren van nieuwe genen en routes die betrokken zijn bij de verschillen tussen individuen op het gebied van hormoon spiegels, menarche leeftijd en menopauze leeftijd.

Deel A laat de associatie van polymorfismen in drie kandidaat genen met oestradiol spiegels, leeftijd van menopauze, BMD en fractuur (fx) risico zien. In **hoofdstuk 2** wordt er gekeken naar een bekend gebied in het oestrogeen receptor alfa (*ESR1*) gen. Dit gebied, ook wel het *Pvu*II-*Xba*I LD-blok genoemd, is geassocieerd met BMD, fractuur risico, oestradiol-spiegels en leeftijd van menopauze. In dit hoofdstuk zijn er 12 extra Single Nucleotide Polymorfismen (SNPs) bestudeerd voor de associatie met deze eindpunten. Dit is gedaan in zowel een subset van vrouwen uit de Rotterdam Studie I als de complete vrouwelijke populatie van Rotterdam Studie I. Met deze analyse en een zogenoemde haplotype analyse (waarbij de combinaties van de allelel van de verschillende SNPs op een chromosoom bestudeerd worden) is er geprobeerd om een variant te vinden die de *Pvu*II-*Xba*I associaties beter kan verklaren (een SNP die een bekende functie heeft) of waarmee we kunnen inzoomen op een functionele variant die we nog niet bestudeerd hebben. Uiteindelijk kwam uit deze

analyse 1 SNP in de promoter regio van het *ESR1* gen die het eerder bestudeerde *Pvu*II-*Xba*I haplotype 1 kan splitsen en dat dit nieuwe haplotype de gevonden associaties kan verklaren.

Hoofdstuk 3 laat zien dat het Pro704 insertie-deletie polymorfisme in de ESR1 co-activator RIZ1 (Retinoblastoma interacting zincfinger domein) niet geassocieerd was met BMD en fractuur risico in mannen en vrouwen van de Rotterdam Studie I, dit in tegenstelling tot eerdere observaties in andere studies. In **hoofdstuk 4** wordt de associatie van het Met158Val polymorfisme in het oestradiol afbrekende enzym COMT (Catechol-O-Methyltransferase) met fractuur risico in mannen en vrouwen, onafhankelijk van BMD beschreven. Onze studie laat als eerste deze associatie met fractuur risico zien, wel heeft een eerdere studie uit Zweden een associatie met BMD laten zien.

In **deel B** van dit proefschrift wordt een kandidaat-route aanpak gepresenteerd. Door gebruik te maken van genotype data van meer dan 14.000 polymorfismen in 61 genen van de oestrogeen metabolisme en signaal route is er gekeken naar de associatie van deze SNPs met spiegels van geslachtshormonen, menopauze leeftijd, en BMD en fractuur risico in de Rotterdam Studie (**hoofdstuk 5**). Verscheidene SNPs in de regio voor het *SHBG* gen zijn geassocieerd met een 2,5% verschil in SHBG spiegels tussen de twee homozygote (extreme) genotype groepen in zowel mannen als vrouwen. Er zijn geen significante associaties voor plasma oestradiol of testosteron spiegels gevonden in mannen en vrouwen en ook niet voor menopauze leeftijd. Eén SNP in het *HSD17B5* gen was geassocieerd met BMD van de lumbale wervels (lwBMD) in postmenopauzale vrouwen, terwijl er geen associaties voor lwBMD in oudere mannen werden gevonden. HSD17B5 is een enzym dat betrokken is bij het omzetten van androstenedion naar tesosteron. Voor geen van de ruim 14.000 bestudeerde SNPs was er een associatie met wervel kolom fx risico.

Daarnaast hebben we in deze dataset SNPs gezocht die een consistente associatie vertoonden met hormoon spiegels of menopauze leeftijd en lwBMD of wervelkolom fx risico. Een consistent associatie wil zeggen dat de SNP geassocieerd is met lagere hormoon spiegels of een eerdere menopauze en een lagere BMD, of hogeren hormoon spiegels en een hogere lwBMD. Verder hebben we geprobeerd om SNPs te vinden die consistent geassocieerd waren met lwBMD en wervelkolom fx risico (i.e., een lagere lwBMD en een verhoogd risico op botbreuken, of een hogere lwBMD en een verlaagd risico op botbreuken). In vrouwen was er een associatie te zien van een SNP in het *NSD1* gen (een androgeen-receptor co-activator) met testosteron spiegels en lwBMD, in mannen waren 4 SNPs in het *ESRRG* gen (een co-activator van

ESR1) geassocieerd met testosteron spiegels en lwBMD. Verder werd er een associatie gevonden van een SNP in het *ESR1* gen zelf die geassocieerd was met oestradiol spiegels en lwBMD in mannen. De consistentie analyse voor lwBMD en fx risico toonde zeven SNPs in drie genen aan (*POR*, *HSD17B5*, *ESRRA*) die consistent geassocieerd waren in vrouwen en 34 SNPs in vier andere genen (*LHCGR*, *ESR1*, *HSD17B3*, *NCOR1*) consistent geassocieerd in mannen. Deze studie liet grote verschillen zien tussen mannen en vrouwen in de genen die consistent geassocieerd zijn met hormoon spiegels en lwBMD en met lwBMD en fx risico. Maar doordat er weinig mensen in deze studie zijn bestudeerd is er een groot aantal fout-positieven (de power is laag), daarom moeten alle resultaten gerepliceerd worden in andere studies om een goede conclusie te kunnen trekken.

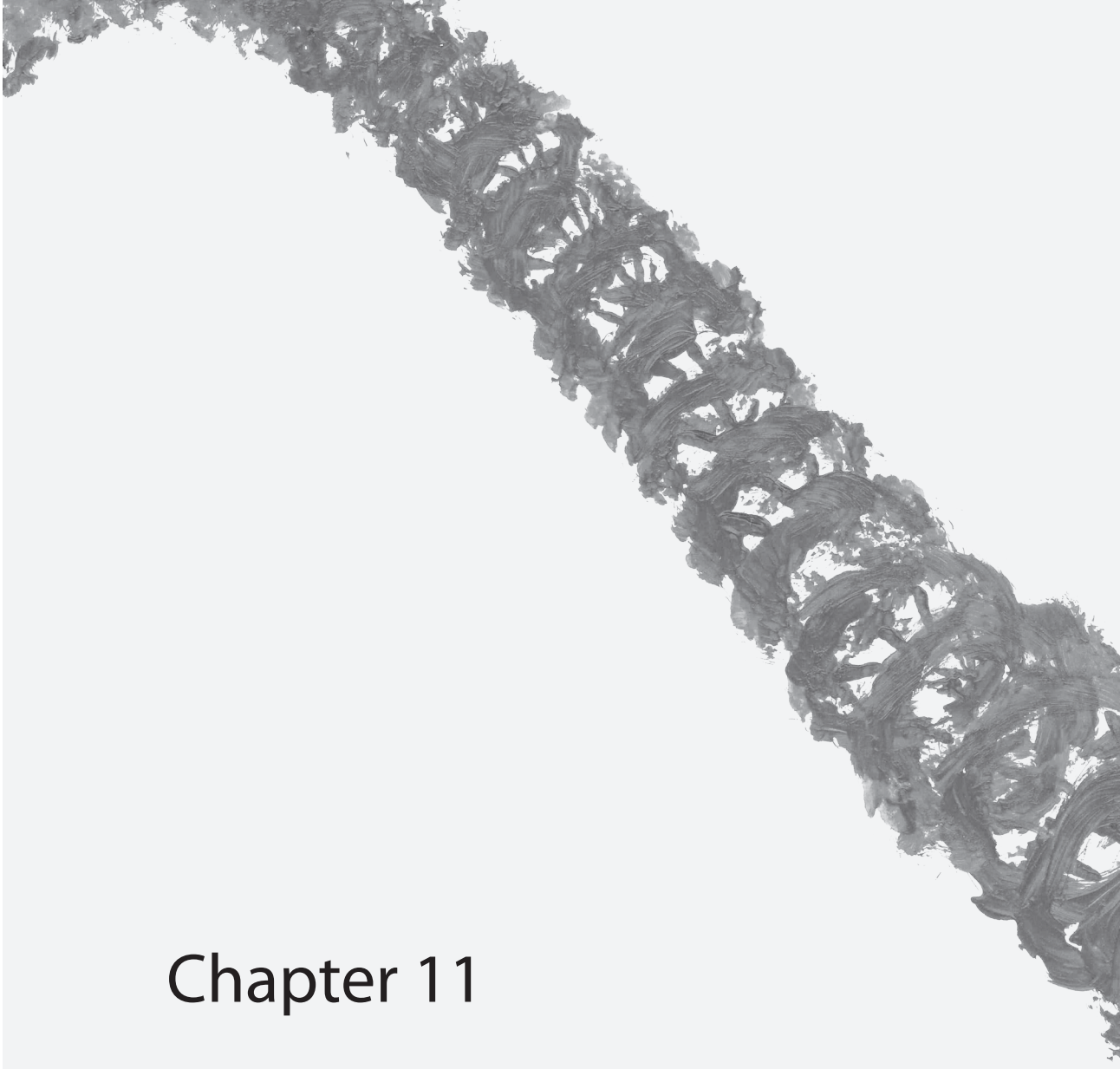
De hiervoor beschreven studies zijn allemaal gebaseerd op bekende biologie bij de keuze van de kandidaat genen, maar misschien zijn niet alle genen betrokken bij de aanmaak, afbraak en effecten van oestrogenen op dit moment bekend. Daarom hebben we ook hypothese-vrije genoom-wijde associatie studies (GWAS) gedaan (**deel C**). In een GWAS wordt de associatie van > 300.000 veel voorkomende genetische variaties in het humane DNA met een oestrogeengerelateerde eigenschap getest, zonder dat er een *a priori* hypothese is over het effect van de genetische variant of het gen op oestrogeengerelateerde eigenschappen. Een GWAS naar geslachtshormoon spiegels in mannen en vrouwen van de Rotterdam Studie I is beschreven in **hoofdstuk 6**. Er werden zeer significante associaties gevonden voor verschillende SNPs in de *SHBG* gen regio met SHBG spiegels in mannen en vrouwen. Deze SNPs zijn dezelfde als eerder al gevonden waren in de kandidaat route studie. Verder werd er een minder significante associatie gevonden voor twee SNPs met plasma DHEAS spiegels in vrouwen, deze SNPs liggen in een gen dat codeert voor een hypothetisch eiwit, waardoor het moeilijk is de biologische achtergrond van deze associatie te verklaren. In de analyse waarbij de resultaten voor mannen en vrouwen gecombineerd werden, was er een associatie te zien van drie SNPs met testosteron spiegels. Twee van deze SNPs liggen in een gebied tussen genen in op chromosoom 6, en de andere SNP ligt in het *HDAC9* gen, dit is een histon deacetylase. Voor deze drie SNPs is het functionele mechanisme achter de associatie niet bekend. Verder is de power in deze studie laag, waardoor vervolg studies nodig zijn om onze resultaten te bevestigen.

Een studie naar menarche leeftijd in 17.510 vrouwen uit zeven verschillende studies van het CHARGE consortium wordt beschreven in **hoofdstuk 7**. Deze studie laat twee gebieden in het DNA zien die een p-waarde hebben die kleiner is dan de $5x10^{-8}$ p-waarde drempel die gebruikt wordt in GWAS. Het eerste gebied (locus) ligt

op chromosoom 9 in een stuk van het DNA tussen genen in, met het dichtstbijzijnde gen op ongeveer 400kb afstand. Dit zorgt ervoor dat het een grotere uitdaging is om voor dit locus het gen te vinden dat voor de associatie zorgt. Het andere locus ligt op chromosoom 6 in de *LIN28B* gen regio, deze SNP was ook geassocieerd met lengte op volwassen leeftijd. Dit is in overeenstemming met de observatie dat een latere menarche een grotere volwassen lengte tot gevolg heeft. **Hoofdstuk 8** laat de resultaten van een GWAS naar natuurlijke menopauze leeftijd in 2.979 Europese vrouwen zien. In deze studie zijn zes SNPs in drie gebieden (loci) gevonden die geassocieerd zijn met natuurlijke menopauze leeftijd (op chromosoom 13, 19 en 20). Het grootste effect werd gevonden voor de chromosoom 13 SNP, die een verschil van 1 jaar tussen de homozygote genotype groepen liet zien. De drie loci zijn nog niet eerder gevonden in een menopauze studie en het functionele mechanisme achter de associaties is nog onbekend op dit moment. Het totale verschil tussen mensen in menopauze leeftijd dat we kunnen verklaren met deze 3 gebieden is 1,1%. Dit houdt in dat er nog heel veel genetische variantie onverklaard blijft. Een deel hiervan kan gevonden worden als het aantal vrouwen vergroot wordt, maar niet alles kan verklaard worden door de veelvoorkomende (minst voorkomende allel >1-5%) genetische variaties die in dit proefschrift bestudeerd zijn.

Een deel van de genetische variantie kan verklaard worden door zeldzame SNPs, met allel frequenties voor het minst voorkomende allel <1%. Een andere soort genetische variatie die een rol kan spelen zijn de structurele variaties; de copy nummer variaties. Dit zijn deleties of duplicaties van grote stukken van een chromosoom en deze kunnen zowel zeldzaam als veelvoorkomend zijn. Weer een ander deel van de genetische variantie zou verklaard kunnen worden door epigenetische veranderingen (verschillen in patronen van DNA-methylering). Naast deze variaties, zal de interactie tussen genetische variaties en tussen genetische variaties en omgevingsfactoren ook een deel van de missende erfelijkheid kunnen verklaren.

In **hoofdstuk 9** worden de resultaten van dit proefschrift bij elkaar gebracht en in een breder perspectief geplaatst. De bevindingen van de kandidaat gen studies en de GWAS worden vergeleken en beide benaderingen worden bediscussieerd. Verder worden er recente bevindingen voor het *ESR1* gen besproken en aan het einde van dit hoofdstuk worden een aantal ideeën voor vervolg onderzoek van de GWAS resultaten gepresenteerd.



Chapter 11

Dankwoord

List of Publications

PhD-portfolio

DANKWOORD

En nu is dit proefschrift af en rest mij nog het dankwoord, misschien wel het lastigste onderdeel van dit boekje. Want hoe kan je iedereen bedanken die de afgelopen jaren op de een of andere manier bijgedragen heeft aan de totstandkoming van dit proefschrift en alles daaromheen. Ik zou graag iedereen persoonlijk willen bedanken, maar dan wordt dit dankwoord te groot. Omdat dit het meest gelezen onderdeel van dit proefschrift gaat zijn, en iedereen op zoek gaat naar zijn eigen naam: ik heb mijn best gedaan niemand te vergeten, mocht dat toch gebeurt zijn, is dat onbewust en zeker niet persoonlijk bedoeld en dus alvast dank je wel!

Allereerst wil ik beginnen met mijn promoter prof.dr. A.G. Uitterlinden. Beste André, al vanaf mijn stage in 2003 ken ik je als een ontzettend gedreven, erg enthousiaste begeleider. Ik wil je erg graag bedanken dat je het helemaal zag zitten om mij, lichte chaoot, te sturen tijdens mijn promotie-onderzoek richting vandaag. Jouw motto, "everything is genetic" kan ik na deze jaren ook omarmen als het mijne. Ik heb van jou niet alleen de grote lijnen van de genetische epidemiologie geleerd, maar ook jouw manier van internationale samenwerkingen, iets wat niet in veel onderzoeksgebieden zo vanzelfsprekend is. Dank je wel dat ik daarbinnen mijn eigen ding mocht doen voor dit proefschrift, en ik hoop nog langer samen met jou hieraan te blijven werken

Natuurlijk als tweede mijn copromoter dr. J.B.J. van Meurs, beste Joyce, ik heb je wel eens mijn steun en toeverlaat op het lab genoemd, en dat was niet eens helemaal een geintje. Waarschijnlijk ben jij degene die na mij het beste weet wat er in dit proefschrift staat. Onze discussies over hele simpele analyzes en resultaten (het eerste jaar) en de wat moeilijkere dingen, zoals "Hoe gaan we nu verder?" (de laatste tijd) en alles wat daar tussen in zat hebben mij een betere onderzoeker gemaakt, daar ben ik van overtuigd! (Alhoewel ik het ook wel fijn vind dat ik tegenwoordig jou met dingen kan helpen). Ik ben ook erg blij dat we samen verder kunnen met het vervolg op GWAS.

Graag wil ik ook de leden van de leescommissie bedanken, prof.dr. H.A.P. Pols, prof.dr. T.D. Spector en dr. J.S.E. Laven. Beste Huib, je begon als mijn promoter later toen je decaan werd, was je wat meer op de achtergrond betrokken bij dit onderzoek. Door jouw vragen en de discussies tijdens artikel besprekingen of de wekelijkse ERGO WB heb ik mijn werk kritischer leren beoordelen, dank daarvoor. Dear Tim, thank you for taking place in the committee and you helpful discussion during the writing of the menopause GWAS paper. Beste Joop, jouw commentaren op mijn menarche en

menopauze bevindingen hebben me erg geholpen de waarde van mijn werk in te schatten.

Of course a thank you to the rest of the committee, prof.dr. C.M. van Duijn, prof.dr. C. Ohlsson, thanks for coming to the Netherlands to be part of my committee and prof.dr. F.H. de Jong. Beste Frank, jouw onuitputtelijke kennis over oestrogenen en androgenen heeft mij heel erg geholpen en ik verheug me op onze samenwerking bij de sex hormone GWAS.

Dan iedereen van het genetisch lab: Fernando heel erg bedankt voor je discussies over mijn papers, jouw papers en je hulp bij analyses, het heeft me een betere statisticus gemaakt. Marjolein en Hanneke, bedankt dat ik jullie (m)oppper-OIO mocht zijn, en natuurlijk voor de gezelligheid op de kamer tijdens koffie en fruit tijd, van jullie heb ik geleerd ook eens nee te zeggen. Ik hoop nog lang met jullie samen te kunnen werken. Karol, muchas gracis por la paciencia que has tenido en ayudarme a mi, bióloga molecular, con algunos de los programas informáticos una y otra vez… (Dasha, gracias por la traducción). Michael, zonder jouw hulp hadden we die Nature Genetics paper nooit zo snel afgekregen en ik ben erg blij dat je vandaag mijn paranimf wil zijn, dank je wel. Saskia, sportmaatje, wereldreiziger, zusje-tomaat en paranifm 2, ik hoop nog lang ons wekelijkse sportuurtje te kunnen voortzetten en dank je wel voor de gesprekken die daar weer bij horen! Pascal, we zijn vanaf het begin kamergenoten en ik wil je af en toe nog wel lastig vallen met mijn verhalen, maar zonder jou stort het hele lab in (al wil je dat niet horen). Ik heb erg veel van je geleerd, al vanaf mijn stage. Marijn, ik zal proberen nooit meer een file kwijt te raken zodat jij weer mag zoeken, maar ik kan niks beloven!! Mila, het functionele geweten van de groep, ik heb een hoop van je geleerd, dank voor het runnen van zoveel arrays (samen met Pascal). Carola, je bent pas zelf gepromoveerd en jouw klinische kijk op dingen is erg handig bij de werkbesprekingen en discussies. Ramazan, bedankt voor de Turske zoetigheid (en het recept ervan)! Liz, your new in our group, but already very essential for all array work. Martha en Tom, dank je wel voor de gezelligheid. Rowena, je bent al eventjes weg, maar samenwerken met jou betekent lol, gelukkig kom je af en toe nog langs. Arnoud, je begon als HLO afstudeerder, toen een jaartje analist en nu zit je in Leiden. Wat een stilte, bedankt voor de gezelligheid en ik heb ook een hoop geleerd. Alle andere studenten: Ulku, Anita, Ashley en nu Debbie. Heel erg bedankt voor jullie bijdrage aan dit proefschrift, hoe klein die ook kan lijken en ik hoop dat ik jullie wat van mijn enthousiasme voor DNA-onderzoek heb kunnen meegeven.

Alle mensen op het ERGO onderzoekscentrum wil ik graag bedanken voor hun werk, omdat zonder jullie dit proefschrift nooit geschreven had kunnen worden.

Frank van Rooij en Jeannette Vergeer jullie wil ik graag apart noemen en bedanken. Frank, voor je hulp als ik weer een file kwijt was, of graag gisteren gehad zou hebben. Jeannette, door jou heb ik inzicht gekregen in het beheer van grote hoeveelheden DNA, RNA, plasma, serum, etc. Abbas, we saved some money, thanks for the help with the black and white figures and SNPTEST.

I also want to thank all colleagues from the (very extended) CHARGE consortium reproductive traits working group. I learned a lot on biostatistics while working with you, especially Kathy and John, thanks for your email and on site support! I hope to be able to continue this collaboration with all of you much longer, speak to you on the next call.

Verder wil ik iedereen van de vijfde verdieping bedanken voor de gezelligheid tijdens de koffie, borrels, bij besprekingen, in Goes of Antwerpen. Piet, Bas, Martin, Arnold, Katy, Esmé en Varsha we hebben een hoop lol gehad bij de organisatie van de labdag 2007. Miriam, Edith en Michel dank voor jullie hulp en geduld als ik weer eens medium, antibiotica, of een techniek nodig had bij het kloneren. En Jenny, dank je wel voor je biologische input in de menarche en menopauze discussies. Ik heb er erg veel van geleerd en ik hoop dat we vanaf nu wat meer functioneel werk in de papers kunnen zetten.

Angela, paranimf op afstand-Costa Rica is toch wat ver, al sinds het eerste jaar van onze studie zijn we goede vriendinnen, en ik vind het jammer dat je er vandaag niet bij kunt zijn. Maar volgend jaar een herkansing bij jouw promotie? Linda, Guus en Anouk, vanaf nu geen samenvattingen meer om door te lezen, of kaften die doorgemaild moeten worden, maar gewoon mails over leuke dingen! Ray, dank je wel voor je email-steun als ik zat te wachten op editor commentaar en wanneer spreken we weer af? René, heel erg bedankt voor, wat ik vind, de mooiste kaft van de wereld!

Tot slot mijn familie en schoonfamilie. Ik wist dat je bij Roderik Alexander, Jeroen en Rogier cadeau kreeg. Jongens heel erg bedankt voor jullie grote bijdrage aan dit proefschrift ". " en natuurlijk de binnenbarbecues, sporten in Veldhoven, filmpjes in Houten, vakanties en heel veel lol. Els en Henny, de beste schoonouders die iemand zich maar kan wensen, bedankt voor jullie enthousiasme als ik weer een poging deed mijn werk uit te leggen. Lieve pap en mam, jullie hebben me geleerd altijd mezelf te zijn, en met beide benen op de grond te blijven. Maar nog belangrijker, jullie hebben mij altijd gesteund en gestimuleerd om vooral zo ver mogelijk te komen in mijn leven en nu zijn we hier. Zonder jullie was dit nooit gelukt, en was ik niet wie ik nu ben. Rogier, van kleine broertje tot de persoon je nu bent, ik ben trots op je! Ik hou van jullie alle 3!

En als allerbelangrijkste mijn lief, Roderik, of Rik zoals mijn meeste collega's je kennen. We zijn al sinds ons eerste studiejaar bij elkaar en je bent de afgelopen jaren mijn nul-punt geweest waar ik omheen bewoog (de laatste maanden kwam ik je soms 5x per dag tegen). Gestresst en hopeloos: "dit wordt nooit wat", "hoe kan ik zo nou werken" en "ik krijg het nooit af!" Of stuiterend van blijdschap: "ik snap het", " het is af", "ze willen mijn paper hebben" (telt een taartje en uiteten ook voor een proefschrift?) Dank je wel voor je onvoorwaardelijke steun, geduld, bloemen, enthousiasme en relativeringsvermogen. Ik hou van je, zullen we nu op vakantie?

One love
One blood
One life
You got to do what you should
One life
With each other

U2 - One

LIST OF PUBLICATIONS

A Genome-Wide Association Study identifies a locus on chromosome 7q22 to influence susceptibility for osteoarthritis. Kerkhof JM, Lories RJ, Meulenbelt I, Jonsdottir I, Valdes AM, Arp PP, Ingvarsson T, Jhamai M, Jonsson H, **Stolk L**, Thorleifsson G, Zhai G, Zhang F, Zhu Y, van der Breggen R, Carr A, Doherty M, Doherty S, Felson DT, Gonzalez A, Halldorsson BV, Hart DJ, Hauksson VB, Hofman A, Ioannidis JPA, Kloppenburg M, Lane NE, Loughlin J, Luyten FP, Nevitt MC, Parimi N, Pols HAP, van de Putte T, Rivadeneira F, Slagboom EP, Styrkársdóttir U, Tsezou A, Zmuda J, Spector TD, Stefansson, K.Uitterlinden AG,van Meurs JBJ. *Arthritis & Rheumatism*. 2009 accepted

Low prevalence of NOD2 SNPs in Behcet's disease suggests protective association in Caucasians. Kappen JH, Wallace GR, **Stolk L**, Rivadeneira F, Uitterlinden AG, van Daele PL, Laman JD, Kuijpers RW, Baarsma GS, Stanford MR, Fortune F, Madanat W, van Hagen PM, van Laar JA. *Rheumatology (Oxford)*. 2009 Sep 11. ahead of print

A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. Estrada K, Krawczak M, Schreiber S, van Duijn K, **Stolk L**, van Meurs JB, Liu F, Penninx BW, Smit JH, Vogelzangs N, Hottenga JJ, Willemsen G, de Geus EJ, Lorentzon M, von Eller-Eberstein H, Lips P, Schoor N, Pop V, de Keijzer J, Hofman A, Aulchenko YS, Oostra BA, Ohlsson C, Boomsma DI, Uitterlinden AG, van Duijn CM, Rivadeneira F, Kayser M. *Hum Mol Genet*. 2009 Sep 15;18(18):3516-24.

Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Perry JR, **Stolk L**, Franceschini N, Lunetta KL, Zhai G, McArdle PF, Smith AV, Aspelund T, Bandinelli S, Boerwinkle E, Cherkas L, Eiriksdottir G, Estrada K, Ferrucci L, Folsom AR, Garcia M, Gudnason V, Hofman A, Karasik D, Kiel DP, Launer LJ, van Meurs J, Nalls MA, Rivadeneira F, Shuldiner AR, Singleton A, Soranzo N, Tanaka T, Visser JA, Weedon MN, Wilson SG, Zhuang V, Streeten EA, Harris TB, Murray A, Spector TD, Demerath EW, Uitterlinden AG, Murabito JM. *Nat Genet*. 2009

Loci at chromosomes 13, 19 and 20 influence age at natural menopause. **Stolk L**, Zhai G, van Meurs JB, Verbiest MM, Visser JA, Estrada K, Rivadeneira F, Williams FM, Cherkas L, Deloukas P, Soranzo N, de Keyzer JJ, Pop VJ, Lips P, Lebrun CE, van der Schouw YT, Grobbee DE, Witteman J, Hofman A, Pols HA, Laven JS, Spector TD, Uitterlinden AG. *Nat*

Genet. 2009

Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, **Stolk L**, Nica A, Inouye M, Hofman A, Stephens J, Wheeler E, Arp P, Gwilliam R, Jhamai PM, Potter S, Chaney A, Ghori MJ, Ravindrarajah R, Ermakov S, Estrada K, Pols HA, Williams FM, McArdle WL, van Meurs JB, Loos RJ, Dermitzakis ET, Ahmadi KR, Hart DJ, Ouwehand WH, Wareham NJ, Barroso I, Sandhu MS, Strachan DP, Livshits G, Spector TD, Uitterlinden AG, Deloukas P. *PLoS Genetics*. 2009 Apr;**5**(4):e1000445

Growth hormone dose in growth hormone-deficient adults is not associated with IGF-1 gene polymorphisms. Meyer S, Schaefer S, Ivan D, **Stolk L**, Arp P, Uitterlinden AG, Nawroth PP, Plöckinger U, Stalla GK, Tuschy U, Weber MM, Weise A, Pfützner A, Kann PH, Buchfelder M, Brabant G, Brunnmüller U, Fassbender W, Faust M, Führer-Sakel D, Kann PH, Kreitschmann-Andermahr I, Koenig J, Said E, Schneider H, Wallaschofski H. *Pharmacogenomics*. 2009;**10**(2):293-302.

A functional polymorphism in the catechol-O-methyltransferase gene is associated with osteoarthritis-related pain. van Meurs JB, Uitterlinden AG, **Stolk L**, Kerkhof HJ, Hofman A, Pols HA, Bierma-Zeinstra SM. *Arthritis Rheum*. 2009;**60**(2):628-9.

The effect of catechol-O-methyltransferase Met/Val functional polymorphism on smoking cessation: retrospective and prospective analyses in a cohort study. Omidvar M, **Stolk L**, Uitterlinden AG, Hofman A, Van Duijn CM, Tiemeier H. *Pharmacogenet Genomics*. 2009;**19**(1):45-51.

The RIZ Pro704 insertion-deletion polymorphism, bone mineral density and fracture risk: the Rotterdam study. **Stolk L**, van Meurs JB, Arp PP, Hofman A, Pols HA, Uitterlinden AG. *Bone*. 2008;**42**(2):286-93.

The catechol-O-methyltransferase Met158 low-activity allele and association with nonvertebral fracture risk in elderly men. **Stolk L**, van Meurs JB, Jhamai M, Arp PP, van Leeuwen JP, Hofman A, de Jong FH, Pols HA, Uitterlinden AG. *J Clin Endocrinol Metab*. 2007;**92**(8):3206-12.

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. Rivadeneira F, van Meurs JB, Kant J, Zillikens MC, **Stolk L**, Beck TJ, Arp P, Schuit SC, Hofman A, Houwing-Duistermaat JJ, van Duijn CM, van Leeuwen JP, Pols HA, Uitterlinden AG. *J Bone Miner Res*. 2006;**21**(9):1443-56.

Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. Schuit SC, de Jong FH, **Stolk L**, Koek WN, van Meurs JB, Schoofs MW, Zillikens MC, Hofman A, van Leeuwen JP, Pols HA, Uitterlinden AG. *Eur J Endocrinol*. 2005;**153**(2):327-34.

PHD-PORTFOLIO

Name PhD student:	Lisette Stolk
Erasmus MC Department:	Internal Medicine – Genetic Laboratory
Research School:	NIHES / MolMed
PhD period:	November 15, 2004 – December 1, 2009
Promotor:	Prof. Dr. A.G. Uitterlinden
Supervisor:	Dr. J.B.J. van Meurs

1. PhD training

	Year	Workload
General academic skills		
- Biomedical English Writing and Communication	2006	13x3 hours
- Annual Course Molecular Medicine	2006	1 week
Research skills		
- Classical Methods for Data-analysis	2005	4 weeks
- Modern Statistical Methods	2005	3 weeks
- Study Design	2006	4 weeks
In-depth courses		
- Genetic-epidemiologic Research Methods	2007	4 weeks
- Advances in Population-based Studies of Complex Genetic Disorders	2007	1 week
- Genetic Linkage Analysis: Model Based Analysis	2006	1 week
- Genetic Linkage Analysis: Model Free Analysis	2006	1 week
- ECTS Training Course: Genetic Aspects of Bone Disease	2006	1 day
- SNP's and Human Disease	2005	1 week
- Micro-array en DNA-chip Technologie, Hogeschool Leiden	2005	2 days
(Inter)national conferences		
- 2nd Joint Meeting of the European Calcified Tissue Society and the International Bone and Mineral Society (Geneva, Switzerland)	2005	June 25-29
- 15th Annual Meeting of the Dutch Society for Calcium and Bone Metabolism (Zeist, the Netherlands)	2005	Oct 27-28
- 33rd European Symposium on Calcified Tissues (Prague, Czech Republic)	2006	May 10-14
- 16th Annual Meeting of the Dutch Society for Calcium and Bone Metabolism (Zeist, the Netherlands)	2006	Nov 9-10
- 34th European Symposium on Calcified Tissues (Copenhagen, Denmark)	2007	May 5-9
- 35th European Symposium on Calcified Tissues (Barcelona, Spain)	2008	May 24-28
- 30th ASBMR Annual Meeting (Montreal, Canada)	2008	Sept 12-16
- 36th European Symposium on Calcified Tissues (Vienna, Austria)	2009	May 23-27
- European Human Genetics Conference (Vienna, Austria)	2009	May 23-26
Seminars and workshops		
- 9th Molecular Medicine Day (Rotterdam, the Netherlands)	2005	Feb 9
- 10th Molecular Medicine Day (Rotterdam, the Netherlands)	2006	Feb 1
- KNAW Colloquium – The Role of DNA Polymorphisms in Complex Traits and Disease + Master Classes (Amsterdam, the Netherlands)	2006	March 14-17
- 11th Molecular Medicine Day (Rotterdam, the Netherlands)	2007	Feb 15
- KNAW Conference – The Role of DNA Polymorphisms in Complex Traits and Diseases + Master Classes (Amsterdam, the Netherlands)	2008	March 18-21
- 3rd CHARGE Face-to-Face meeting (Rotterdam, the Netherlands)	2009	April 16-17
- Endo retreat (Rotterdam, the Netherlands)	2009	May 14
- 4th CHARGE Face-to-Face meeting (Washington, DC, USA)	2009	Oct 29-30

Presentations

- Genetic Variations in the Estrogen Pathway. *Molecular Medicine Day 2005* — poster
- The Low Activity 158Met Variant of the *COMT* Gene is Associated with Increased Fracture Risk in Elderly Men. *European Calcified Tissue Society and the International Bone and Mineral Society 2005* — poster
- The Low Activity 158Met Variant of the *COMT* Gene is Associated with Increased Fracture Risk in Elderly Men. *Dutch Society for Calcium and Bone Metabolism 2005* — oral
- High Resolution Genotyping of the *ESR1* Gene and Association with Bone Parameters and Osteoporosis. *European Symposium on Calcified Tissues 2006* — oral
- No effect of the RIZ1 Pro704 insertion deletion polymorphism on bone mineral density and fracture risk in elderly men and women. *Dutch Society for Calcium and Bone Metabolism 2006* — oral
- No Effect of the *RIZ1* Pro704 Insertion Deletion Polymorphism on Bone Mineral Density and Fracture Risk in Elderly Men and Women. *Molecular Medicine Day 2007* — poster
- Multilocus Analysis of Estrogen Pathway Polymorphisms and Vertebral Fracture Risk in Elderly Women: An Exercise *European Symposium on Calcified Tissues 2007* — poster
- Novel loci influencing menopause are identified by a genome wide association study. *European Symposium on Calcified Tissues 2008* — oral
- Genetic variation of the SHBG gene, SHBG levels and bone parameters. *ASBMR Annual Meeting 2008* — poster
- Chromosomal Loci Associated with Age at Natural Menopause. *Endo retreat 2009* — oral
- Gender-specific associations of estrogen pathway genes with bone endpoints. *European Symposium on Calcified Tissues 2009* — poster
- Loci on Chromosome 19 and 20 are Associated with Age at Natural Menopause: a Meta-analysis of 10,399 Women. *European Human Genetics Conference 2009* — oral

2. Teaching activities

	Year	
Lecturing		
- Vaardigheids Onderwijs second year Medical students – A woman with Hypercortisolism	2006	March
- Honours Progam – Human Genotyping Facility Visit	2008	Feb 19
- Molecular Medicine – Applied Bioinformatics course – Analyzing SNPs using Bioinformatic Tools	2009	April 24
- Vaardigheids Onderwijs second year Medical students – A woman with Hypercortisolism	2009	March
- Erasmus Summer Program – Genomics in Molecular Medicine – SNP databases	2009	Aug 19
Supervising practicals		
- Molecular Medicine Postgraduate School - SNP's and Human Disease	2005	Nov
- Molecular Medicine Postgraduate School - SNP's and Human Disease	2006	Oct 30 – Nov 3
- NIHES – Practical MSc Clinical Research	2007	Feb 12-13
- Molecular Medicine Postgraduate School - SNP's and Human Disease	2007	Nov 5-9
- Molecular Medicine Postgraduate School - SNP's and Human Disease	2008	Nov 3-7
- Molecular Medicine Postgraduate School - SNP's and Human Disease	2009	Nov 2-6
Supervising Bachelor's thesis		
- Arnoud Boot – Hogeschool Rotterdam – Biology and Medical Laboratory Research – Quantitative methylation detection with the Sequenom MassARRAY EpiTYPER	2007 / 2008	7 months