Genetic studies of the regulation of bone parameters and serum testosterone

Joel Eriksson

Department of Internal Medicine and Clinical Nutrition

Institute of Medicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2018

Genetic studies of the regulation of bone parameters and serum testosterone

ABSTRACT

Osteoporosis is a common disease characterized by low bone mass and microarchitectural deterioration of bone tissue leading to increased risk of fracture and represents a huge economic burden on health care systems. The main aim of this thesis was therefore to try to identify new genetic variants associated with different bone parameters that could serve as potential pharmaceutical targets in the future and to evaluate the clinical utility of these variants for fracture prediction.

We used several well-characterized cohorts and performed the largest genome-wide association studies to that date on DXA-derived areal bone mineral density (aBMD), which is used clinically, and trabecular and cortical volumetric BMD, measured by the more specific peripheral quantitative computed tomography. We identified many genetic variants associated with bone parameters and the clinical endpoint fractures. The genetic variants associated with aBMD predicted incident fractures, but the magnitude of these associations was substantially reduced after adjustment for aBMD. Thus, the clinical utility of these genetic variants for fracture prediction is limited when aBMD is known.

Low serum testosterone (T) levels have been linked to an increased risk of osteoporosis in men. Observational studies have also demonstrated that obesity is strongly associated with low serum T, but the direction and causality of this relationship is unclear. The second objective of this thesis was therefore to determine whether low T causes obesity, or vice versa. Hence, using a bi-directional Mendelian randomization analysis, we found evidence of a causal effect of body mass index (BMI) on serum T, whereas no evidence was found supporting a causal effect of serum T on BMI in men. The studies herein have identified a number of novel loci associated with different bone parameters and, hence, fracture risk. These findings may result in the development of novel pharmaceutical therapies for osteoporosis and the improvement of prediction models with new biomarkers to identify patients at risk. In addition, we demonstrated that there is a causal effect of BMI on serum T in men, suggesting that population-level interventions to reduce BMI are expected to increase serum T in men.

Keywords: osteoporosis, bone mineral density, genetics, genome-wide association study, testosterone, body mass index, Mendelian randomization

SAMMANFATTNING PÅ SVENSKA

Benskörhet (osteoporos) är en folksjukdom som kostar samhället enorma summor årligen. Den kännetecknas av mikrostrukturella förändringar i benet samt av att benmassan minskar, vilket tillsammans ökar risken för fraktur. Det huvudsakliga syftet med den här avhandlingen har därför varit att identifiera nya genetiska varianter som skulle kunna utgöra mål för framtida läkemedel mot benskörhet, samt att utvärdera den kliniska nyttan av dessa för frakturprediktion.

I första delarbetet använde vi ett flertal väldefinierade kohorter och genomförde den då största genomtäckande associationsstudien (eng. GWAS) på bentäthet mätt med tvådimensionell röntgenteknik (s.k. DXA). 56 genetiska områden associerade med bentäthet i höften och/eller ländryggen identifierades, varav 14 även var associerade med fraktur. Dessa fynd kan i framtiden bidra till upptäckten av nya läkemedel mot benskörhet och till en förbättring av de modeller som används idag för att identifiera patienter med hög risk för en framtida fraktur.

I det andra delarbetet utvärderades den kliniska nyttan av de genetiska varianter som identifierats i delarbete I för tvådimensionell bentäthet, förlust av tvådimensionell bentäthet över tid, samt för frakturer i en population bestående av äldre män och kvinnor. För detta beräknades två genetiska risksummor för varje individ. En för de genetiska varianter som var associerade med bentäthet och en för de som var associerade med frakturer. Båda risksummor var associerade med bentäthet, men inte med bentäthetsförlust vilket talar för att olika genetiska mekanismer styr vår maximala bentäthet kontra hur snabbt vi tappar i bentäthet med stigande ålder. Båda genetiska risksummor var associerade med fraktur, men denna association försvagades markant när modellerna justerats för uppmätt bentäthet. Den kliniska nyttan av dessa genetiska varianter för att prediktera frakturer är därför begränsad när bentätheten är känd.

Då tvådimensionell bentäthet inte kan skilja på kortikalt (kompakt) och trabekulärt (spongiöst) ben utförde vi i delarbete III en genomtäckande associationsstudie på tredimensionell bentäthet mätt med datortomografi som kan separera kortikalt och trabekulärt ben. Fyra kohorter med kaukasiska män och kvinnor i olika åldrar ingick i studien och vi identifierade olika genetiska områden för kortikalt ben jämfört med trabekulärt ben.

Låga nivåer av det manliga könshormonet testosteron har visats vara associerat med benskörhet och frakturer hos män. Observationsstudier har också visat att övervikt är associerat med lågt testosteron hos män, men det är oklart om det rör sig om ett orsakssamband och isåfall i vilken riktning det sker. Det andra syftet med den här avhandlingen var därför att avgöra om övervikt orsakar låga testosteronnivåer eller tvärtom. I delarbete IV använde vi därför mendelsk randomisering på fem kohorter av kaukasiska män. Vi fann att en standarddeviation lägre BMI höjde testosteronet med 13-15%, men fann inga bevis för att låga testosteronnivåer gav övervikt.

Sammantaget så har studierna i denna avhandling identifierat ett antal områden i det mänskliga DNA:t som är associerade med olika benparametrar och därmed frakturrisk. Dessa fynd kan resultera i nya läkemedelsbehandlingar för benskörhet och en förbättring av modellerna som används för att identifiera patienter med hög risk. Utöver detta har vi visat att övervikt orsakar en sänkning av det manliga könshormonet testosteron hos män. Detta innebär att interventioner som minskar övervikt på populationsnivå förväntas höja testosteronnivåerna hos män.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP, Koller DL, Li G, Liu CT, Minster RL, Moayyeri A, Vandenput L, Willner D, Xiao SM, Yerges-Armstrong LM, Zheng HF, Alonso N, **Eriksson J**, Kammerer CM, Kaptoge SK, Leo PJ, Thorleifsson G, Wilson SG, Wilson JF, Aalto V, Alen M, Aragaki AK, Aspelund T, Center JR, Dailiana Z, Duggan DJ, Garcia M, Garcia-Giralt N, Giroux S, Hallmans G, Hocking LJ, Husted LB, Jameson KA, Khusainova R, Kim GS, Kooperberg C, Koromila T, Kruk M, Laaksonen M, Lacroix AZ, Lee SH, Leung PC, Lewis JR, Masi L, Mencej-Bedrac S, Nguyen TV, Nogues X, Patel MS, Prezelj J, Rose LM, Scollen S, Siggeirsdottir K, Smith AV, Svensson O, Trompet S, Trummer O, van Schoor NM, Woo J, Zhu K, Balcells S, Brandi ML, Buckley BM, Cheng S, Christiansen C, Cooper C, Dedoussis G, Ford I, Frost M, Goltzman D, González-Macías J, Kähönen M, Karlsson M, Khusnutdinova E, Koh JM, Kollia P, Langdahl BL, Leslie WD, Lips P, Ljunggren Ö, Lorenc RS, Marc J, Mellström D, Obermayer-Pietsch B, Olmos JM, Pettersson-Kymmer U, Reid DM, Riancho JA, Ridker PM, Rousseau F, Slagboom PE, Tang NL, Urreizti R, Van Hul W, Viikari J, Zarrabeitia MT, Aulchenko YS, Castano-Betancourt M, Grundberg E, Herrera L, Ingvarsson T, Johannsdottir H, Kwan T, Li R, Luben R, Medina-Gómez C, Palsson ST, Reppe S, Rotter JI, Sigurdsson G, van Meurs JB, Verlaan D, Williams FM, Wood AR, Zhou Y, Gautvik KM, Pastinen T, Raychaudhuri S, Cauley JA, Chasman DI, Clark GR, Cummings SR, Danoy P, Dennison EM, Eastell R, Eisman JA, Gudnason V, Hofman A, Jackson RD, Jones G, Jukema JW, Khaw KT, Lehtimäki T, Liu Y, Lorentzon M, McCloskey E, Mitchell BD, Nandakumar K, Nicholson GC, Oostra BA, Peacock M, Pols HA, Prince RL, Raitakari O, Reid IR, Robbins J, Sambrook PN, Sham PC, Shuldiner AR, Tylavsky FA, van Duijn CM, Wareham NJ, Cupples LA, Econs MJ, Evans DM, Harris TB, Kung AW, Psaty BM, Reeve J, Spector TD, Streeten EA, Zillikens MC, *Thorsteinsdottir U, *Ohlsson C, *Karasik D, *Richards JB, *Brown MA, *Stefansson K,

*Uitterlinden AG, *Ralston SH, *Ioannidis JP, *Kiel DP, *Rivadeneira F.

* shared senior authorship

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

Nature Genetics **2012**;44(5):491-501.

II.   **Eriksson J**, Evans D, Nielson C, Shen J, Srikanth P, Hochberg M, McWeeney S, Cawthon P, Wilmot B, Zmuda J, Tranah G, Mirel D, Challa S, Mooney M, Crenshaw A, Karlsson M, Mellström D, Vandenput L, Orwoll E, Ohlsson C.

Limited clinical utility of a genetic risk score for the prediction of fracture risk in elderly subjects.

Journal of Bone and Mineral Research **2015**;30(1):184-94.

III.   *Paternoster L, *Lorentzon M, *Lehtimäki T, ***Eriksson J**, Kähönen M, Raitakari O, Laaksonen M, Sievänen H, Viikari J, Lyytikäinen LP, Mellström D, Karlsson M, Ljunggren O, Grundberg E, Kemp JP, Sayers A, Nethander M, Evans DM, Vandenput L, Tobias JH, Ohlsson C.

* shared first authorship

Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure.

PLoS Genetics **2013**;9(2):e1003247.

IV.   **Eriksson J**, Haring R, Grarup N, Vandenput L, Wallaschofski H, Lorentzen E, Hansen T, Mellström D, Pedersen O, Nauck M, Lorentzon M, Nystrup Husemoen LL, Völzke H, Karlsson M, Baumeister SE, Linneberg A, Ohlsson C.

Causal relationship between obesity and serum testosterone status in men: A bidirectional mendelian randomization analysis.

PLoS One **2017**;12(4):e0176277.

Article reprints were used with permission from the publishers.

CONTENT

ABBREVIATIONS

aBMD   areal Bone Mineral Density

AUC   Area Under (the receiver operating characteristic) Curve

AR   Androgen Receptor

BMD   Bone Mineral Density

BMI   Body Mass Index

BMP   Bone Morphogenetic Proteins

CI   Confidence Interval

CVD   CardioVascular Disease

DXA   Dual-energy X-ray Absorptiometry

E2   Estradiol

eQTL   expression Quantitative Trait Loci

ERα   Estrogen Receptor alpha

ERβ   Estrogen Receptor beta

FRAX   Fracture Risk Assessment Tool

FSH   Follicle-Stimulating Hormone

GOOD   Gothenburg Osteoporosis and Obesity Determinants

GWAS   Genome-Wide Association Study

HR-MRI   High Resolution Magnetic Resonance Imaging

HR-pQCT	High-Resolution peripheral Quantitative Computed Tomography
IDI	Integrated Discrimination Improvement
LH	Luteinizing Hormone
LHRH	Luteinizing Hormone-Releasing Hormone
LOH	Late-Onset Hypogonadism
MAF	Minor Allele Frequency
MR	Mendelian Randomization
MrOS	Osteoporotic Fractures in Men
NRI	Net Reclassification Improvement
ROC	Receiver Operating Characteristic
SNP	Single Nucleotide Polymorphism
T	Testosterone
TRT	Testosterone Replacement Therapy
vBMD	volumetric Bone Mineral Density

1 INTRODUCTION

Osteoporosis is a common disease characterized by low bone mass and microarchitectural deterioration of bone tissue leading to increased risk of fracture. The risk of an osteoporotic fracture is believed to be as high as 46.4% for women and 22.4% for men in Scandinavia after the age of 50 (1). It has been estimated that the disease accounts for more than one million new fracture cases each year, representing a huge economic burden on health care systems, with costs of several billions of dollars each year in the United States alone and expected to rise considerably by the year 2025 (2).

Today´s available osteoporosis treatments have led to a substantial reduction in vertebral fracture risk in patients with osteoporosis. However, non-vertebral fracture risk has only been marginally improved (3). Hence, there is a dire need for new pharmaceutical targets for non-vertebral fractures as well as improved prediction models that identify those patients who would benefit most from osteoporotic treatment.

Sex hormones have been linked to a number of diseases, including osteoporosis (4-7) and an increased risk of falls (8). Loss of estrogens or androgens increases the rate of bone remodeling by removing restraining effects on osteoblastogenesis and osteoclastogenesis, and also causes a focal imbalance between bone resorption and bone formation (9). In fact, low serum estradiol (E2) and low serum testosterone (T) predict clinical vertebral fractures, nonvertebral osteoporosis fractures, and hip fractures in older men (8, 10).

Low T has also been linked to high body mass index (BMI) and high risk of cardiovascular disease (CVD) (11), but causality has not yet been established. This constitutes an important clinical challenge, since we are either aiming to reduce weight to increase T, or to increase T, via T treatment, to reduce BMI and risk of CVD. Potentially dangerous side effects also need to be addressed. For instance, some studies have indicated potentially dangerous side effects of T treatment on the prostate as well as an increased risk of CVD (12).

1.1 BONE STRUCTURE

Bones can be categorized as flat (skull, scapulae, sternum etc) or long, tubular bones (vertebras, appendicular bones etc). Regardless of their shape and localization, virtually all bones consist of two types of bone tissue: the compact outer surface called cortical bone and the spongy inside called trabecular bone (Figure 1). Cortical bone is stiffer, harder and due to its compact structure, heavier, than trabecular bone. It is composed of lamellae concentrically arranged around a centrally situated canal. This is referred to as an osteon, or a Haversian system. Between the lamellae are cavities, where bone cells called osteocytes are embedded (13). Each cavity is connected to others through small channels called canaliculi. This structure makes up a porous appearance. The volume fraction of these pores, referred to as cortical porosity, correlates well with the natural decrease in bone density in adults (14). Microscopically, trabecular bone consists of plates (trabeculae) and bars of bone adjacent to small, irregular cavities that contain bone marrow. The trabeculae are organized in a way to provide maximum strength similar to braces that are used to support a building and are aligned towards the mechanical load distribution that the bone experiences (15). Due to different requirements, bones differ in their distribution of cortical and trabecular bones.

Bone biology

Bones undergo constant reconstruction, where resorption (performed by osteoclast cells) and formation (performed by osteoblast cells) occur at different sites of the bone simultaneously. During the first two decades of our lives, known as the modeling phase, formation exceeds resorption, resulting in a net increase in bone mass. During this time there are also major changes to the gross morphology of the bone, including longitudinal growth of the long bones by bone formation at the endplates of the bones (epiphyseal growth plates) as well as radial growth due to bone formation on the outer surface (periosteal apposition) of the cortex and resorption on the inner surface (endosteal resorption). Following the modeling phase, is the remodeling phase, where there is a balance between bone formation and bone resorption. Remodeling enables the bone to respond and adapt to load-induced strain, replace old or damaged bone tissue and to maintain calcium homeostasis (16). While it occurs in both cortical and trabecular bone, there is evidence that suggests that cortical and trabecular bone are affected differently throughout life, indicating that the genetic control of cortical and

trabecular bone might differ (17). For example, in mice it has been demonstrated that the WNT16 gene has a large effect on cortical bone, via increased cortical thickness and decreased cortical porosity, and the risk of non-vertebral fractures, while no substantial effect has been seen on trabecular bone volume fraction (18).
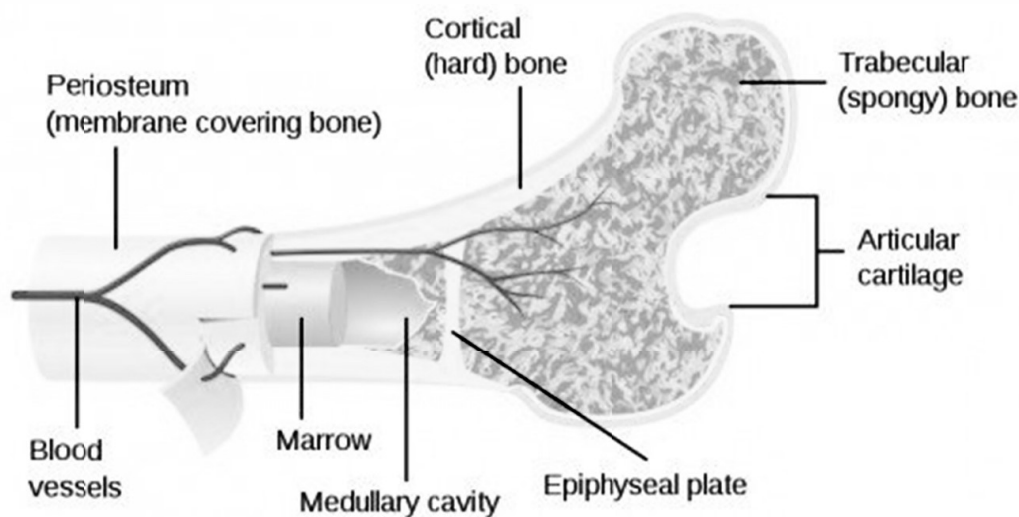

Figure 1. Bone structure. Source: Pbroks13, via Wikimedia Commons, CC BY 3.0 License

1.2   BONE ASSESSMENT

Dual-energy X-ray Absorptiometry (DXA) is the golden standard for assessing areal bone mineral density (aBMD), i.e. the amount of bone mineral in bone tissue. It uses two low radiation X-ray beams with different energy levels. Using the fact that absorbed energy is a function of density, it is possible to differentiate between different tissues. The resulting image produced by the 2-dimensional DXA, therefore provides aBMD as grams per square centimeter, as well as bone mineral content and bone area. Although the DXA method is a robust method that provides reproducible results that strongly correlate to fracture risk, it cannot provide information on the geometrical structure and true volumetric bone mineral density (vBMD) and, thus, cannot distinguish between cortical and trabecular bone. With the three dimensional technique quantitative computed tomography (QCT) it is possible to study macro-structural properties like cortical geometry and vBMD (mg per cm^3) of both the cortical and trabecular compartment (19). Due to factors such as price, inconvenience for the patient and radiation dose,

peripheral QCT (pQCT) is often only used in a research setting for assessment of appendicular bones, e.g. arm or leg. A high resolution pQCT (HR-pQCT) or high resolution magnetic resonance imaging (HR-MRI) offers higher resolution (50-100 μm) enabling quantification of trabecular microstructure and an estimate of cortical porosity in humans. Whereas DXA can be found in a clinical setting, pQCT, HR-pQCT and HR-MRI are presently solely used in research (20-22).

1.3   DEFINITION OF OSTEOPOROSIS

Osteoporosis is defined as having an areal BMD (aBMD) at the hip or lumbar spine at least 2.5 SD values below the population average in young healthy individuals as measured by DXA (23). Although osteoporosis increases the risk of fractures in general, typical osteoporotic sites include the hip, wrist, humerus and vertebra (24). Of all fractures, hip and vertebral fractures have the greatest negative impact on quality of life and mortality (25, 26). The risk for hip fractures increases exponentially with age, which is believed to be due to both a decrease in aBMD at the proximal femur, decreased bone quality, as well as an increased risk of falls (27)

Although aBMD explains about 60-70% of the variance in bone strength (28, 29), only about half of the women with a hip fracture had total hip aBMD values consistent with osteoporosis (30). Part of the explanation might be that DXA cannot distinguish between cortical and trabecular bone. Another explanation might be found in differences in bone size as we've demonstrated that although individuals with a constitutive predisposition to higher rates of bone resorption have a lower areal and/or cortical BMD, any adverse effect on bone strength and fracture risk may be at least partially compensated for by greater bone size (31).

1.4   HERITABILITY OF OSTEOPOROSIS

There is compelling evidence that our genetic heritage is a major contributor to overall risk of osteoporosis and fractures. In fact, twin and family studies have provided evidence of substantial (50-85 %) heritability for aBMD and, thus, risk of osteoporosis (32, 33). Twin and family studies have also demonstrated a clear heritability for aBMD loss (40-50%) (34-36), $\cong$50% for hip and forearm fractures and lower ($\cong$24%) for vertebral fractures (37-39).

1.5 FRACTURE PREDICTION

In order to identify patients with the highest risk of osteoporotic fractures, tools, such as Fracture Risk Assessment Tool (FRAX, https://www.sheffield.ac.uk/FRAX/), have been developed to help clinicians determine when treatment is indicated. These tools integrate the risks associated with clinical risk factors with or without BMD (40-42) and currently include age, sex, weight, height, previous fractures, parent fractured hip, smoking status, use of glucocorticoids, rheumatoid arthritis, secondary osteoporosis and alcohol intake above 3 units per day with or without addition of femoral neck BMD. Identifying genetic determinants of osteoporosis and fracture risk might improve prediction models such as FRAX, and/or, serve as a basis for new targets for pharmaceutical intervention.

1.6 TESTOSTERONE, BONE MASS AND FRACTURE RISK

T is a steroid synthesized from cholesterol in several steps. It can be transformed into the more potent androgen dihydrotestosterone by the 5a-reductase enzymes or converted into E2 by the aromatase enzyme (Figure 2). It exerts its action through binding to and activation of the androgen receptor (AR) or, indirectly, after aromatization to E2 via estrogen receptor alpha (ERα) or estrogen receptor beta (ERβ) (43). The testes in males and, to a lesser extent, the ovaries in females are the main producers of T in males and females, respectively, but small amounts are also produced by the adrenal glands. Serum T levels in males are approximately seven times higher than in females (44).
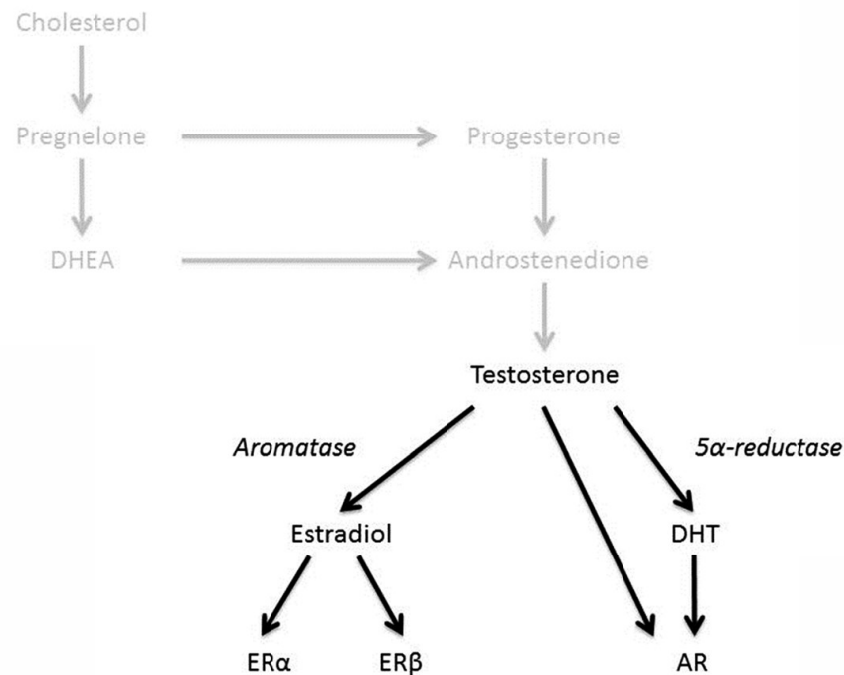
Figure 2. Testosterone.pathway. Testosterone can be converted to DHT or Estradiol. ERα = Estrogen receptor alpha, ERβ = Estrogen receptor beta, AR = Androgen receptor, DHT = Dihydrotestosterone.

T is largely bound to two plasma proteins. Most of the circulating T (50–60%) is bound with high affinity to sex hormone-binding globulin (SHBG), while a smaller fraction (40–50%) is bound loosely to albumin, and 1–3% is unbound and termed free T (45).

Although animal studies have shown that T, via activation of the AR, regulate bone mass in male rodents, it seems that aBMD is mainly affected by E2 and not T in males (8, 46, 47). In fact, the more modest effect of T on fracture risk is proposed to be mediated by effects on muscle strength and risk of falls rather than due to an effect on BMD (48, 49).

1.7 DEFINITION OF HYPOGONADISM

Hypogonadism is classically defined as primary or secondary. In primary hypogonadism (hypergonadotrophic hypogonadism), it is the testes that fail to produce adequate amounts of T, despite elevated gonadotropin levels (low T and increased luteinizing hormone [LH] and follicle-stimulating hormone [FSH]). In secondary hypogonadism (hypogonadotrophic hypogonadism) failure occurs at the hypothalamus–pituitary level (low T and low gonadotropin or LHRH levels) (50). Many chronic illnesses are associated with low T levels but do not fit into the two classical endocrine situations described above. These syndromes, with clinical symptoms of hypogonadism, are acquired in adulthood and often exhibit functional hyposecretion at the level of both pituitary and testis (51). It should be noted, however, that although suppressed serum T is common in ageing men, only a small proportion of them develop the genuine syndrome of low T associated with diffuse sexual (e.g., erectile dysfunction), physical (e.g. loss of vigor and frailty) and psychological (e.g., depression) symptoms (52). The European Male Ageing Study (EMAS) has recently defined a strict diagnostic criteria for late onset hypogonadism (LOH) which includes three sexual symptoms (lessened sexual thoughts, weakened morning erections and erectile dysfunction), and either repeated (at least twice) serum total T level <8 nmol/l, or serum total T of 8–11 nmol/l and free T <220 pmol/l (53). By these criteria, only about 2% of 40- to 80-year-old men have LOH. In particular obesity, but also impaired general health, is a more common cause of low T than chronological age *per se*.

1.8 HERITABILITY OF SERUM TESTOSTERONE

It is well established that serum T is negatively associated with age and BMI and positively associated with smoking (54). Of these, BMI has the strongest correlation with serum T. Studies in male twins indicate that there is a strong heritability for serum T, with genetic factors accounting for 65% of the variation in serum T (55), but few genetic variants associated with T have been identified. The largest genome-wide association study (GWAS) to date explaining less than 5% of total variance, was led by our research group and identified two single nucleotide polymorphisms (SNPs) at the SHBG locus and one near the FAM9B locus on the X chromosome independently associated with serum T in males (56). Interestingly, the SNP (or one which completely correlates with it) at the FAM9B locus has later been shown to be associated with BMD and E2 in males (57, 58).

7

Also, one of the SHBG SNPs identified was non-synonymous, meaning that the polymorphism gives rise to a change in amino acid sequence, which resulted in an affected binding affinity of serum T to SHBG (56).

1.9   MEASUREMENT OF SERUM TESTOSTERONE

Serum T is commonly measured by immunoassay-based techniques. These techniques have, however, a questionable specificity, especially at lower concentrations (59, 60). Mass spectrometry (MS) is the golden standard for the quantification of sex steroids in serum samples (61).

1.10 OBESITY

Obesity is often defined simply as a condition of abnormal or excessive fat accumulation in adipose tissue, to the extent that health may be impaired (62). Overweight is defined as having a BMI between 25 and 30 while obesity is defined as having a BMI above 30 kg/m^2. BMI is a low cost population-level measure of obesity. Although it is the most widely used metric, it does not account for the wide variation in body fat distribution, and may not correspond to the same degree of fatness or associated health risk in different individuals and populations (62).

Using a hypothesis-free approach, a GWAS offers a technique that could identify previously unknown genetic markers associated with a trait through multiple linear regression models, where each regression tests the association between the trait and an individual SNP. Most GWASs focusing on obesity have used BMI as the phenotypic trait, but there have also been studies with somewhat smaller sample sizes that target other metrics such as waist circumference and waist-hip-ratio. In a recent co-authored study, six anthropometric traits (BMI, height, weight, waist and hip circumference and waist-to-hip ratio) were combined using a principal components analysis (63), revealing six new loci associated with body shape. Hence, our metrics for obesity does not single-handedly capture the nature of body shape and obesity. Furthermore, we have also found evidence of age-dependent genetic effects on obesity (64, 65).

Although using specific and/or multiple metrics in a study would provide more detailed information on the different aspects of obesity than BMI alone, the availability of subjects with that metric might be an issue since a small sample size might result in a lack of statistical power. Hence, despite its obvious shortcomings as a detailed measurement of obesity, BMI is often used as it allows for large sample sizes.

1.11 TESTOSTERONE AND OBESITY-RELATED TRAITS

It has been known for about forty years that obese men have lower T compared to lean men (66). Since then, multiple cross-sectional and prospective studies have consistently found inverse correlations between both total and free T levels and adiposity in men (67). Total T levels decrease as BMI increases, partly because sex hormone binding globulin (SHBG) concentrations are reduced. Free and non-SHBG-bound T levels, however, may also decline, especially with massive obesity (68). In fact, in the HERITAGE Family Study, the well-established inverse relationship between age and total T could no longer be demonstrated after adjusting for body fat mass (69).

Low serum T has also been found to be associated with CVD (70), however, as for the association between serum T and obesity, the question of cause and effect between obesity (and the resulting obesity-related diseases) and T remains.

1.12 GENETICS

DNA

Deoxyribonucleic acid (DNA) is a molecule that stores all information necessary for the growth, development, functioning and reproduction of all known living organisms. In humans, DNA molecules consist of two strands, containing the same biological information, coiled around each other to form a double helix(71). Each strand is composed of simpler monomer units called nucleotides, where each nucleotide is composed of one of four nitrogen-containing nucleobases — cytosine (C), guanine (G), adenine (A), or thymine (T) — a sugar called deoxyribose, and a phosphate group. The nitrogenous bases of the two separate strands are bound together, according to base pairing rules (A with T, and C with G), with hydrogen bonds to make the double-stranded DNA. It is the sequence of these four nucleobases that encodes biological information. However, most of the DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve as patterns for protein sequences, but might still influence transcription.

GWAS

There is a varying degree of variation of single nucleotides within the human population. At a specific location in the DNA, 60% of all males might

have a C, whereas the remaining 40% might have an A. These single nucleotide variations are called SNPs and are common throughout our DNA. In genetics, these SNPs have enabled the study of diseases in a completely new way. Rather than looking at a few genes and their association with a disease, GWASs use millions of regression analyses, where each analysis focuses on the association between the disease and one SNP. However, rather than genotyping the whole genome (i.e. listing all of the nucleotides in order from one end of the DNA strand to the other), most previous GWASs have used chips with a specified number of specific SNPs throughout the genome. Since each SNP usually explains a very small amount of the variation in a trait (such as a disease), a very large number of subjects are needed to obtain statistical significance for some of these SNPs. This is especially so since the threshold for significance (usually $p<0.05$) needs to take into account multiple testing. After an adjustment for the number of tests, $5*10^{-8}$ was the significance threshold for GWASs for many years. Recently, due to the reduced cost of genotyping and improved imputation, the number of available SNPs for testing has increased drastically, resulting in a lowered significance threshold. Due to the massive costs of obtaining both genetic and phenotype data for a sufficiently large number of subjects, GWASs are seldom performed in single cohorts. Rather, the results of many research groups are combined together using a method called meta-analysis, which attaches weight to each group's result based on the number of subjects and standard error of the analysis. This enables researchers to share summary statistics rather than raw data between groups.

Since the chips used for genotyping differ between manufacturers and because these chips only provide a small fraction of all SNPs, a method called imputation (Figure 3) emerged as a way to calculate a predefined set of SNPs based on the inherent relationships between different SNPs (i.e. by knowing the exact value of X SNPs it is possible to calculate another Y SNPs in close vicinity). It is achieved by using known haplotypes (sets of tightly linked SNPs that tend to always occur together) in a population, for instance from the HapMap (approximately 2 million variants; https://www.genome.gov/10001688/international-hapmap-project/), or the 1000 Genomes Project (approximately 80 million variants in phase 3; http://www.internationalgenome.org/about#1000G_PROJECT) in humans.
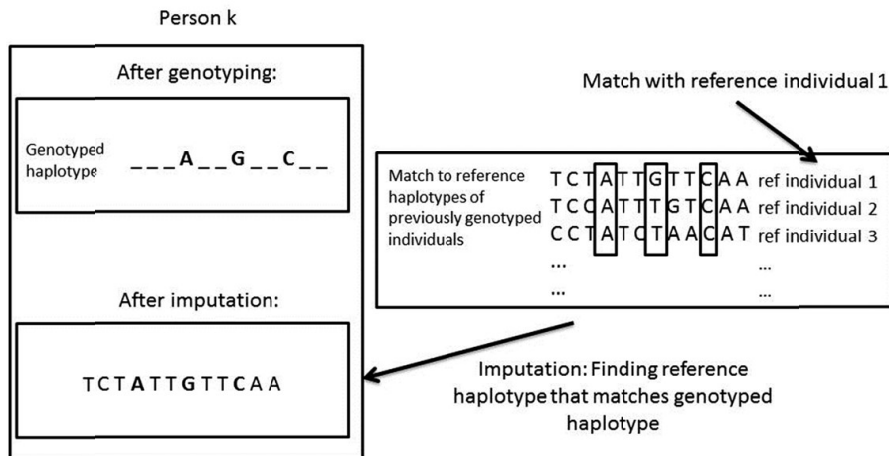
Figure 3. Genotype imputation. Variants are imputed by matching reference haplotypes to genotyped haplotypes. If more than one reference haplotype matches the genotyped haplotype, the corresponding values are added probabilistically (1/3T and 2/3 G, for example).

Today, it is even possible to get the whole DNA sequenced, which is referred to as Whole Genome Sequencing (WGS). The haplotype reference consortium ((http://www.haplotype-reference-consortium.org/) constitutes another recent effort which aims at building a much larger combined haplotype reference panel. This, in turn, enables more detailed haplotypes which could be used for the imputation of SNPs with low minor allele frequency (MAF) (72).

The imputed set of SNPs thus provides the necessary framework needed to be able to compare the results from different groups that use a variety of genotyping chips. Although quite a lot of significant associations between different SNPs and a disease have been identified, it is only in rare cases that the significant SNPs cause the decrease or increase in risk for developing the disease. Rather, it is often only associated with the causative SNP. It does, however, provide a clue to which gene(-s) that might be involved in the pathogenesis. In order to achieve more than associations between loci identified in a GWAS and the trait of interest, the results needs to be combined with other analyses as well. Translational research, where animal models are used to knock out the gene of interest is one way forward. Another way is to use expression quantitative trait loci (eQTL) analyses that focus on how different genomic loci contribute to variation in expression levels of mRNAs (which is later translated to proteins).

Two tools for graphically displaying the results from the numerous regression analyses performed in a GWAS are the Quantile-Quantile (QQ) (Figure 4) and Manhattan plots (Figure 5). The QQ plot shows the expected distribution of association test statistics (X-axis) across the million SNPs compared to the observed values (Y-axis). Any deviation from the X=Y line implies a consistent difference between cases and controls across the whole genome, suggesting bias from population stratification etc. However, if the plot shows a dotted line matching X=Y until the dotted line sharply deviates at the end, the deviating dots are likely to represent one, or more, true associations.
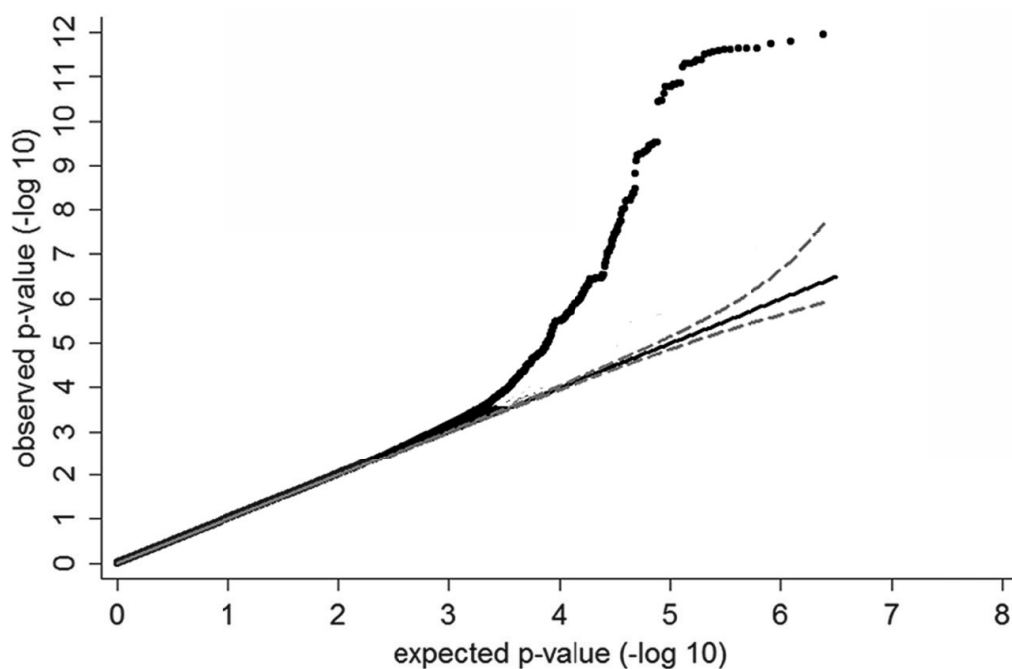

Figure 4. QQ plot. The straight line corresponds to expected p values. The demolished lines correspond to 95% confidence intervals. The dotted line corresponds to observed p values. The deviation of observed vs expected p values corresponds to significant association(-s).

Unlike the QQ plot that displays the results from the many regression analyses over the whole genome, Manhattan plots depict the results on chromosomes individually. Since there are many correlated SNPs, Manhattan plots provide an easy way to determine on which chromosome the most significantly associated SNPs reside and, since there is no correlation between SNPs on different chromosomes, whether significant SNPs constitute obviously independent signals.
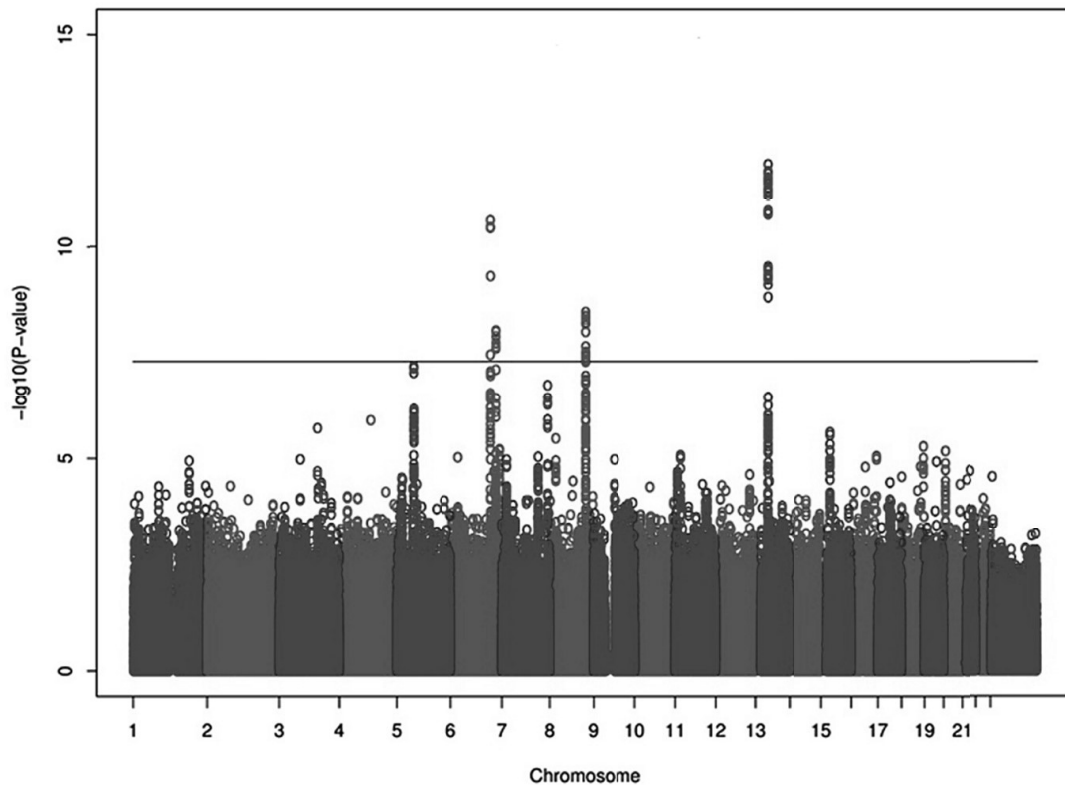
Figure 5. Manhattan plot. Observed p values per chromosome. Significant associations are above the horizontal black line ($p <= 5x10^{-8}$).

Mendelian randomization

A method called Mendelian randomization (MR) uses genetic variants in observational epidemiology to make causal inferences about modifiable (non-genetic) risk factors for disease and health-related outcomes (73). Since our genetic variants are determined at conception and remain constant throughout life, MR is not influenced by reverse causation (74). Furthermore, since random assortment of alleles occurs during gamete formation, genetic variants with effect on a modifiable exposure, for example BMI, are randomly distributed in relation to potential confounders. Under the assumption that the genetic instrument (or instrumental variable, IV, based on genetic variants) is not directly associated with the outcome, or any potential confounding variable, but rather, that the association is with the risk factor of interest, the genetic instrument divides the population into subgroups which systematically differ in the risk factor, but not in any competing risk factor. The genetically-defined subgroups are then analogous to treatment arms in a randomized controlled trial (75).

Despite the fact that MR studies are less susceptible to reverse causation and confounding than observational studies, there are limitations to the approach. These include population stratification (genetic associations reflect

latent strata in the population), pleiotropy (genes influencing multiple phenotypes), canalization (the ability of a population to produce the same phenotype regardless of variability of its environment or genotype), inadequate power and linkage disequilibrium (interdependence between genetic variants included in the same genetic instrument, such as an allele score) etc (73-75). MR analyses could be categorized as one-sample analyses (when the same cohorts are used for the IV-exposure and IV-outcome analyses) or two-sample analyses (when the set of cohorts used for the IV-exposure and IV-outcome analyses are different), or a hybrid of the two (when there is a partial overlap between the cohorts used in the IV-exposure and IV-outcome analyses).

The assumptions of MR have been discussed to great length and the exclusion restriction criteria, relating to pleiotropy (i.e. that the instrumental variable should not affect the outcome independently of the exposure) remains the most critical one (76), Possible solutions have been proposed, including that of Davey-Smith and Hemani used in paper IV that suggests that it is highly unlikely that independent IVs produce similar causal effects (74). Another method, MR-Egger, developed by Bowden et al has gained a lot of interest in the genetic epidemiology field lately (77). Rather than developing a risk sum based on individual SNPs, MR-Egger considers each SNP individually as a single instrumental variable. Under the assumption that across all genetic variants, the covariance between the effect of the IV on the outcome and the effect of the IV on the exposure is zero ('InSIDE assumption'), IVs with a stronger effect on the exposure should give less-biased MR estimates. A regression of the MR estimates on the first stage coefficients including an intercept then provides a consistent estimate of the causal effect (77). Despite its widespread use today, however, MR-Egger has some serious drawbacks of its own. Recent simulations have shown that estimates from the MR-Egger method can be more biased and have greater Type 1 error rates compared with traditional methods in settings when pleiotropic effects of multiple genetic variants act through the same confounder (78). Hence, the InSIDE assumption is crucial to the interpretation of causal inferences from the MR-Egger method in the case of pleiotropy. However, the InSIDE assumption cannot be tested and may not hold if the genetic variants used as IVs are correlated with confounders of the association between exposure and outcome. Moreover, in one-sample settings (where the association between instrument(-s) and exposure and the association between instrument(-s) and outcome are tested in the same sample) MR-Egger may suffer heavily from weak instrument bias (78).
    A new interesting, but far less studied, approach called pleiotropy-robust Mendelian randomization (PRMR) have been demonstrated using

simulations and real data to provide unbiased estimates of causal effects even when all genetic instruments violate the exclusion restriction. In order to do so, however, it requires that there is a subsample where the first stage regression (between instrumental variable and exposure) is zero. If no such subsample is available, unbiased estimates are not guaranteed, but the method could still be used as a sensitivity analysis to determine how strong the violation of the exclusion restriction would have to be in order to render the causal effect β to be 0 (78).

2 AIM

The aim of this thesis can be divided in two. The aim of paper I-III was to increase our understanding of the genetic architecture underlying various bone parameters and to evaluate the clinical use of these genetic variants for fracture risk prediction.

Low T has been shown to be associated with an increased risk of osteoporosis and fractures in men. In addition, observational studies have demonstrated that obesity is strongly associated with low serum T, but the direction and causality of this relationship is unclear. Therefore, the aim of paper IV was to evaluate the potentially causal relationship and direction between low T and obesity in men.

Specific aims:

Paper I: Identify novel genetic loci associated with DXA-derived two-dimensional aBMD

Paper II: Determine the clinical usefulness of these genetic findings for prediction of bone loss and fractures

Paper III: Use a three-dimensional pQCT to identify novel genetic loci associated with cortical and trabecular bone parameters separately

Paper IV: Determine if high BMI causes low T, or if low T causes high BMI in men using a bi-directional MR approach

3 METHODOLOGICAL CONSIDERATIONS

Several international cohorts contributed to the data used in the analyses for this thesis (see Table 1). For details regarding each cohort, please see the methods section in each paper.

Table 1. Main cohorts included in each paper.

Cohort	Country	Paper I	Paper II	Paper III	Paper IV
Gothenburg Osteoporosis and Obesity Determinants Study (GOOD)	Sweden	x		x	x
The Osteoporotic Fractures in Men Study (MrOS Sweden)	Sweden	x*	X	x*	x
The Osteoporotic Fractures in Men Study (MrOS US)	US		X		
The Study of Osteoporotic Fractures (SOF)	US		X		
The Avon Longitudinal Study of Parents and Children (Alspac)	UK			x	
Young Finns Study (YFS)	Finland	x*		x	
The Intervention 1999 Study (INTER99)	Denmark				x
The Study of Health in Pomerania (SHIP-TREND)	Germany				x
The Study of Health in Pomerania (SHIP)	Germany				x
Amish Family Osteoporosis Study (AFOS)	US	x			
Anglo-Australasian Osteoporosis Genetics Consortium (AOGC)	Australia, New Zealand, UK	x			
Cardiovascular Health Study (CHS)	US	x			
DeCODE Genetics Study (DeCODE)	Iceland	x			
Erasmus Rucphen Family (ERF)	Netherlands	x			
European Prospective Investigation into Cancer, Norfolk study (EPICNOR)	UK	x			

Framingham Heart Study (FHS)	US	x			
Health Aging and Body Composition (HABC)	US	x			
Hong Kong Osteoporosis Study (HKOS)	China	x			
Indiana Genetics of Bone Fragility Study (Indiana)	US	x			
The Orkney Complex Disease Study (ORCADES)	UK	x			
Rotterdam Study-I (RS-1)	Netherlands	x			
Rotterdam Study-II (RS-II)	Netherlands	x			
Rotterdam Study-I (RS-III)	Netherlands	x			
TwinsUK (TUK-1)	UK	x			
TwinsUK (TUK-23)	UK	x			

* part of the replication phase

3.1 DEVELOPMENT OF A GENETIC RISK SCORE

Genetic risk scores for aBMD were created in paper I and II, and for serum T and BMI in paper IV. Creating a genetic risk score combines the risk of all included SNPs. The genetic risk scores developed in paper I, II and IV were based on independent SNPs. As a result, the calculated combined risk adds the risk of each SNP. Depending on whether a weighted or an un-weighted model is used, combining the risk either means calculating the risk simply by counting the number of risk alleles (un-weighted model),

$$Score_i = \sum_{j=1}^{m} x_{ij}$$

, where m = the number of SNPs and x_{ij} = the number of risk alleles of individual i at locus j

18

or by attaching weight to each risk allele based on the effect, relative to the other SNPs, that the risk allele has on the phenotype in question.

$$Score_i = \sum_{j=1}^{m} b_j \, x_{ij}$$

, where m = the number of SNPs, b_j = weighted risk for each risk allele at locus j, x_{ij} = the number of reference alleles of individual i at locus j.

Weights for the weighted GRS on BMD, BMI and serum T were based on each SNP's effect size with BMD in project I, with BMI in the meta-analysis by Locke et al and with serum T in earlier studies of ours (56, 58, 79).

Although using a weighted model incorporates more information on the association between SNPs and phenotype, it also requires that the data set where the effect size of the association between a SNP and the phenotype is estimated and the data set where the genetic risk score is evaluated are disjoint. In other words, one should not evaluate a weighted genetic risk score and estimate the SNPs' effect sizes in the same cohorts as this might lead to biased results. Unfortunately, because of the small effect sizes, large samples are required, which usually means that GWASs include most, if not all, of the cohorts with available phenotype and genotype data. Finding cohorts with relevant data, that were not part of the original study, might therefore become a challenging task. In paper II and IV there is an overlap in cohorts between the original study identifying and estimating the effect sizes and the study where the genetic risk score (-s) is calculated. For aBMD and BMI this overlap is less than 4.5%, while it is considerably larger for T in paper IV(40%). As this could potentially bias the results, we calculated both a weighted and an un-weighted risk score and arrived at similar results.

3.2 META-ANALYSIS

Due to ethical constraints regarding the sharing of individual subject data, summary statistics (sample size, standard error and estimated effect size) from locally performed analyses are often shared instead. In the papers included in this thesis, analyses using pooled data have been performed when possible (due to the ethical constraints mentioned above, we have not shared any individual data from the Gothenburg Osteoporosis and Obesity Determinants [GOOD] and Osteoporotic Fractures in Men [MrOS Sweden]

cohorts). It has been suggested, though, that there is no, or minimal, loss in power for linear regression analyses used in most GWASs when summary statistics is meta-analyzed instead of performing a pooled analysis of all individual data (80). In paper IV, where raw data was available, results were very similar for both pooled and meta-analyzed data.

If all studies had the same variation in their results, combining their results to take advantage of the larger number of subjects, would only amount to computing the mean effect. Unfortunately, this is never the case. Some studies will have more precise effect estimates than others. A meta-analysis combines these results using a weighted mean, where some studies will be given more weight than others. Two commonly used models are the fixed effect and the random effects models. Both models assign weights based on each study's variance. Whereas the fixed effect model assumes that there is one true effect size underlying all studies, the random effects model allow for multiple true effect sizes. As a consequence, in the fixed effect model there is only one source of error in the estimate of the combined effect and that is the random error within studies. Hence, given a sufficiently large sample this error will tend towards zero.

The weight assigned to each study, w_i, in the fixed effect model is

$$w_i = 1/V_i$$

, where V_i is the within-study variance for study (i). The weighted mean T is then computed as:

$$T = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}$$

, which is the sum of the products $w_i T_i$ (effect size multiplied by weight) divided by the sum of the weights. The variance of the combined effect is defined as the reciprocal of the sum of the weights:

$$V = \frac{1}{\sum_{i=1}^{k} w_i}$$

In contrast, in the random effects model, there are two sources of sampling and two sources of error. The first relates to estimating the true effect size within each study (which is similar to the fixed effect model). Given a large enough sample size the sampling error will tend toward zero. The second source of sampling, however, relates to estimating the mean of the true effects. Here, the number of studies, rather than the size of each

study, will decrease the second source of sampling error. Since the variance is a sum of both errors, the random effects model will always have a larger variance, standard error and confidence interval (CI) for the combined effect than the fixed effect model unless between-study variation is zero.

Concretely, the weights assigned to each study in the random effects model are

$$w_i^* = \frac{1}{v_i^*}$$

, where v_i is the within-study variance for study (i) plus the between-studies variance, τ^2.

$$v_i^* = v_i + \tau^2$$

The weighted mean, T^*, is then calculated as

$$T^* = (\sum_{i=1}^{k} w_i^* T_i)/\sum_{i=1}^{k} w_i^*$$

, which is the sum of the products divided by the sum of the weights. The variance of the combined effect is defined as the reciprocal of the sum of the weights.

$$v^* = \frac{1}{\sum_{i=1}^{k} w_i^*}$$

In the papers included in this thesis, the choice of model was based primarily on a test of heterogeneity, where a random effects model was used when there was evidence of high heterogeneity and a fixed effect model otherwise. Although this is common practice, it should be noted that if the number of studies is small and the within-studies variance is large, this test may have low power (81).

3.3 AN ALTERNATIVE METHOD FOR DEVELOPING A GENETIC RISK SCORE

Today, the total number of GWASs performed is quite large, but it is often the case that it is not possible to get access to individual data due to ethical constraints. Thus, the ways of calculating genetic risk scores described above, which requires raw data, is not an option.

Interestingly, another use of the meta-analysis model was recently presented (82). As it turns out, small effect sizes of multiple SNPs that are in linkage equilibrium with each other are effectively identical when estimated jointly by a multiple linear regression model and in a series of single SNP regression models. Thus, it is possible to mimic the regression of a genetic risk score based on a number of SNPs by single regressions on each of the included SNPs, thus removing the need for data on single individuals for the analysis. Each SNP is then treated as a cohort in the meta-analysis setting, with weights for individual effect sizes based on the precision of each effect size (e.g. variance). In paper IV this was used to evaluate a potentially causal effect of serum T on BMI. As it was possible that a lack of a significant causal effect in the cohorts with available data could be the result of a too small sample size, we used the method described above to evaluate a potential causal effect of T on BMI using the very large GIANT consortium. This enabled us to increase the sample size more than tenfold.

3.4 RECLASSIFICATION

A number of metrics have been proposed in order to evaluate a model's discrimination capabilities, but the area under the receiver operating characteristic (ROC) curve (AUC) has traditionally been the most popular metric (88). It is the probability that given two subjects, one who will develop an event and the other who will not, the model will assign a higher probability of an event to the former (89) and is defined as follows: Let X represent the predicted probability of developing an event and D be the event indicator. If f is the probability density function of X, for any cut-off point u, 0<u<1, we can express sensitivity and one minus specificity (negSpec) as

$$Sensitivity\ (u) = S(u) = P(X > u | D = 1) = \int_0^1 f(x|D=1)dx$$

$$negSpec\ (u) = P(u) = P(X > u | D = 0) = \int_0^1 f(x|D=0)dx$$

Then, AUC can be expressed as

$$AUC = \int_0^1 \frac{S(u)\ d}{du} P(u)du = \int_0^1 S(u)f(u|D=0)du = P(X_i > X_j | D_i = 1, D_j = 0).$$

In paper II, we evaluated the clinical usefulness of a genetic risk score for fracture prediction based on the aBMD-associated SNPs identified in paper I. For this task, we used three different metrics: AUC (c-statistics), continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI).

All three metrics have their own weaknesses. For instance, the AUC calculates a model's sensitivity and specificity for all cut-offs for the probability of an outcome, including cut-offs that would never have found their way into a clinical setting. After all, why would anyone care for a model's ability not to cause false alarms (false positives) at points where its ability to correctly identify a true positive is close to zero (Figure 6; lower left corner on the diagram)?
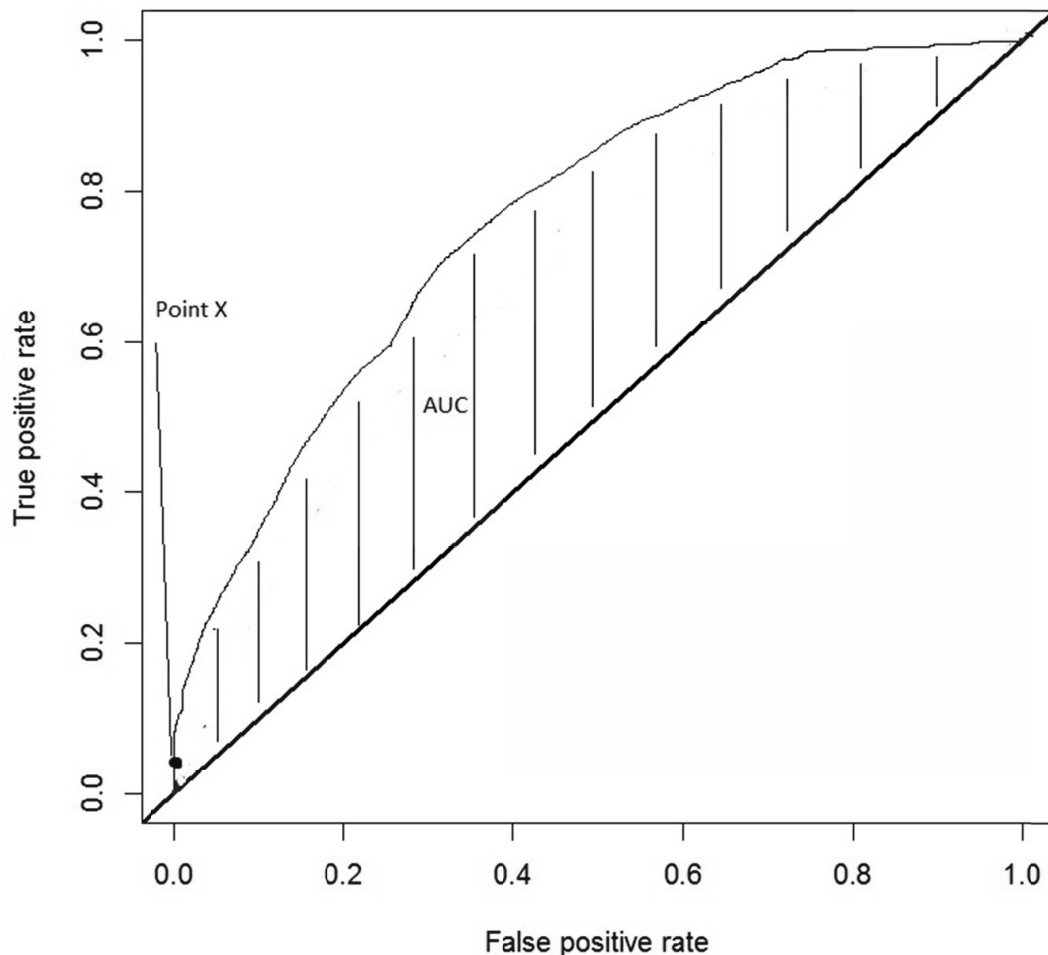

Figure 6. Receiver Operating Characteristic (ROC) curve. Area under curve (AUC) is the area between the ROC curve and the straight diagonal line. Point X close to (0,0) corresponds to a cutoff with close to 0 false and true positives (model will predict almost no events).

Another issue with using the c-statistics to evaluate a new marker is that it has been shown that in order to arrive at a meaningful increase in c-statistics, when starting out with a reasonably good model, a very strong, independent association between the new marker and the outcome is required, thus potentially failing to identify a marker as one that should be incorporated into the prediction model (83-85).

Reclassification quantified in terms of NRI is the sum of differences in proportions of individuals moving up minus the proportion moving down for those with the outcome, and the proportion of individuals moving down minus the proportion moving up for those without the outcome. Formally,

$$NRI = \frac{P(event|up) * P(up) - P(event|down) * P(down)}{P(event)}$$
$$+ \frac{P(nonevent|down) * P(down) - P(nonevent|up) * P(up)}{P(nonevent)}$$

Where $P(event|up) * P(up) = \frac{\#events\ with\ increased\ predicted\ probability}{\#\ events}$

P(event|down) * P (down), P(nonevent|up) * P(up) and P(nonevent|down) * P(down) are all calculated in a similar way.

Originally, NRI was designed to evaluate a marker at a specified cut-off. In Sweden, using the well-established fracture prediction tool, FRAX, a 15% risk of osteoporotic fracture within the next 10 years is the cut-off where a DXA scan (to measure aBMD) is warranted (86). Unfortunately, with our study's prevalence of fractures, this would be inappropriate. Therefore, we chose to use a continuous NRI. As a consequence, it is susceptible to the same criticism as the AUC: there is no guarantee that reclassification occurs at cut-offs that are clinically relevant. Furthermore, NRI says nothing about the magnitude of the change in predicted risk for a subject. Rather, it merely counts the number of subjects with changed risk prediction. It is, however, considerably more sensitive to improvements of the original model. In contrast, IDI, is equivalent to the difference in discrimination slopes of two models, meaning it quantifies the change between the new and old model in the gap in predicted risk between subjects with and without an event. Formally,

If $P_{new,events}$ denotes the mean of the new model-based predicted probabilities of an event for those who develop events and $P_{old,events}$ denotes

the corresponding quantity based on the old model. Let $P_{new,nonevents}$ and $P_{old,nonevents}$ denote the corresponding quantities for nonevents. IDI can then be estimated as

$$IDI = \left(P_{new,events} - P_{new,nonevents}\right) - \left(P_{old,events} - P_{old,nonevents}\right).$$

In order to minimize the drawbacks of each respective metric, we calculated AUC, NRI, and IDI in paper II.

3.5   Z SCORES

The issue of obtaining comparable results involves not only genotyping with different chips giving different SNPs, but also the phenotyping itself. In genetics, imputation was developed to ensure comparable results across study populations. Comparable continuous phenotypes call for a far less complex solution. In this thesis we used standard scores, z-scores, which express the raw score (e.g. serum T levels, weight etc) in terms of the number of standard deviations away from the mean:

Z-score = (Observed − sample mean) / sample standard deviation

Z-scores have been used in the papers herein on all phenotypes, including T, BMI and different types of metrics on BMD. Furthermore, we have used log transformation of non-normally distributed traits.

A concern regarding the use of z-scores relates to the fact that the sample variance can differ quite a lot from the overall population variance in some instances, resulting in a overestimated, or underestimated, difference (e.g. standard deviations from the mean) between subjects. This might pose a problem in some situations where the study population of interest is chosen based on the individuals characteristics in some regard (e.g. personal bests of international pro marathon runners), as the variance within this group might be far smaller than in the general population. It is less obvious, however, how this would be a problem in cohorts where non-performance metrics are used and subjects are chosen randomly.

3.6 IMPUTATION

The HapMap (https://www.genome.gov/10001688/international-hapmap-project/) has been the main reference panel used for the papers in this thesis. The 1000G reference panel (http://www.internationalgenome.org/about#1000G_PROJECT) was used as a secondary analysis to increase the resolution in the areas close to a significant SNP in paper I. The rationale behind this is that the SNPs found in a GWAS are seldom causally linked to the phenotype. Rather, they are correlated with the causal variant. Increasing the resolution, could, then, provide the causal variant, or provide a SNP which is more highly correlated with it.

Imputed, rather than genotyped, information has been used for all SNPs including those which where genotyped originally. This procedure has previously been shown to provide consistent and reliable results for common variants (87). At the time of paper I and paper III, GWASs were performed almost exclusively using HapMap with reference panels for specific ethnic groups such as Caucasians. Although the growth of the genetic research area owes much to imputation in general and HapMap in particular, there are limits. One of the most important aspects is related to the concept of MAF. MAF corresponds to the percentage of risk alleles in a population. A low MAF means that a larger sample is needed for an acceptable power. Historically (including paper I and III), a 5% cutoff has often been used, where SNPs with a lower MAF are excluded from the analyses due to power issues. Although having a lower MAF cutoff means performing regression analyses on more SNPs, and thus lowering the significance threshold, it is not the increased number of tests that presents the greatest challenge (partly because many SNPs are correlated and thus do not give rise to a lower significance threshold because they do not constitute independent tests). Rather, it is the common-disease, common-variant hypothesis, which holds that genetic influences on diseases of high population prevalence are old, and are thus typically very common, together with the issue of having low power to detect a significant association for rare variants (88). Recently, concerns have also been raised regarding the challenges on the analytical side. For instance, the rare variants might be too few for traditional statistical tests and multicollinearity might make it difficult to identify independent associations (89). Moreover, until recently the imputation accuracy for SNPs with allele frequency below 5% has been unsatisfying (90). Hence, the choice to focus on common variants provided a necessary compromise between imputation quality and power on one side and the number of significant findings and variance explained on the other, and reflects the resources available at that

time. With larger data sets and reduced cost of genotyping new possibilities arise. For example, in a more recent co-authored publication, focusing on less common (MAF 1-5%) and rare variants (<1%), we found a SNP in the EN1 gene (MAF = 1.7%) with an effect on lumbar spine BMD four times stronger than the mean of previously reported common variants and two times stronger than that of the largest previously reported effect on lumbar spine BMD (91).

Since paper I and III, focus has shifted towards rare variants and the number of individuals included in nonspecific-population reference panels as well as the number of smaller population-specific reference panels have increased. To date, there has been conflicting evidence regarding whether to use large reference panels, smaller population-specific panels or combined reference panels that uses more than one reference panel for SNPs with MAF<1% (92-94).

3.7   MENDELIAN RANDOMIZATION

In paper IV we used an MR approach to evaluate a potential causal relationship between BMI and serum T. Although MR analyses offer epidemiologists a tool that has the potential to detect relationships beyond mere correlation, they rely on three key assumptions (Figure 7):

1.   The instrumental variable (in our case the genetic risk score) is associated with the exposure.

2.   The instrumental variable is independent of any confounders of the exposure and the outcome.

3.   The association between the instrumental variable and the outcome exists only because of the association between the instrumental variable and the exposure; the instrumental variable is independent of the outcome given the exposure (no pleiotropy).
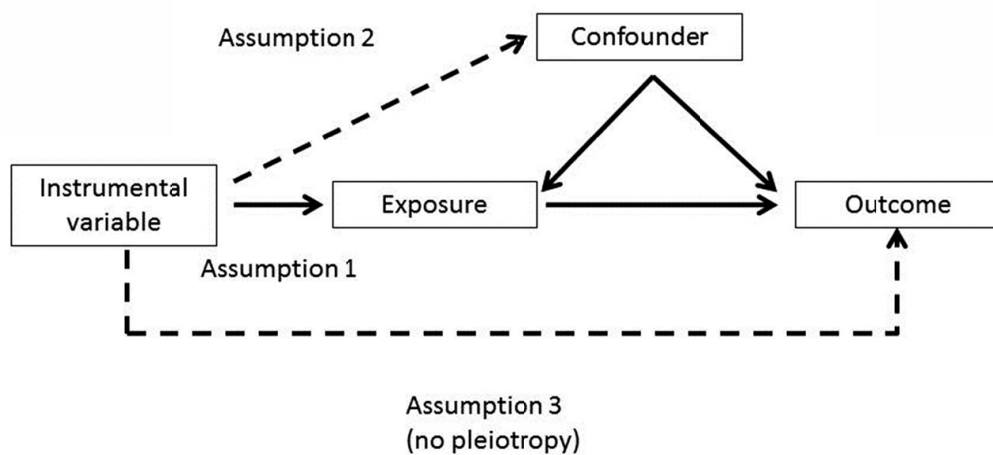
Figure 7. Assumptions of Mendelian Randomization analyses. Assumption 1 assumes that the instrumental variable (IV) is associated with the exposure. Assumption 2 assumes that the IV is independent of the unmeasured confounding factors. Assumption 3 assumes that the IV is independent of the outcome given the exposure and unmeasured confounding factors (no pleiotropy).

Of these, only the first is easily tested. Using BMI-associated and serum T-associated SNPs from the largest GWASs on BMI and serum T, respectively, we developed genetic risk scores for both phenotypes. The association between each risk score and risk factor were then assessed in order to evaluate if assumption one was violated.

Age and smoking are known to be associated with serum T levels. An association between any of these, or other unknown confounders, and the genetic risk score, thus, risks violating assumption two. We therefore tested the association between each genetic risk score and age and smoking. However, as it is not possible to test unknown confounders, it is possible that assumption two is still violated.

To evaluate assumption number three, the assumption of no pleiotropy (i.e. no effect on the outcome independent of the exposure or unmeasured confounding factors), we used two separate, independent, genetic risk scores (one comprised of the well-established BMI-related FTO SNP and the other the remaining 96 SNPs) as a way to detect a pleiotropic effect. Although it doesn't remove the possibility of pleiotropy, it would require similar pleiotropic effects starting out from two independent genetic "sites". With that said, had we performed additional sensitivity analyses such as MR-Egger with the same results, it would have further strengthened our conclusions.

It is well established that an instrumental variable that fulfills the assumptions above, could still be inappropriate because its association with

the risk factor is too weak, i.e. the variance explained of the risk factor is low. The F statistics is a way to measure how well an instrumental variable captures the variation of a risk factor and depends on sample size and the variance explained of the exposure variable by the genetic instrument. A weak instrument (usually interpreted as an F statistics below 11) risks introducing a bias away from the null in a one-sample MR analysis, due to the fact that a weak instrumental variable is biased in the same direction as the observational association (95). For BMI and serum T, this means that a weak instrument would have an increased risk of suggesting a causal relationship where there is none. However, all genetic risk scores used in paper IV had high F statistics that makes weak instrument bias highly unlikely.

The F statistics and the power to detect a significant association are related. While the power is dependent on variance explained of the exposure by the genetic risk score as well as sample size, it also depends on the correlation between exposure and outcome. It is thus possible to have a suitable genetic risk score (which does not suffer from weak instrument bias), but still have low power due to a low correlation between the exposure and outcome. In paper IV, we had very good power assuming the causal effect would be similar to the observed association between risk factor and outcome. However, it is by no means certain that the un-confounded causal effect is of the same magnitude. Thus, for the genetic risk score for T, we used a larger sample from the GIANT consortium to test the association. Had this association been statistically significant, it would have been possible to perform something close to a two-sample MR analysis as there was very little overlap between the cohorts used for the IV-exposure and IV-outcome analyses. Also, due to the very large sample size of the GIANT consortium, this meant effectively reducing the risk of low power as the reason for the lack of a significant causal effect of T on BMI.

Population stratification, often the result of geographical restrictions, refers to the systematic differences in allele frequencies between sub-populations within a population followed by genetic drift in each group. If present, it could affect the results. The subjects included for the analyses in paper IV were all Caucasians, where population stratification based on our sample was considered to be a minor problem. The similar allele frequencies for the SNPs in this study provide further support of this (96). If present, but properly adjusted for, it is unlikely to be a major problem where no family connection is allowed between subjects (88, 97, 98). Thus, in paper I, II and III where population stratification might be an issue, models were further adjusted for population stratification.

We did not test for canalization, which means that the genetic effect would be compensated for by some feedback mechanism. However, since such feedback mechanism would bias the results toward the null, it cannot explain the effect of BMI on serum T.

4 RESULTS

4.1 PAPER I

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

Aim

To identify novel genetic loci associated with DXA-derived two-dimensional aBMD

Using 17 cohorts including 32,961 individuals of European and East Asian ancestry, we performed the largest meta-analysis to date on lumbar spine (LS) and femoral neck (FN) BMD. In addition, we also tested the top BMD-associated markers for replication in 50,933 independent subjects and for association with risk of low-trauma fracture in 31,016 individuals with a history of fracture (cases) and 102,444 controls.

Main results

- 56 loci (32 new) associated with BMD at genome-wide significance ($P < 5 \times 10^{-8}$). Most SNPs were associated with BMD at both FN and LS.

- Fourteen BMD-associated loci were also associated with fracture risk ($P < 5 \times 10^{-4}$, Bonferroni corrected)

- Two of the newly identified loci were discovered in the sex-stratified meta-analysis: 8q13.3 in women and Xp22.31 in men

- In general, the effect of these SNPs on BMD was larger than on fracture risk

Conclusions

We identified 56 loci (64 SNPs), including 32 novel loci, that were independently associated with FN and/or LS BMD.

Discussion

Although the effect size of each BMD-associated SNP is small, this study identified previously known as well as new genes within the pathways Wnt factors, mesenchymal stem cell differentiation and endochondral ossification.

Fracture risk and BMD are correlated, but other bone characteristics not captured by BMD as well as balance and muscular strength etc also influence the risk for fracture. Hence, some important fracture variants may have limited impact on BMD and vice versa. This is true for the SNP at 18p11.21, which was the most significantly associated SNP with fracture (OR = 1.08, 95% CI = 1.06–1.10; $P = 8.8 \times 10^{-13}$), despite a modest effect on BMD. This is in sharp contrast with the majority of variants identified in this study that were found to have strong effects on BMD, but lacked a significant association with fracture. Hence, given the complex nature of fracture risk, future well-powered GWAS meta-analyses should focus on fracture risk as the primary end point.

Interestingly, although not unexpected, our study also found evidence of sex and site specificity with regard to BMD variation. rs5934507 at Xp22.31 was only significant in men. In a previous study of ours, this SNP was found to be associated with serum T in men (56). It is likely that it affects serum T, which, in turn, regulates BMD directly, or perhaps more likely, via E2.

We also found evidence of site specificity, where some markers were only associated with BMD at the femoral neck or the lumbar spine. This is in line with the different bone characteristics at these sites where trabecular bone is the dominant type at the lumbar spine, whereas the femoral neck consists mostly of cortical bone. These findings further highlight the importance of a future GWAS focusing on specific cortical and trabecular bone phenotypes.

4.2 PAPER II

Limited Clinical Utility of a Genetic Risk Score for the Prediction of Fracture Risk in Elderly Subjects

Aim

To determine the clinical usefulness of the BMD-associated SNPs found in paper I for prediction of BMD loss and fracture

Using two male (MrOS US, MrOS Sweden) and one female (Study of Osteoporotic Fractures [SOF]) large prospective cohorts of older subjects, we studied the clinical utility of a genetic risk score based on 63 autosomal BMD-associated SNPs (GRS63) and a genetic risk score based on 16 autosomal fracture-associated SNPs (GRS16) for the prediction of BMD, BMD change and radiographically and/or medically confirmed incident fractures (8,067 subjects, 2,185 incident fractures).

Main results

- GRS63 was associated with BMD, but not with BMD change.

- Similar significant associations with fractures were found for both GRS63 and GRS16. For both GRSs, the associations were substantially attenuated after BMD adjustment.

- Net reclassification improvements with the addition of the GRSs to a base model (age, weight and height) were modest and substantially attenuated in BMD-adjusted fracture prediction models.

- No significant improvements in C-statistics were found when the GRSs were added to a fracture prediction model including age, weight, height and BMD.

Conclusions

GRS63 is associated with BMD, but not BMD change, suggesting that the genetic determinants of BMD differ from those of BMD change. When BMD is known, the clinical utility of the two GRSs for fracture prediction is limited in elderly subjects.

Discussion

Previous studies have shown that both BMD and BMD change are highly heritable traits (32-36).

In the present study, GRS63, which is based on BMD-associated SNPs, was highly significantly associated with BMD, but not with BMD change, suggesting that the genetic architecture underlying BMD change differs from that of peak BMD.

Although osteoporosis is defined in terms of BMD, neither BMD nor osteoporosis is important in their own rights. Rather, their importance is derived from their connection to the risk of fractures. Today's tools for fracture risk prediction combine clinical risk factors with BMD measurements. Although helpful for clinicians, they are still in need of improvement. Therefore, the present study evaluated the GRSs association with and ability to predict fractures and found that both GRS63 and GRS16 are significantly associated with hip, non-vertebral and all fractures. However, adding FN-BMD as a covariate substantially reduced the effect sizes. This was not surprising given the fact that all included SNPs for both GRSs were identified initially using BMD. That both GRSs remain significantly associated with non-vertebral and all fractures might be partly explained by the fact that a single BMD measurement does not capture all of the BMD information during lifetime. It is also possible that some of the SNPs have an effect on other bone parameters such as specific cortical and trabecular bone traits not quantified by the 2-dimensional DXA technique.

In most cases, BMD measurements are readily available in developed countries. Because of this, a potential clinical utility of a GRS must take this into consideration. We found only minor improvements in AUC for fracture prediction (both hip and all fractures) when the GRSs were added to a base model adjusted for age, height and weight. After adjustment for BMD, there was not a significant improvement in AUC for any fracture type in any of the cohorts for GRS63 or GRS16. Although a significant improvement in reclassification metrics could be seen in two of the cohorts for all fractures, the quantitative changes in assigned risk, on average, were small (a 1.1% increase in predicted risk in those with a fracture and a 0.2% decrease in predicted risk in those without a fracture), thus limiting the clinical utility of the GRSs when BMD is known. In countries where BMD is not available, the utility of these genetic risk scores would be higher. Improving the GRSs clinical utility in the future might include incorporating future GWASs aimed at identifying SNPs associated with fracture, lean mass and risk of falls as well as more bone-specific traits.

4.3 PAPER III

Genetic determinants of trabecular and cortical volumetric bone mineral densities and bone microstructure

Aim

To identify novel genetic loci associated with specific cortical and trabecular bone parameters separately using three-dimensional pQCT

Trabecular and cortical vBMD were measured using pQCT. Separate GWA meta-analyses for both traits were then performed using up to 5,878 subjects, including both men and women followed by replication. The identified SNPs were further analyzed in a subset (n=729) consisting of young men with available data on HR-pQCT, where the impact of these SNPs on trabecular bone microstructure and cortical porosity was determined. Finally, in an attempt to assess the underlying functional mechanism of the identified loci, we examined their potential role in regulating gene expression using eQTL in primary human osteoblasts.

Main results

- Four separate loci (RANKL rs1021188, p = 3.6 x 10^{-14}; LOC285735, rs271170, p = 2.7 x 10^{-12}; OPG, rs7839059, p = 1.2 x 10^{-10}; and ESR1/C6orf97, rs6909279, p = 1.1 x 10^{-9}) were genome-wide significantly associated with cortical vBMD (p<= 5 x 10^{-8}).

- One locus, FMN2/GREM2 (rs9287237, p = 1.9 x 10^{-9}), reached genome-wide significance in the analysis on trabecular vBMD.

- rs1021188 (RANKL locus) was significantly associated with cortical porosity.

- rs9287237 (FMN2/GREM2) was significantly associated with trabecular bone volume fraction, number and thickness, as well as fracture risk and prevalent x-ray verified vertebral fractures in the MrOS Sweden cohort and with GREM2 expression in human osteoblasts.

- There was a low correlation between femoral neck aBMD and cortical vBMD (correlation= 0.04) and a modest correlation between femoral neck aBMD and trabecular vBMD (correlation=0.65).

Discussion

A number of GWA meta-analyses have been performed identifying a large number of SNPs associated with aBMD (58, 99-103). Although aBMD decreases with age, it is well established that age is also a major predictor of fracture independently of aBMD (27). The increased risk of fractures later in life is believed to be due to a deterioration of bone quality that is not detectable by DXA. This age-associated deterioration is associated with trabecular perforation, thinning, and loss of connectivity, as well as increased cortical porosity (27, 104).

Thus, the objective of the present study was to identify genetic determinants of vBMDs and bone microstructure parameters separately for the cortical and trabecular bone compartments. Due to the fact that only a few of the subjects included in this study had data available from a HR-pQCT analysis, we started by performing the GWAS on cortical and trabecular vBMD as analysed by standard pQCT and then evaluated the associations for identified genetic signals on bone microstructure using the HR-pQCT with higher spatial resolution.

Analyses using bone microstructure have previously enabled us to identify a missense variant in the WNT16 gene to be associated with cortical bone thickness (105). Translational research later demonstrated that mice lacking the WNT16 gene had reduced cortical thickness (18).

Of the five identified significant vBMD loci, one of the four cortical vBMD loci (LOC285735) and the trabecular vBMD locus (FMN2) are novel bone-related loci. The remaining three significant cortical vBMD loci have previously been shown to be associated with aBMD in paper I and elsewhere (58, 100).

The low to modest correlation between the vBMD phenotypes and aBMD means that there is information on bone-specific traits that cannot be quantified by DXA measurements. This might also help explain why this study was able to identify novel bone-associated genetic loci despite the low number of subjects included compared to that of paper I.

Some recent studies have confirmed the effect of RANKL on cortical vBMD, whereas others have also shown an effect on the trabecular traits (106-109).

The significant SNP rs9287237 resides in the FMN2 locus. FMN2 has not previously been described to be associated with skeletal phenotypes. However, rs9287237 is located only slightly downstream of GREM2, which is an extracellular antagonist of bone morphogenetic proteins (BMPs) and it inhibits osteoblastic differentiation (110, 111). Interestingly, the eQTL analyses in human osteoblasts demonstrated that this SNP was significantly associated with the expression of the nearby GREM2 gene. In fact, the allele that was associated with increased trabecular vBMD, increased trabecular bone volume fraction and reduced risk of vertebral fractures was also associated with a decreased expression of GREM2 in human osteoblasts. This suggests that GREM2 might be the gene affecting the trabecular vBMD.

4.4   PAPER IV

Causal relationship between obesity and serum testosterone status in men: A bidirectional Mendelian randomization analysis

Aim

To determine if high BMI causes low T, or if low T causes high BMI in men using a bi-directional MR approach

Serum levels of T were measured in 7,446 men from Denmark, Germany and Sweden. The study subjects also had genotype data on 97 BMI-associated SNPs and 3 T-associated SNPs readily available. Based on the results from our previous GWAS we developed both weighted and un-weighted genetic risk scores for both BMI and T. Using a bi-directional MR analysis we then examined the direction and causality of the relationship between BMI and T. Sub-analyses included testing the association between the weighted genetic risk scores for both BMI and T with SHBG and dividing the GRS for T into SNPs residing within and outside the SHBG locus.

Main results

- Both the weighted ($_w$GRS$_{BMI}$, p = 2.0 x 10^{-3}) and the un-weighted ($_{uw}$GRS$_{BMI}$, p = 1.7 x 10^{-3}) genetic risk score for BMI were significantly and inversely associated with serum T in the meta-analyzed combined cohort. A pooled analysis showed similar results.

- For a body weight reduction, where BMI declines from 30 (cut off for obesity) to 25 (cut off for overweight) kg/m^2, the effect would equal roughly a 13% to 15% increase in serum T.

- Neither the $_w$GRS$_T$ nor the $_{uw}$GRS$_T$ (or the GRSs developed for T SNPs within and outside the SHBG locus) were associated with BMI in the included cohorts. Furthermore, the autosomal SNPs were not individually, or combined, associated with BMI using the GIANT consortium of up to 104,349 men.

Discussion

The prevalence of having a low total serum T (defined as <300 ng/dl) based on one sample ranges from 24 - 77% (112-117). Adding further requirements such as an additional sample, that the samples are taken in the

morning and that the patient has clinical symptoms lowers the prevalence to somewhere around 6% (112, 113).

Although Sweden is still trailing the US, the number of patients receiving TRT in Västra Götalandsregionen, a part of Sweden, has increased steadily over the last few years, which coincides at least partly with the timing of aggressive advertisement campaigns depicting TRT as somewhat of a miracle pill for (ageing) men (118, 119).

Observational studies demonstrate that obesity is associated with low serum T (54), but the direction and causality of this relationship is unclear. Although most randomized, placebo-controlled trials have indicated that T treatment increases lean mass and reduces fat mass in men with low serum T (120-125), the effect of different T levels on BMI and body weight is inconsistent. Reverse causation has been proposed as one possible explanation (126-128).

In light of the obesity epidemic in the western world, the inverse association between T and obesity-related diseases (70, 129, 130) has initiated a discussion as to whether T supplementation could be used as a means to reduce the risk of developing obesity-associated cardiometabolic diseases in men with low serum T. Reverse causation as well as safety concerns regarding increases in cardiovascular risk still need to be addressed, however (131, 132).

In the present study, we found evidence of a causal effect of BMI on serum T. Each SD genetically instrumented increase in BMI was associated with a 0.25 SD decrease in serum T, which is similar to the effect of the observational association. For someone reducing their BMI from 30 to 25 kg/m^2, this equals a 13 - 15% increase in serum T. The finding is also in line with a recent meta-analysis by Corona et al that revealed that body weight reductions as a result of both low-calorie diet and bariatric surgery are associated with significantly increased serum T (133).

The identified causal effect of BMI on serum T was rather similar regardless of if the calculations were performed by pooling the samples or meta-analyzing the results from each cohort, or whether weighted, or un-weighted, genetic risk scores were used.

In contrast, we found no evidence of a causal effect of serum T on BMI, regardless of whether a weighted or un-weighted model was used. Furthermore, since SHBG is known to be associated with serum T levels, we also performed a sub-analysis where the SNPs constituting the GRS for T where divided based on whether they were within or outside the SHBG locus.

Despite the fact that the study was well-powered according to the power analysis, a key assumption underlying this result was that the causal effect of T on BMI was close to the observed association. Since it is possible that the causal effect is smaller, implicating that the study might still be underpowered, we also used the GIANT consortium to test the association between T-related autosomal SNPs and BMI in a much larger sample including as many as 104,349 men. Despite the considerably larger sample size, we failed to identify a causal role of T on BMI.
One of the greatest challenges when applying an MR analysis to answer questions of causality in biological systems is the question of pleiotropy, where one gene affects the outcome independent of the exposure of interest. By using as many as 97 recently reported independent SNPs to index BMI, we were able to minimize the risk of shared pleiotropy and linkage-disequilibrium-induced confounding pleiotropic effects (134, 135). Also, the use of two separate independent genetic instruments with similar point estimates of the causal effect of BMI on serum T further reduced the risk of pleiotropy (134, 135). Despite the efforts to minimize the risk of pleiotropy described above, it is a limitation of this study that none of the more recently developed sensitivity analyses such as MR-Egger were performed.

5 SUMMARY

In the largest GWAS on DXA-derived BMD to that date, we identified 56 genetic loci, of which 32 where novel, associated with two-dimensional aBMD at the femoral neck and/or lumbar spine.

Based on these findings, we developed two genetic risk scores. One based on those SNPs that were significantly associated with BMD (GRS63) and one that was based on those SNPs that were significantly associated with fractures (GRS16). Both GRSs were associated with fractures, but the estimated effect sizes were substantially reduced after adjustments for BMD. The clinical utility, as assessed by AUC and reclassification, was limited when BMD is known.

As the 2-dimensional DXA technique cannot differentiate between cortical and trabecular bone, we performed GWASs using the more specific 3-dimensional pQCT. Using this technique and HR-pQCT, we managed to identify one novel locus significantly associated with cortical vBMD and one significantly associated with trabecular vBMD, trabecular number and thickness as well as GREM2 expression in human osteoblasts and fracture risk.

Low serum T levels are associated with an increased risk of osteoporosis and fractures. but the determinants of serum T are to a large extent unknown. Observational studies have demonstrated that obesity is strongly associated with low serum T, but the direction and causality of this relationship is unclear. As a second objective of this thesis, we therefore applied an MR approach and found evidence of a causal effect of BMI on serum T, where 1 SD lower BMI increased serum T by 13 - 15%. No evidence was found supporting a causal effect of serum T on BMI.

6 GENERAL DISCUSSION AND FUTURE PERSPECTIVES

When we performed the largest GWAS on FN and LS BMD at that time, we identified 32 novel loci. These findings could be of use as potential new pharmaceutical targets and/or for improving prediction models to identify patients at risk of osteoporotic fractures. Admittedly, in our search for new pharmaceutical targets, a GWAS is merely the beginning, but a potentially important one. In fact, evidence suggests that drug targets implicated by GWAS are twice as likely to succeed in clinical trials (136), which, given the costs of clinical trials, is of substantial worth. Interestingly, the WNT16 identified in paper I and elsewhere might well turn out to be such an example. The work of our research group and others (18, 137, 138) has revealed at least part of the mechanisms by which it exerts its effect on BMD. Also, the FMN2/GREM2 gene identified in paper III, with an effect on both trabecular bone and fracture, also constitutes an interesting object for further mechanistic studies.

In contrast to the two promising examples discussed above from paper I and III, the total genetic variance explained (5-6%) of BMD in paper I, and elsewhere, has been low despite large sample sizes. This is not solely an issue for BMD, but seems to be the case for most complex traits in human populations (139). The debate on the missing heritability has fueled the discussion regarding the success or failure of GWAS. Interestingly, Yang et al demonstrated that, for height at least, most of the heritability is not missing. Rather, it has still to be detected because of the small individual effects that are too small to pass stringent significance tests. Moreover, the remaining heritability is due to incomplete linkage disequilibrium between causal variants and genotyped SNPs, exacerbated by causal variants having lower MAF than the SNPs explored to date (139). Thus, still larger sample sizes that allows for identification of SNPs with smaller effects and SNPs with lower allele frequencies, together with population-specific imputation panels is likely to increase the variance explained. Also, several theoretical studies have suggested that by selecting subjects from the extremes of the population distribution, power can be increased considerably (140-142) given the same sample size. Hence, the fact that the UK Biobank (http://www.ukbiobank.ac.uk/) was made available to researchers world-wide with its large sample size and well-characterized phenotypes and state-of-the-art whole genome sequencing, it is not unlikely that the variance explained will increase dramatically for a number of different phenotypes. In fact, a

recent analysis on heel BMD using ultrasound in the UK Biobank, revealed 153 new loci significantly associated with BMD at this site. This explained approximately 20% of the total genetic variance (143).

In paper II, we aimed at identifying the clinical utility of the findings in paper I. Although associated with BMD and fractures, the GRSs did not improve AUC significantly when BMD was part of the base model. This is in line with the results of Lee et al (144). Although both Lee and the present study found significant improvements of NRI, it is questionable whether this translates into a meaningful increase in clinical utility.

A future GRS for fracture prediction might benefit from the addition of markers identified in other fracture-associated traits such as risk of falls and our recent co-authored GWAS on lean mass (145). Using the results from larger GWASs on more detailed specific bone-related phenotypes such as trabecular and cortical vBMD will most likely also yield important contributions. The weight assigned to each marker could then be assigned in a similar fashion as in paper II using each SNP's effect on fracture risk. These extended GRSs could then potentially improve fracture risk prediction when combined with the clinical risk factors, enabling personalized fracture risk assessment.

There are several alternatives to the prediction model used in paper II. One of the more obvious is to use a less stringent p-value threshold for the inclusion of SNPs (i.e. resulting in more true positives and false positives in the model). This approached has been adopted by many studies of late (146-152). Results indicate that this could improve the prediction capability of the model (153-156). The optimal threshold, however, depends on a number of factors, including the ratio of true positives to false positives, sample size (possibly due to the fact that true positives become more enriched in the SNP sets with lower p-value thresholds in larger sample sizes) and the metrics used for the evaluation of risk prediction (151, 154).

The findings in paper III explained 1.5% of total genetic variance of cortical vBMD and 0.7% of trabecular vBMD in the replication cohort. If the ratio between genetic variance and total variance is taken to be roughly the same for these traits as for aBMD, this equals a genetic variance of about 1.8-2% and 0.9-1.2%, respectively. Hence, despite having a sample size of less than one tenth of that used in paper I, in paper III we managed to identify new loci with an explained genetic variance of 1/3 of that explained by both new and already known loci in paper I. Part of the explanation might lie in the fact that the cohorts used in paper III were based almost exclusively on Caucasian men, whereas the study population used in paper I was highly

genetically heterogenous. However, the most plausible explanation is that the complex nature of aBMD with a mixture of different variables including cortical and trabecular vBMD and bone dimensions introduces more noise than do more specific individual phenotypes such as cortical, or trabecular vBMD. Regardless, rather than just increasing the sample size for the readily available aBMD by more genetically heterogeneous cohorts, it would be well advised to consider moving forward with more detailed studies using pQCT or HR-pQCT in homogeneous populations.

The findings in paper IV are important due to the increasing trend of T prescription and safety concerns related to T treatment. While we found no evidence of a causal effect of serum T on BMI, it should be remembered that BMI is a metric of weight and does reflect neither the distribution of body fat nor the distribution between muscle and fat. Hence, it cannot be ruled out that an effect on fat loss is at least partly compensated for by a gain in lean mass. Future studies on more specific phenotypes with known connections to CVD are warranted.

7 CONCLUSION

Osteoporosis and its related fractures are a global public health concern, accounting for huge costs to society, with costs expecting to increase further with an aging population. By using the largest sample size to date then, we managed to identify 32 novel genetic loci associated with aBMD at the femoral neck and/or lumbar spine. By developing genetic risk scores based on these findings we showed, unfortunately, that they had limited clinical utility for fracture prediction, when aBMD is known. The utility of these findings should not only be discussed in terms of fracture prediction, however, as evident by the WNT16 gene. The WNT16 gene identified in paper I and elsewhere have recently been the subject of intense research as our research group and others have identified mechanisms and effects that make it interesting to evaluate its merits as a potential new pharmaceutical target.

Bone at different skeletal sites consists of a mixture of trabecular and cortical bone. At some sites trabecular bone is the predominantly bone type, whereas cortical bone is more abundant at other sites. Although aBMD is the golden standard for diagnosing osteoporosis in a clinical setting, it cannot differentiate between cortical and trabecular vBMD. We therefore performed a successful GWAS on cortical and trabecular vBMD that, despite its modest sample size, identified two novel loci associated with cortical and trabecular vBMD, respectively.

Finally, in paper IV we applied an MR approach and found evidence of a causal effect of BMI on serum T, but not the other way around. This supports the idea that the increasing number of obese non-hypogonadal men with low serum T should be offered lifestyle interventions rather than T treatment as a first intervention. In fact, based on the trends of reduced serum T and increased BMI in men, together with these results, it is quite possible that successful population level interventions reverting the obesity epidemic might also lead to a reversal of the secular trend of reduction in serum T.

RELATED PUBLICATIONS NOT INCLUDED IN THESIS

1. Kemp JP, Sayers A, Paternoster L,..., **Eriksson J**, et al. Does bone resorption stimulate periosteal expansion? A cross-sectional analysis of beta-C-telopeptides of type I collagen (CTX), genetic markers of the RANKL pathway, and periosteal circumference as measured by pQCT. J Bone Miner Res. 2014;29(4):1015-24

2. Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JR, Koster A, Petersen AK, **Eriksson J**, et al. Genetic determinants of serum testosterone concentrations in men. PLoS Genet. 2011;7(10):e1002313.

3. Ried JS, Jeff MJ, Chu AY, Bragg-Gresham JL, van Dongen J, Huffman JE, Ahluwalia TS, Cadby G, Eklund N, **Eriksson J** et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun. 2016;7:13357.

4. Winkler TW, Justice AE, Graff M,..., **Eriksson J**, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. 2015;11(10):e1005378.

5. Zheng HF, Forgetta V, Hsu YH,..., **Eriksson J**, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature. 2015;526(7571):112-7.

6. Zheng HF, Tobias JH, Duncan E, Evans DM, **Eriksson J**, et al. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. PLoS Genet. 2012;8(7):e1002745.

7. Zillikens MC, Demissie S, Hsu YH,..., **Eriksson J**, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. Nat Commun. 2017;8(1):80.

ACKNOWLEDGEMENT

REFERENCES

1.	Kanis JA, Johnell O, Oden A, Sembo I, Redlund-Johnell I, Dawson A, et al. Long-term risk of osteoporotic fracture in Malmo. Osteoporos Int. 2000;11(8):669-74.

2.	Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. J Bone Miner Res. 2007;22(3):465-75.

3.	Baron R, Hesse E. Update on bone anabolics in osteoporosis treatment: rationale, current status, and perspectives. J Clin Endocrinol Metab. 2012;97(2):311-25.

4.	Haring R, Volzke H, Steveling A, Krebs A, Felix SB, Schofl C, et al. Low serum testosterone levels are associated with increased risk of mortality in a population-based cohort of men aged 20-79. Eur Heart J. 2010;31(12):1494-501.

5.	Laughlin GA, Barrett-Connor E, Bergstrom J. Low serum testosterone and mortality in older men. J Clin Endocrinol Metab. 2008;93(1):68-75.

6.	Tivesten A, Vandenput L, Labrie F, Karlsson MK, Ljunggren O, Mellstrom D, et al. Low serum testosterone and estradiol predict mortality in elderly men. J Clin Endocrinol Metab. 2009;94(7):2482-8.

7.	Mellstrom D, Vandenput L, Mallmin H, Holmberg AH, Lorentzon M, Oden A, et al. Older men with low serum estradiol and high serum SHBG have an increased risk of fractures. J Bone Miner Res. 2008;23(10):1552-60.

8.	Vandenput L, Ohlsson C. Estrogens as regulators of bone health in men. Nat Rev Endocrinol. 2009;5(8):437-43.

9.	Vanderschueren D, Vandenput L, Boonen S, Lindberg MK, Bouillon R, Ohlsson C. Androgens and bone. Endocr Rev. 2004;25(3):389-425.

10.	Sinnesael M, Boonen S, Claessens F, Gielen E, Vanderschueren D. Testosterone and the male skeleton: a dual mode of action. J Osteoporos. 2011;2011:240328.

11.	Liu PY, Death AK, Handelsman DJ. Androgens and cardiovascular disease. Endocr Rev. 2003;24(3):313-40.

12.	Spitzer M, Huang G, Basaria S, Travison TG, Bhasin S. Risks and benefits of testosterone therapy in older men. Nat Rev Endocrinol. 2013;9(7):414-24.

13.	Datta HK, Ng WF, Walker JA, Tuck SP, Varanasi SS. The cell biology of bone metabolism. J Clin Pathol. 2008;61(5):577-87.

14.	Burghardt AJ, Kazakia GJ, Ramachandran S, Link TM, Majumdar S. Age- and gender-related differences in the geometric properties and

biomechanical significance of intracortical porosity in the distal radius and tibia. J Bone Miner Res. 2010;25(5):983-93.

15.   Barak MM, Lieberman DE, Hublin JJ. A Wolff in sheep's clothing: trabecular bone adaptation in response to changes in joint loading orientation. Bone. 2011;49(6):1141-51.

16.   Raisz LG. Physiology and pathophysiology of bone remodeling. Clin Chem. 1999;45(8 Pt 2):1353-8.

17.   Khosla S, Melton LJ, 3rd, Riggs BL. The unitary model for estrogen deficiency and the pathogenesis of osteoporosis: is a revision needed? J Bone Miner Res. 2011;26(3):441-51.

18.   Moverare-Skrtic S, Henning P, Liu X, Nagano K, Saito H, Borjesson AE, et al. Osteoblast-derived WNT16 represses osteoclastogenesis and prevents cortical bone fragility fractures. Nat Med. 2014;20(11):1279-88.

19.   Brandi ML. Microarchitecture, the key to bone quality. Rheumatology (Oxford). 2009;48 Suppl 4:iv3-8.

20.   Laib A, Hauselmann HJ, Ruegsegger P. In vivo high resolution 3D-QCT of the human forearm. Technol Health Care. 1998;6(5-6):329-37.

21.   Majumdar S, Genant HK. Assessment of trabecular structure using high resolution magnetic resonance imaging. Stud Health Technol Inform. 1997;40:81-96.

22.   Nishiyama KK, Macdonald HM, Buie HR, Hanley DA, Boyd SK. Postmenopausal women with osteopenia have higher cortical porosity and thinner cortices at the distal radius and tibia than women with normal aBMD: an in vivo HR-pQCT study. J Bone Miner Res. 2010;25(4):882-90.

23.   Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. World Health Organ Tech Rep Ser. 1994;843:1-129.

24.   Kanis JA, Oden A, Johansson H, Borgstrom F, Strom O, McCloskey E. FRAX and its applications to clinical practice. Bone. 2009;44(5):734-43.

25.   Sernbo I, Johnell O. Consequences of a hip fracture: a prospective study over 1 year. Osteoporos Int. 1993;3(3):148-53.

26.   Wiktorowicz ME, Goeree R, Papaioannou A, Adachi JD, Papadimitropoulos E. Economic implications of hip fracture: health service use, institutional care and cost in Canada. Osteoporos Int. 2001;12(4):271-8.

27.   Nicks KM, Amin S, Atkinson EJ, Riggs BL, Melton LJ, 3rd, Khosla S. Relationship of age to bone microstructure independent of areal bone mineral density. J Bone Miner Res. 2012;27(3):637-44.

28.   Cody DD, Gross GJ, Hou FJ, Spencer HJ, Goldstein SA, Fyhrie DP. Femoral strength is better predicted by finite element models than QCT and DXA. J Biomech. 1999;32(10):1013-20.

29.   Muller ME, Webber CE, Bouxsein ML. Predicting the failure load of the distal radius. Osteoporos Int. 2003;14(4):345-52.

30.	Wainwright SA, Marshall LM, Ensrud KE, Cauley JA, Black DM, Hillier TA, et al. Hip fracture in women without osteoporosis. J Clin Endocrinol Metab. 2005;90(5):2787-93.

31.	Kemp JP, Sayers A, Paternoster L, Evans DM, Deere K, St Pourcain B, et al. Does bone resorption stimulate periosteal expansion? A cross-sectional analysis of beta-C-telopeptides of type I collagen (CTX), genetic markers of the RANKL pathway, and periosteal circumference as measured by pQCT. J Bone Miner Res. 2014;29(4):1015-24.

32.	Peacock M, Turner CH, Econs MJ, Foroud T. Genetics of osteoporosis. Endocr Rev. 2002;23(3):303-26.

33.	Ralston SH, Uitterlinden AG. Genetics of osteoporosis. Endocr Rev. 2010;31(5):629-62.

34.	Makovey J, Nguyen TV, Naganathan V, Wark JD, Sambrook PN. Genetic effects on bone loss in peri- and postmenopausal women: a longitudinal twin study. J Bone Miner Res. 2007;22(11):1773-80.

35.	Shaffer JR, Kammerer CM, Bruder JM, Cole SA, Dyer TD, Almasy L, et al. Genetic influences on bone loss in the San Antonio Family Osteoporosis study. Osteoporos Int. 2008;19(12):1759-67.

36.	Zhai G, Andrew T, Kato BS, Blake GM, Spector TD. Genetic and environmental determinants on bone loss in postmenopausal Caucasian women: a 14-year longitudinal twin study. Osteoporos Int. 2009;20(6):949-53.

37.	Andrew T, Antioniades L, Scurrah KJ, Macgregor AJ, Spector TD. Risk of wrist fracture in women is heritable and is influenced by genes that are largely independent of those influencing BMD. J Bone Miner Res. 2005;20(1):67-74.

38.	Michaelsson K, Melhus H, Ferm H, Ahlbom A, Pedersen NL. Genetic liability to fractures in the elderly. Arch Intern Med. 2005;165(16):1825-30.

39.	Wagner H, Melhus H, Pedersen NL, Michaelsson K. Heritable and environmental factors in the causation of clinical vertebral fractures. Calcif Tissue Int. 2012;90(6):458-64.

40.	Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int. 2008;19(4):385-97.

41.	Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int. 2007;18(8):1033-46.

42.	Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. Osteoporos Int. 2008;19(10):1431-44.

43.   Luetjens CM WG. "Chapter 2: Testosterone: Biosynthesis, transport, metabolism and (non-genomic) actions". In: Nieschlag E, Behre, Hermann M, editor. Testosterone: Action, Deficiency, Substitution (4th ed) Cambridge: Cambridge University Press. pp. 15–32; 2012. p. pp. 15-32.

44.   Torjesen PA, Sandnes L. Serum testosterone in women as measured by an automated immunoassay and a RIA. Clin Chem. 2004;50(3):678; author reply -9.

45.   Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. Endocr Rev. 2005;26(6):833-76.

46.   Moverare S, Venken K, Eriksson AL, Andersson N, Skrtic S, Wergedal J, et al. Differential effects on bone of estrogen receptor alpha and androgen receptor activation in orchidectomized adult male mice. Proc Natl Acad Sci U S A. 2003;100(23):13573-8.

47.   Lindberg MK, Moverare S, Skrtic S, Alatalo S, Halleen J, Mohan S, et al. Two different pathways for the maintenance of trabecular bone in adult male mice. J Bone Miner Res. 2002;17(4):555-62.

48.   Ohlsson C, Borjesson AE, Vandenput L. Sex steroids and bone health in men. Bonekey Rep. 2012;1:2.

49.   Vandenput L, Mellstrom D, Laughlin GA, Cawthon PM, Cauley JA, Hoffman AR, et al. Low Testosterone, but Not Estradiol, Is Associated With Incident Falls in Older Men: The International MrOS Study. J Bone Miner Res. 2017;32(6):1174-81.

50.   Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536-59.

51.   Khera M, Broderick GA, Carson CC, 3rd, Dobs AS, Faraday MM, Goldstein I, et al. Adult-Onset Hypogonadism. Mayo Clin Proc. 2016;91(7):908-26.

52.   Huhtaniemi I. Late-onset hypogonadism: current concepts and controversies of pathogenesis, diagnosis and treatment. Asian J Androl. 2014;16(2):192-202.

53.   Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, Finn JD, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. N Engl J Med. 2010;363(2):123-35.

54.   Wu FC, Tajar A, Pye SR, Silman AJ, Finn JD, O'Neill TW, et al. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. J Clin Endocrinol Metab. 2008;93(7):2737-45.

55.   Bogaert V, Taes Y, Konings P, Van Steen K, De Bacquer D, Goemaere S, et al. Heritability of blood concentrations of sex-steroids in

relation to body composition in young adult male siblings. Clin Endocrinol (Oxf). 2008;69(1):129-35.

56.   Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JR, Koster A, et al. Genetic determinants of serum testosterone concentrations in men. PLoS Genet. 2011;7(10):e1002313.

57.   Eriksson AL, Perry JRB, Coviello AD, Delgado GE, Ferrucci L, Hoffman AR, et al. Genetic Determinants of Circulating Estrogen Levels and Evidence of a Causal Effect of Estradiol on Bone Density in Men. J Clin Endocrinol Metab. 2018;103(3):991-1004.

58.   Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet. 2012;44(5):491-501.

59.   Huhtaniemi IT, Tajar A, Lee DM, O'Neill TW, Finn JD, Bartfai G, et al. Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men. Eur J Endocrinol. 2012;166(6):983-91.

60.   Lee JS, Ettinger B, Stanczyk FZ, Vittinghoff E, Hanes V, Cauley JA, et al. Comparison of methods to measure low serum estradiol levels in postmenopausal women. J Clin Endocrinol Metab. 2006;91(10):3791-7.

61.   Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME. Challenges to the measurement of estradiol: an endocrine society position statement. J Clin Endocrinol Metab. 2013;98(4):1376-87.

62.   Consultation RoaW. Obesity: preventing and managing the global epidemic. WHO Technical Report Series 894.6-7.

63.   Ried JS, Jeff MJ, Chu AY, Bragg-Gresham JL, van Dongen J, Huffman JE, et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun. 2016;7:13357.

64.   Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. 2015;11(10):e1005378.

65.   Winner SJ, Morgan CA, Evans JG. Perimenopausal risk of falling and incidence of distal forearm fracture. BMJ. 1989;298(6686):1486-8.

66.   Glass AR, Swerdloff RS, Bray GA, Dahms WT, Atkinson RL. Low serum testosterone and sex-hormone-binding-globulin in massively obese men. J Clin Endocrinol Metab. 1977;45(6):1211-9.

67.   Allan CA, McLachlan RI. Androgens and obesity. Curr Opin Endocrinol Diabetes Obes. 2010;17(3):224-32.

68.   Zumoff B, Strain GW, Miller LK, Rosner W, Senie R, Seres DS, et al. Plasma free and non-sex-hormone-binding-globulin-bound testosterone

are decreased in obese men in proportion to their degree of obesity. J Clin Endocrinol Metab. 1990;71(4):929-31.

69.   Couillard C, Gagnon J, Bergeron J, Leon AS, Rao DC, Skinner JS, et al. Contribution of body fatness and adipose tissue distribution to the age variation in plasma steroid hormone concentrations in men: the HERITAGE Family Study. J Clin Endocrinol Metab. 2000;85(3):1026-31.

70.   Ohlsson C, Barrett-Connor E, Bhasin S, Orwoll E, Labrie F, Karlsson MK, et al. High serum testosterone is associated with reduced risk of cardiovascular events in elderly men. The MrOS (Osteoporotic Fractures in Men) study in Sweden. J Am Coll Cardiol. 2011;58(16):1674-81.

71.   Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953;18:123-31.

72.   McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.

73.   Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.

74.   Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89-98.

75.   Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42(4):1134-44.

76.   Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep. 2017;4(4):330-45.

77.   Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.

78.   van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. Int J Epidemiol. 2017.

79.   Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.

80.   Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. Genet Epidemiol. 2010;34(1):60-6.

81.   Michael Borenstein LVH, J. P. T. Higgins and H. R. Rothstein. Introduction to Meta-Analysis: John Wiley & Sons, Ltd; 2009.

82.   International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9.

83.	Greenland P, O'Malley PG. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. Arch Intern Med. 2005;165(21):2454-6.

84.	Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882-90.

85.	Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med. 2006;355(25):2615-7.

86.	(Socialstyrelsen) TNBoHaW. Vid en frakturrisk enligt FRAX på 15 % eller högre i kombination med ett bentäthetsvärde (T-värde) på -2 eller sämre.

87.	Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. Hum Genet. 2009;125(2):163-71.

88.	Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet. 2005;366(9492):1223-34.

89.	Schwantes-An TH, Sung H, Sabourin JA, Justice CM, Sorant AJM, Wilson AF. Type I error rates of rare single nucleotide variants are inflated in tests of association with non-normally distributed traits using simple linear regression methods. BMC Proc. 2016;10(Suppl 7):385-8.

90.	Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

91.	Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature. 2015;526(7571):112-7.

92.	Chou WC, Zheng HF, Cheng CH, Yan H, Wang L, Han F, et al. A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. Sci Rep. 2016;6:39313.

93.	Mitt M, Kals M, Parn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. Eur J Hum Genet. 2017;25(7):869-76.

94.	Zhou W, Fritsche LG, Das S, Zhang H, Nielsen JB, Holmen OL, et al. Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. Genet Epidemiol. 2017;41(8):744-55.

95.	Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.

96.   Larsson SC, Burgess S, Michaelsson K. Association of Genetic Variants Related to Serum Calcium Levels With Coronary Artery Disease and Myocardial Infarction. JAMA. 2017;318(4):371-80.

97.   Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst. 2000;92(14):1151-8.

98.   Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev. 2002;11(6):513-20.

99.   Duncan EL, Danoy P, Kemp JP, Leo PJ, McCloskey E, Nicholson GC, et al. Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. PLoS Genet. 2011;7(4):e1001372.

100.   Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet. 2008;371(9623):1505-12.

101.   Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat Genet. 2009;41(11):1199-206.

102.   Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, et al. Multiple genetic loci for bone mineral density and fractures. N Engl J Med. 2008;358(22):2355-65.

103.   Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, et al. New sequence variants associated with bone mineral density. Nat Genet. 2009;41(1):15-7.

104.   Zebaze RM, Ghasem-Zadeh A, Bohte A, Iuliano-Burns S, Mirams M, Price RI, et al. Intracortical remodelling and porosity in the distal radius and post-mortem femurs of women: a cross-sectional study. Lancet. 2010;375(9727):1729-36.

105.   Zheng HF, Tobias JH, Duncan E, Evans DM, Eriksson J, Paternoster L, et al. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. PLoS Genet. 2012;8(7):e1002745.

106.   Bonani M, Meyer U, Frey D, Graf N, Bischoff-Ferrari HA, Wuthrich RP. Effect of Denosumab on Peripheral Compartmental Bone Density, Microarchitecture and Estimated Bone Strength in De Novo Kidney Transplant Recipients. Kidney Blood Press Res. 2016;41(5):614-22.

107.   Kostenuik PJ, Nguyen HQ, McCabe J, Warmington KS, Kurahara C, Sun N, et al. Denosumab, a fully human monoclonal antibody to RANKL,

inhibits bone resorption and increases BMD in knock-in mice that express chimeric (murine/human) RANKL. J Bone Miner Res. 2009;24(2):182-95.

108. McClung MR, Zanchetta JR, Hoiseth A, Kendler DL, Yuen CK, Brown JP, et al. Denosumab densitometric changes assessed by quantitative computed tomography at the spine and hip in postmenopausal women with osteoporosis. J Clin Densitom. 2013;16(2):250-6.

109. Tsai JN, Nishiyama KK, Lin D, Yuan A, Lee H, Bouxsein ML, et al. Effects of Denosumab and Teriparatide Transitions on Bone Microarchitecture and Estimated Strength: the DATA-Switch HR-pQCT study. J Bone Miner Res. 2017;32(10):2001-9.

110. Ideno H, Takanabe R, Shimada A, Imaizumi K, Araki R, Abe M, et al. Protein related to DAN and cerberus (PRDC) inhibits osteoblastic differentiation and its suppression promotes osteogenesis in vitro. Exp Cell Res. 2009;315(3):474-84.

111. Im J, Kim H, Kim S, Jho EH. Wnt/beta-catenin signaling regulates expression of PRDC, an antagonist of the BMP-4 signaling pathway. Biochem Biophys Res Commun. 2007;354(1):296-301.

112. Araujo AB, Esche GR, Kupelian V, O'Donnell AB, Travison TG, Williams RE, et al. Prevalence of symptomatic androgen deficiency in men. J Clin Endocrinol Metab. 2007;92(11):4241-7.

113. Araujo AB, O'Donnell AB, Brambilla DJ, Simpson WB, Longcope C, Matsumoto AM, et al. Prevalence and incidence of androgen deficiency in middle-aged and older men: estimates from the Massachusetts Male Aging Study. J Clin Endocrinol Metab. 2004;89(12):5920-6.

114. Jasuja GK, Bhasin S, Reisman JI, Berlowitz DR, Rose AJ. Ascertainment of Testosterone Prescribing Practices in the VA. Med Care. 2015;53(9):746-52.

115. Layton JB, Li D, Meier CR, Sharpless JL, Sturmer T, Jick SS, et al. Testosterone lab testing and initiation in the United Kingdom and the United States, 2000 to 2011. J Clin Endocrinol Metab. 2014;99(3):835-42.

116. Mulligan T, Frick MF, Zuraw QC, Stemhagen A, McWhirter C. Prevalence of hypogonadism in males aged at least 45 years: the HIM study. Int J Clin Pract. 2006;60(7):762-9.

117. Walsh TJ, Shores MM, Fox AE, Moore KP, Forsberg CW, Kinsey CE, et al. Recent trends in testosterone testing, low testosterone levels, and testosterone treatment among Veterans. Andrology. 2015;3(2):287-92.

118. Lindhé A AEH, Gustafsson l, Delphin S, Mellén A, Stoopendahl A, Pouragheli D, Quester J, Lundberg L. Prognosrapport: Kostnadsutvecklingen för läkemedel i Västra Götalandsregionen 2018-2020. 2018.

119. Lindhé A LC, Nylén K, Assadi-Nejad B, Gustafsson L, Mellén A, Wall U. Prognosrapport: kostnadsutveckling för läkemedel i Västra Götalandsregionen 2015-2016. 2015.

120.   Blackman MR, Sorkin JD, Munzer T, Bellantoni MF, Busby-Whitehead J, Stevens TE, et al. Growth hormone and sex steroid administration in healthy aged women and men: a randomized controlled trial. JAMA. 2002;288(18):2282-92.

121.   Kenny AM, Prestwood KM, Gruman CA, Marcello KM, Raisz LG. Effects of transdermal testosterone on bone and muscle in older men with low bioavailable testosterone levels. J Gerontol A Biol Sci Med Sci. 2001;56(5):M266-72.

122.   Page ST, Amory JK, Bowman FD, Anawalt BD, Matsumoto AM, Bremner WJ, et al. Exogenous testosterone (T) alone or with finasteride increases physical performance, grip strength, and lean body mass in older men with low serum T. J Clin Endocrinol Metab. 2005;90(3):1502-10.

123.   Saad F, Aversa A, Isidori AM, Gooren LJ. Testosterone as potential effective therapy in treatment of obesity in men with testosterone deficiency: a review. Curr Diabetes Rev. 2012;8(2):131-43.

124.   Snyder PJ, Peachey H, Hannoush P, Berlin JA, Loh L, Lenrow DA, et al. Effect of testosterone treatment on body composition and muscle strength in men over 65 years of age. J Clin Endocrinol Metab. 1999;84(8):2647-53.

125.   Srinivas-Shankar U, Roberts SA, Connolly MJ, O'Connell MD, Adams JE, Oldham JA, et al. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. J Clin Endocrinol Metab. 2010;95(2):639-50.

126.   Haring R, Ittermann T, Volzke H, Krebs A, Zygmunt M, Felix SB, et al. Prevalence, incidence and risk factors of testosterone deficiency in a population-based cohort of men: results from the study of health in Pomerania. Aging Male. 2010;13(4):247-57.

127.   Laaksonen DE, Niskanen L, Punnonen K, Nyyssonen K, Tuomainen TP, Valkonen VP, et al. The metabolic syndrome and smoking in relation to hypogonadism in middle-aged men: a prospective cohort study. J Clin Endocrinol Metab. 2005;90(2):712-9.

128.   Travison TG, Araujo AB, Kupelian V, O'Donnell AB, McKinlay JB. The relative contributions of aging, health, and lifestyle factors to serum testosterone decline in men. J Clin Endocrinol Metab. 2007;92(2):549-55.

129.   English KM, Mandour O, Steeds RP, Diver MJ, Jones TH, Channer KS. Men with coronary artery disease have lower levels of androgens than men with normal coronary angiograms. Eur Heart J. 2000;21(11):890-4.

130.   Haffner SM, Shaten J, Stern MP, Smith GD, Kuller L. Low levels of sex hormone-binding globulin and testosterone predict the development of non-insulin-dependent diabetes mellitus in men. MRFIT Research Group. Multiple Risk Factor Intervention Trial. Am J Epidemiol. 1996;143(9):889-97.

131.   Basaria S, Coviello AD, Travison TG, Storer TW, Farwell WR, Jette AM, et al. Adverse events associated with testosterone administration. N Engl J Med. 2010;363(2):109-22.

132.   Vigen R, O'Donnell CI, Baron AE, Grunwald GK, Maddox TM, Bradley SM, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. JAMA. 2013;310(17):1829-36.

133.   Corona G, Rastrelli G, Monami M, Saad F, Luconi M, Lucchese M, et al. Body weight loss reverts obesity-associated hypogonadotropic hypogonadism: a systematic review and meta-analysis. Eur J Endocrinol. 2013;168(6):829-43.

134.   Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. Genes Nutr. 2011;6(1):27-43.

135.   Taylor PN, Richmond R, Davies N, Sayers A, Stevenson K, Woltersdorf W, et al. Paradoxical Relationship Between Body Mass Index and Thyroid Hormone Levels: A Study Using Mendelian Randomization. J Clin Endocrinol Metab. 2016;101(2):730-8.

136.   Sabik OL, Farber CR. Using GWAS to identify novel therapeutic targets for osteoporosis. Transl Res. 2017;181:15-26.

137.   Alam I, Alkhouli M, Gerard-O'Riley RL, Wright WB, Acton D, Gray AK, et al. Osteoblast-Specific Overexpression of Human WNT16 Increases Both Cortical and Trabecular Bone Mass and Structure in Mice. Endocrinology. 2016;157(2):722-36.

138.   Moverare-Skrtic S, Wu J, Henning P, Gustafsson KL, Sjogren K, Windahl SH, et al. The bone-sparing effects of estrogen and WNT16 are independent of each other. Proc Natl Acad Sci U S A. 2015;112(48):14972-7.

139.   Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-9.

140.   Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. Am J Hum Genet. 2001;68(6):1463-74.

141.   Chen Z, Zheng G, Ghosh K, Li Z. Linkage disequilibrium mapping of quantitative-trait Loci by selective genotyping. Am J Hum Genet. 2005;77(4):661-9.

142.   Huang BE, Lin DY. Efficient association mapping of quantitative trait loci with selective genotyping. Am J Hum Genet. 2007;80(3):567-76.

143.   Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youlten SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet. 2017;49(10):1468-75.

144.   Lee SH, Lee SW, Ahn SH, Kim T, Lim KH, Kim BJ, et al. Multiple gene polymorphisms can improve prediction of nonvertebral fracture in postmenopausal women. J Bone Miner Res. 2013;28(10):2156-64.

145.   Zillikens MC, Demissie S, Hsu YH, Yerges-Armstrong LM, Chou WC, Stolk L, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. Nat Commun. 2017;8(1):80.

146.   Chen GB, Lee SH, Montgomery GW, Wray NR, Visscher PM, Gearry RB, et al. Performance of risk prediction for inflammatory bowel disease based on genotyping platform and genomic risk score method. BMC Med Genet. 2017;18(1):94.

147.   Desikan RS, Fan CC, Wang Y, Schork AJ, Cabral HJ, Cupples LA, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. PLoS Med. 2017;14(3):e1002258.

148.   Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet. 2015;11(4):e1004969.

149.   Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ. 2018;360:j5757.

150.   Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007;17(10):1520-8.

151.   Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry. 2014;55(10):1068-87.

152.   Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. 2013;9(2):e1003264.

153.   Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9(3):e1003348.

154.   Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. Genet Epidemiol. 2018;42(1):4-19.

155.   International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.

156.   Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421-7.