

# The latest news from the GENOMOS study

André G. Uitterlinden<sup>a,b,c</sup>

<sup>a</sup> Department of Internal Medicine, <sup>b</sup> Department of Clinical Chemistry, <sup>c</sup> Department of Epidemiology & Biostatistics  
Erasmus MC, Rotterdam, The Netherlands

Address for correspondence:  
Prof. Dr. A.G. Uitterlinden  
Professor of Complex Genetics  
Head Human Genotyping Facility  
Ee575 Genetic Laboratory  
Department of Internal Medicine  
Erasmus MC, PO Box 2040, NL-3000-CA Rotterdam,  
The Netherlands  
Ph. +31 10 7043573  
Fax +31 10 7035430  
E-mail: a.g.uitterlinden@erasmusmc.nl

## Summary

Most common age-related diseases such as osteoporosis, have strong genetic influences and therefore intense efforts are ongoing to identify the underlying genetic variants. Knowledge of these variants can help in understanding the disease process and might benefit development of interventions and diagnostics. Association studies have now become the standard approach to uncover the genetic effects of common variants. Yet, in all fields of complex disease genetics – including osteoporosis – progress in identifying these genetic factors has been hampered by often controversial results. Because of the complicated genetic architecture of the diseases and the small effect size for each individual risk alleles, this is mostly due to low statistical power and limitations of analytical methods. It is now recognised that association analysis followed by replication and prospective multi-centred meta-analysis is currently the best way forward to identify genetic markers for complex traits, such as osteoporosis. To accomplish this, large (global) collaborative consortia have been established that have large collections of DNA samples from subjects with a certain phenotype and that use standardized methodology and definitions, to quantify by meta-analysis the subtle effects of the responsible gene variants. The GENOMOS consortium has played such a role in the field of osteoporosis and has initially identified and refuted associations of well known candidate genes. This consortium is now expected to play an important role in validation of risk alleles coming from Genome Wide Association Studies (GWAS) for osteoporosis, some of which have just been published. Together with genetic studies on more rare syndromes, the GWA approach in combination with the GENOMOS consortium, is likely to help in clarifying the genetic architecture of complex bone traits such as BMD, and – eventually – in understanding the genetics of clinically relevant endpoints in osteoporosis, i.e., fracture risk. Such genetic insights will be useful in understanding biology and are likely to also find applications in clinical practice.

KEY WORDS: genetic, osteoporosis, polymorphisms, GWAS.

## Osteoporosis has genetic influences

Certain aspects of osteoporosis have been found to have strong genetic influences. This can be derived, for example, from genetic epidemiological analyses which showed that, in women, a maternal family history of fracture is positively related to fracture risk (1). Most evidence, however, has come from twin studies on bone mineral density (BMD) (2-6). For BMD the heritability has been estimated to be high: 50-80% (2-5). Thus, although twin studies can overestimate the heritability, a considerable part of the variance in BMD values might be explained by genetic factors while the remaining part could be due to environmental factors and to gene-environment interactions. This also implicates that there are “bone density” genes, variants of which will result in BMD levels that are different between individuals. These differences can become apparent in different ways, for example, as peak BMD or as differences in the rates of bone loss at advanced age. While this notion has resulted in much attention being paid to the genetics of BMD in the field of osteoporosis, it is likely that this attention is also due simply to the wide-spread availability of devices to measure BMD. This does not necessarily imply that BMD is the most important biological parameter of bone strength to consider. At the same time it is important to realize that (low) BMD is but one of many risk factors for osteoporotic fracture, the clinically most relevant endpoint of the disease.

Heritability estimates of fracture risk have been – understandably – much more limited due to the scarcity of good studies allowing precise estimates. Collecting large collections of related subjects with accurate standardized fracture data is notoriously difficult in view of the advance age at which they occur. While documenting a fracture event is now possible in several longitudinal studies, excluding a fracture event in those who report no fracture (“the controls”) is more difficult because they could still suffer a fracture later in life. One option to overcome this might be to take controls which are (much) older. In the case of hip fracture patients (with a mean age of 80 yrs) this would require control subjects of 90-100 yrs. It is questionable whether such healthy survivors are proper controls for fracture cases.

Andrew et al. (5) recently studied 6570 white healthy UK female volunteer twins between 18 and 80 years of age, and identified and validated 220 non traumatic wrist fracture cases. They estimated a heritability of 54% for the genetic contribution to liability of wrist fracture in these women. Interestingly, while BMD was also highly heritable, the statistical models showed very little overlap of shared genes between the two traits in this study.

Michealsson et al. (6) studied 33,432 Swedish twins (including 6,021 twins with any fracture, 3,599 with an osteoporotic fracture, and 1,055 with a hip fracture after the age of 50 years) and concluded that heritability of hip fracture overall was 48% but was 68% in twins younger than 69 years, and decreased to 3% in elderly twins 79 years and older. Indeed, another Finish study of elderly twins showed very little heritability for risk of fracture (7). Altogether, this suggests that although fracture risk is genetically determined, at older age other factors, perhaps

environmental factors, are more important in explaining variance in fracture risk.

While it might be difficult to demonstrate fracture risk is heritable, one can also argue that it follows from simple logic reasoning that aspects of osteoporosis, including fracture risk, must have a genetic influence. We know that DNA is the blueprint of life, and that the genotype differs between individuals, and that phenotypes differ between individuals. Thus, the difficulties in demonstrating heritability of fracture risk are probably due to limitations of our methods and approaches of measuring it.

The heritability estimates of osteoporosis indicate a considerable influence of environmental factors which can be modifying the effect of genetic predisposition. Gene-environment interactions one can think of, in this respect, include diet, exercise and exposure to sunlight (for vitamin D metabolism), for example. While genetic predisposition will be constant during life, environmental factors tend to change during the different periods of life resulting in different "expression levels" of the genetic susceptibility. Ageing is associated with a general functional decline resulting in, for example, less exercise, less time spent outdoors, changes in diet, etc. This can result in particular genetic susceptibilities being revealed only later on in life after a period when they went unnoticed due to sufficient exposure to one or more environmental factors.

Taking all this into account it becomes evident that osteoporosis is, not very surprisingly, considered a truly "complex" genetic trait. This complex character is shared with other common and often age-related traits with genetic influences such as diabetes, schizophrenia, Alzheimer's disease, osteoarthritis, cancer, etc. "Complex" means that a trait is multifactorial as well as multi-genic. Thus, genetic risk factors (i.e., certain alleles or gene variants) will be transmitted from one generation to the next, but the expression of these genotype factors in the final phenotype ("the penetrance") will be dependent on interaction with other gene variants and with environmental factors.

Given that the Human Genome Project has now resulted in the identification of nearly all genes in the human genome, it is not very surprising that most attention in the analysis of gene-environment interactions has gone to the genes, also referred to as the "genocentric" approach. The idea behind this is that once we know which gene variants are involved, it will be more

straightforward to analyse the contribution of environmental factors and their interplay with genetic factors.

### Risk gene identification in complex genetic diseases

Most common diseases such as diabetes, osteoporosis and cardiovascular diseases as well as many disease-related so-called intermediate traits or endophenotypes such as cholesterol levels, glucose levels, and bone mineral density, have strong genetic influences meaning that genetic variants will exist that contribute to explain this heritability. Yet, the identification of genetic factors underlying these disorders and traits and clarifying their genetic architecture has been very problematic, given the complex nature of the phenotypes and the limited molecular tools available at the time to identify the underlying genetic factors.

Complex diseases are typically influenced by many genetic variants each with modest effect size while the variability in expression of the disease phenotype is most likely also influenced by environmental factors in interaction with the genetic factors. Figure 1 shows the approaches most commonly used in the past two decades to identify genetic susceptibility factors for such complex diseases: the top-down genome-wide approaches and the bottom-up candidate gene approaches. It is safe to say now that linkage approaches in related subjects have been unsuccessful to identify genetic factors in complex disease. This is most likely due to the low power of this approach to detect the subtle effects and to the low "genetic resolution" meaning that very large chromosomal areas were potentially identified but with many possible candidate genes in them. On the other hand, the candidate gene approach in association studies has frequently suffered from irreproducible results mostly due to limited samples size and lack of standardization in phenotyping and genotyping. The GENOMOS consortium was started to address the problems in the candidate gene association analysis in particular, but since then has shown to be useful also for other approaches such as Genome Wide Association Studies (GWAS). I will first briefly discuss the classical association study design, followed by a description of the GENOMOS consortium, and then end with the recent GWAS on osteoporosis and the start of the GEFOS project.

Type of approach	Resolution	Effectiveness	
		Common risk alleles	Rare risk alleles
<i>"Top-down" / hypothesis free</i>			
<b>* Genome-Wide Linkage Analysis</b> - Pedigrees - Sib-pairs - Human, mouse	5-20 million bp	-	+/-
<b>* Genome-Wide Association (GWA) Analysis</b> - 100K - 2000K SNP analysis in cases/controls	5-50 thousand bp	+	-
<b>* Genome-Wide Sequencing</b> - full genome not yet feasible on large scale	1 bp	(?)	(?)
<i>"Bottom-up" / up-front hypothesis</i>			
<b>* Association Analyses of Candidate Gene Polymorphisms (based on biology)</b>	1 bp	+/-	+/-
<b>* Candidate Sequencing</b> - Selected regions (e.g., exons, gene regions)	1 bp	+	+

Figure 1 - Some characteristics of the most commonly used molecular approaches to identify susceptibility alleles for complex disorders. "Resolution" indicates the size of the chromosomal area that is identified as being linked/associated to the phenotype of interest, and which can vary from one base pair to many millions of base pairs. "Effectiveness" indicates the success rate of the method to identify risk alleles for complex genetic diseases and phenotypes, either common or rare, as derived from publications.

### Association analysis of candidate gene polymorphisms

The *bottom-up* approach to identify genetic risk factors for osteoporosis builds upon biology, i.e., the known involvement of a particular gene in aspects of osteoporosis, e.g., bone metabolism. This gene is then referred to as a “candidate gene”. The candidacy of such a gene can be established by several lines of evidence:

1. Cell biological and molecular biological experiments indicating for example bone cell-specific expression of the gene.
2. Animal models in which a gene has been mutated (e.g., natural mouse mutants), over-expressed (transgenic mice), or deleted (knock-out mice) and which result in a bone-phenotype.
3. Naturally occurring mutations of the human gene resulting in monogenic Mendelian diseases with a bone phenotype.
4. More recently, any “hit” from a Genome Wide Association Study (GWAS).

Subsequently, in the candidate gene frequently occurring sequences variants (polymorphisms) have to be identified which supposedly lead to subtle differences in level and/or function of the encoded protein. We distinguish mutations from polymorphisms purely on the basis of frequency: polymorphisms occur in at least 1% of the population, mutations in less. The most common DNA sequence variant now being studied is the Single Nucleotide Polymorphism or SNP, which is the most common type of variation in the human genome, but of course several other types of sequence variation need consideration such as variable number of tandem repeats (VNTR) and copy number variations (CNV). Yet, these require specialised technology to study in large populations, and therefore will await later studies, where as for SNPs most technology seems now in place, resulting in many studies on SNPs in relation to complex diseases.

Several databases are now available which contain information on DNA sequence variation, especially on the common variants in any gene of the Human Genome (e.g., dbSNP from, NCBI, Celera, HapMap, and several more specialized databases such as from the Program for Genomic Analysis (PGA). Common DNA sequence variations were usually regarded as just polymorphic (so called “anonymous” polymorphisms) until proven otherwise, but this view is changing. Many of them have now been shown to have consequences for the level and/or activity of the protein encoded (functional polymorphisms). These can include, e.g., sequence variations leading to alterations in the amino acid composition of the protein, changes in the 5' promoter region leading to differences in mRNA expression, and/or polymorphisms in the 3' region leading to differences in mRNA degradation. In particular, the GWAS (see below) have identified polymorphisms which can be very far away from the actual gene of interest and most likely are involved in fine regulation of the gene of interest. As a result of this large amount of evidence that is being accumulated for DNA polymorphisms we are now regarding all of them as potentially functional, until proven otherwise.

Polymorphisms of interest are usually first tested in population-based and/or case-control “association studies”, to evaluate their contribution to the phenotype of interest at the population level. However, association studies do not establish cause and effect; they just show correlation or co-occurrence of one with the other. Cause and effect has to be established in truly functional cellular and molecular biological experiments involving, e.g., transfection of cell lines with allelic constructs and testing activities of the different alleles. This can occur at different levels of organization and depends on the type of protein analysed, e.g., enzymes, vs. matrix molecules vs. transcription factors. Acknowledging these complexities it will remain a challenge, once an association has been observed, to identify the

correct test of functionality. And vice versa once functionality has been established, to identify the correct endpoint in an epidemiological study.

Because functional polymorphisms lead to meaningful biological differences in function of the encoded “osteoporosis” protein this also makes the interpretation of association analyses using these variants quite straightforward. For example, for functional polymorphisms it is expected that the same allele will be associated with the same phenotype in different populations. This can even be extended to similar associations being present in different ethnic groups, although allele frequencies can of course differ by ethnicity.

Out of the lines of evidence mentioned above, numerous candidate genes for risk of osteoporosis have emerged. These include “classical” candidate genes for osteoporosis such as collagen type I, the vitamin D receptor, and the oestrogen receptors. Yet, also recently identified “bone” genes, such as LRP5, can become candidate genes because their involvement in bone biology has now been established. These early genetic studies on monogenetic pedigrees in which an LRP5 mutation was segregating (such as in the High Bone Mass phenotype pedigrees or in osteoporosis pseudoglioma pedigrees) have identified LRP5 as a candidate gene for osteoporosis, but of course not established its role as a genetic risk factor for osteoporosis at the population level. Yet, very interestingly, work from the GENOMOS consortium as well as the recent GWAS have also identified LRP5 as a risk gene for osteoporosis (see below).

### Genetic effects: large vs. small and common vs. rare

From the analysis of the few successfully identified genetic risk factors for complex disorders it is by now clear that for complex disorders in general the risks associated with each individual genetic variant are generally modest in terms of effect size. For a number of DNA variants for several complex disorders a trend can be discerned whereby the more common variants are associated with smaller risks (such as PPAR $\gamma$  Pro12Ala in type 2 diabetes) than the more rare variants (such as FactorV Leiden and thrombosis).

While the risk of disease for a human subject is indeed small for such individual genetic risk variants, because there are so many of these common variants in the human genome, the combined effect – or genetic load – of these risk variants can be substantial both for the individual as well as for the population. One can speculate that evolution has allowed these common variants to float around in the human population because they do not compromise reproductive success (or might even enhance it) and only start to affect fitness of the individual carrying such variants late in life, far after the reproductive period. On the other end of the spectrum more rare variants will be selected out in evolution because they do affect reproductive success and/or will be private to individuals as newly arisen mutations.

Overall the current thought about underlying genetic risk variants of complex diseases such as osteoporosis, is that there will be several (maybe up to a hundred) common variants conferring risk, but any given individual will also carry several genetic variants that are very rare in the population and might have bigger associated genetic risks. As we will see below we now have sufficient technology to start identifying these more common effects with the smaller effect sizes. We will have to wait though until cost effective total human genome sequencing techniques become available to identify in individuals the collections of the much rarer sequence variants that perhaps confer larger effects.

These small effect sizes also explain why it has been difficult to

identify such risks convincingly, in spite of these genetic variants being so common. Common in this respect means allele frequencies of a genetic risk factor of 5-50% and modest effect sizes means odds ratios of 1.1-2.0. Statistical power calculations show that indeed very large study populations, of 1000 to 10,000 subjects of case-control collections and/or population-based cohorts need to be studied in order to demonstrate convincingly such small effects by association analysis. Only recently such large study populations have become available and consortia have been assembled to address these challenges in a robust manner. Yet, such large combined sets of data require meta-analysis to estimate true effect sizes of individual variants.

**Meta-analysis**

In the coming years we can expect to see more and more association analyses to be performed of an ever increasing list of candidate gene polymorphisms. It will therefore be necessary to put all these data in perspective by performing meta-analyses of the individual association analyses. Meta-analysis can quantify the results of various studies on the same topic and estimate and explain their diversity. Recent evidence indicates that a systematic meta-analysis approach can estimate population-wide effects of genetic risk factors for human disease (8) and that large studies are more conservative in these estimates and should preferably be used (9). An analysis of 301 studies on genetic associations (on many different diseases) concluded that there are many common variants in the human genome with modest, but real effects on common disease risk, and that studies using large samples will be able to convincingly identify such variants (10). This notion in the field of complex genetics has led to the creation of consortia of investigators working on the same disease, and then in particular on the genetics of complex diseases and traits. Whereas these consortia first operated in isolation, they are now sharing experience through the HuGENET (<http://www.cdc.gov/genomics/hugenet>) instigated network of networks (11,12). Among such consortia, GENOMOS (<http://www.genomos.eu>) as the network of investigators working on genetics of osteoporosis, was one of the first starting in 2003 and was involved in the first Network of Networks meetings.

Meta-analysis initially had some drawbacks because it was mostly based on combining sets of existing data resulting in sometimes substantial bias in the outcome. This is mainly because there is publication bias in the literature (positive studies reporting exaggerated effects) and there was virtually no standardization among investigators in methods of genotyping or phenotyping and data analysis. Yet, with the advent of growing consortia of investigators working on the subtle effects in complex genetics, the concept of meta-analysis has developed into one of *prospective* meta-analysis. Here, the investigators collectively perform genotyping under standardized conditions and agree on the outcomes, well before any outcome of individual studies is known. This approach will therefore include positive as well as negative studies on the polymorphism of interest. The GENOMOS consortium was one of the first networks to use such a prospective meta-analytic approach to start a systematic test of candidate gene polymorphisms in the field of osteoporosis.

**The GENOMOS consortium**

The EU-sponsored GENOMOS (Genetic Markers for Osteoporosis) consortium attempts to perform such studies using standardized methods of genotyping and phenotyping. The GENOMOS project involves the large-scale study of several candidate gene polymorphisms in relation to osteoporosis-related outcomes in subjects drawn from several European centers. Its main outcomes are fractures and femoral neck and lumbar spine BMD. The general research program is presented in Figure 2, while an overview of the participating centers and groups is given in Figure 3. Design details are further described in the first meta-analysis of individual-level data on the *ESR1* gene (13).

Apart from it being a very large study of genetics of complex disease with at the moment > 25,000 subjects included, an important aspect of this study is its prospective multi-center design. This means the genotype data are generated for all centers only *after* which the association analysis is done, thereby rendering it immune to possible publication bias. The targets of the study are polymorphisms for which some *a priori* evidence for involvement in osteoporosis is present already; it is not designed to be a risk gene-discovery tool and currently therefore

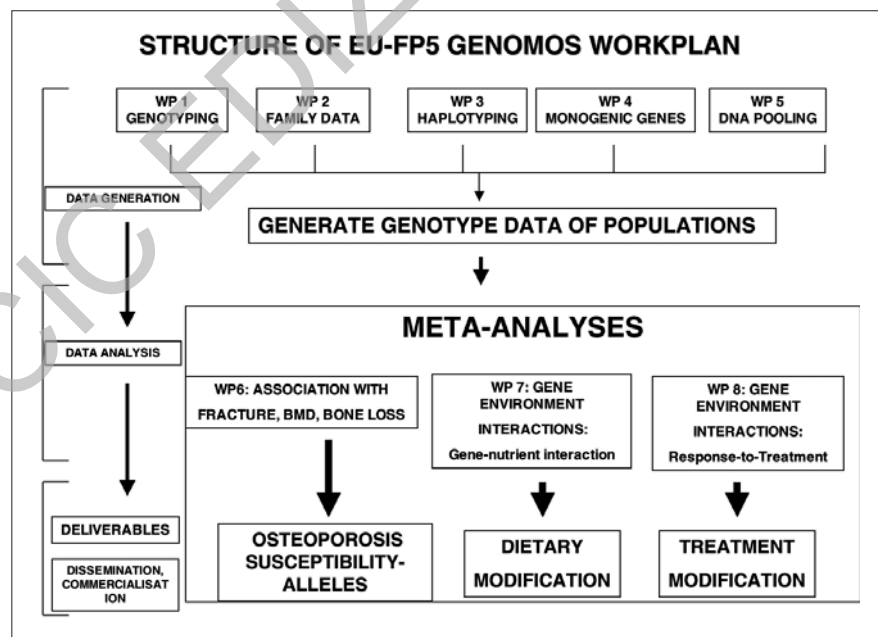


Figure 2 - An overview of the research program of the EU-FP5 sponsored GENOMOS project. There are several work packages that have helped in the phase of data generation, one of which is the actual genotyping of the participating populations (WP1). The other work packages were investigating other methods to optimize the process of genotyping and finding new candidate genes. In the phase of meta-analysis GENOMOS investigated the contribution of candidate gene polymorphisms to the main endpoints BMD and fracture risk, and explored the possibilities to determine the contribution of gene-environment interactions, in gene-drug interaction with HRT and in gene diet interaction with dietary calcium intake. In the final phase we identified the true risk alleles and their effect size.

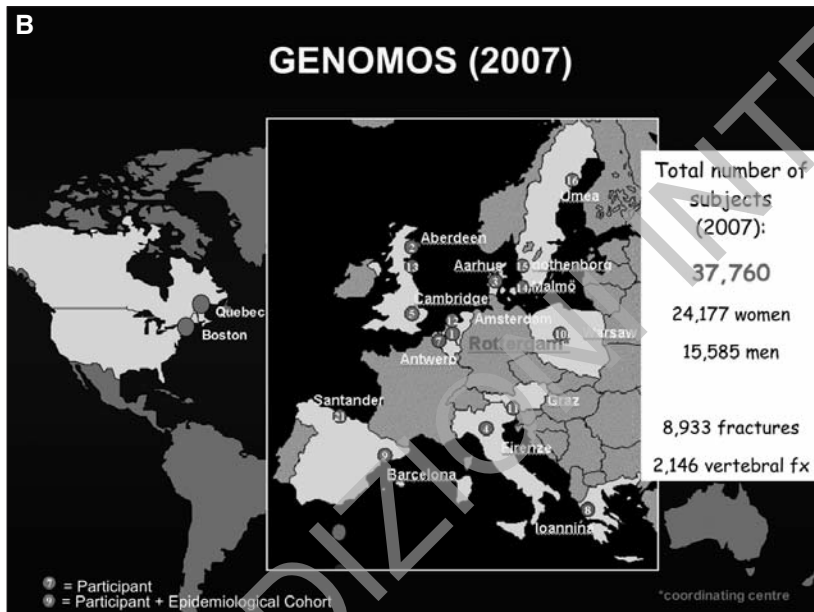
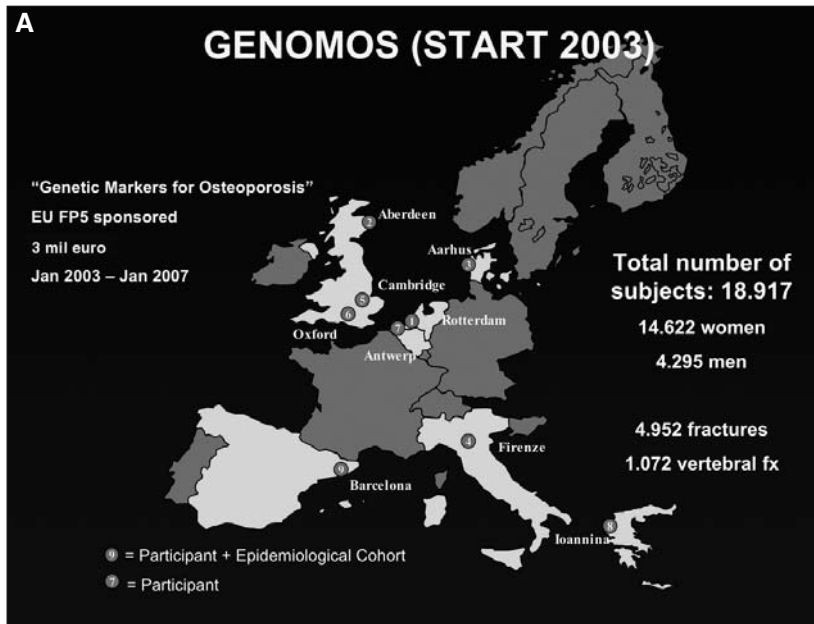


Figure 3 - A) A geographical overview of the initial participating centers in the GENOMOS consortium at the start in 2003. These include the centers with study populations used for genotyping and association analysis, and a centre specializing in finding causative genes in monogenetic bone disorders (Antwerp, Belgium), and the statistical centre for meta-analysis (Ioannina, Greece). B) Idem, but at the end of the contractual period in early 2007. The number of centers with study populations has grown to now include > 36.000 subjects.

cannot, for example, assess all genetic diversity across a gene. While fracture has been debated as an endpoint in genetics of osteoporosis studies, this was chosen in the GENOMOS study because it is clinically the most relevant endpoint. Statistical power of the GENOMOS study to detect genetic effects on fracture risk is high with > 5,000 fractures.

An overview of all the meta-analyses published by the GENOMOS consortium so far is presented in Table I. The very first GENOMOS meta-analysis of three polymorphisms in the *ESR1* gene (intron 1 polymorphisms XbaI and PvuII and the promoter (TA) variable number of tandem repeats micro-satellite) and haplotypes thereof, among 18 917 individuals in 8 European centres, demonstrated no effects on BMD but a modest effect on fracture risk (19-35% risk reduction for XbaI homozygotes), independent of BMD (13). We then went on to study the Sp1 *COL1A1* gene polymorphism (14), 5 polymorphisms in the vitamin D receptor gene including the Cdx2 promoter variant, the FokI variant, and the BsmI, Apal and the TaqI variants (15),

5 polymorphisms in the *TGFbeta* gene (16), and the exon 9 and exon 13 variants in *LRP5* and the exon 9 variant in *LRP6* (17).

Overall, the major results of the GENOMOS study included the identification of the *LRP5* variants as true osteoporosis risk variants but with modest effects size. The *LRP5* effects were very consistent across different populations rendering very low p-values for the overall effect (although it was small), probably indicating that this is a universal genetic effect for osteoporosis which can be expected to pop-up in close to every population studied. In addition, we identified the *ESR1* SNPs as fracture risk factors and not so much as BMD associated variants and also showed that the Sp1 *COL1A1* variant was associated with a modest increase in vertebral fracture risk, as we did for the Cdx2 variant in the *VDR* gene. These results show that the candidate gene approach is fruitful in identification of osteoporosis risk alleles, but only when applied as rigorously as we did in GENOMOS. In addition, it showed that the effect size of

Table I - Overview of results on candidate gene associations from the GENOMOS study.

Gene (n=6)	SNPs (n=17)	Sample n	Associations with OP phenotypes				Publication
			BMD (SD)		Fracture (Odds Ratio)		
			Femoral neck	Lumbar spine	Vertebral	Non vertebral	
ESR1	3	18,917	–	–	1.2-1.3	1.1-1.2	Ioannidis et al., JAMA 2004 (13)
COLIA1	1	20,786	0.15	0.15	1.1 (Sp1)	–	Ralston et al., PLOS Med 2006 (14)
VDR	5	26,242	–	–	1.1 (Cdx1)	–	Uitterlinden et al., Ann Int Med 2006 (15)
TGFbeta	5	28,924	–	–	–	–	Langdahl et al., Bone 2008 (16)
LRP5	2	37,760	0.15	0.15	1.12 -1.26	1.06 -1.14	Van Meurs et al., JAMA 2008 (17)
LRP6	1	37,760	–	–	–	–	Ibid. (17)

SNPs = single nucleotide polymorphisms; OP = osteoporosis.

the risk alleles is modest. This could be simply due to our choice of candidate genes (not being able to picking out the most important risk genes for osteoporosis), but might also signal the general allelic architecture of BMD. In view of the recent GWAS results we think the latter is indeed the case.

Finally, we excluded these 5 TGFbeta variants to contribute to osteoporosis and most likely also other variants in the coding region of this prominent candidate gene for osteoporosis. This is equally important as finding risk alleles, as it signals to the scientific research community to not study TGFbeta variants further in relation to osteoporosis.

In the course of the GENOMOS project we have also evaluated several different approaches to find new potential genetic markers for osteoporosis. Of these, the work package on analysing monogenetic families (WP4) has been very successful in identifying new candidate genes, including LRP5 which was also identified in our consortium as a prominent risk gene or osteoporosis. Other approaches such as linkage analysis in families or TDT testing in sib pairs selected on (mild) osteoporosis (WP2) was found to be *not* successful. In addition, we tested some techniques for genetic association studies which were found to be helpful (WP3 on haplotyping) or not so helpful (WP5 LD mapping in pooled samples). These are all important messages to the scientific community on how to progress in the most efficient way in complex genetics.

No major effect of hormone replacement therapy (HRT) use was seen for the effect of ESR1 genotype on BMD or fracture risk, but the study was hampered by lack of standardised methods to assess HRT use and quality of the datasets. No major effect of dietary Ca intake was seen for the effect of VDR genotype on BMD or fracture risk, but the study was hampered by lack of standardised methods and quality of the datasets.

In conclusion we can now say that:

- GENOMOS is now the largest consortium of research groups working on genetics of osteoporosis
- GENOMOS has fulfilled a pioneering role in setting the stage for research in complex genetics.
- GENOMOS has identified and refuted genetic risk factors for osteoporosis
- The effects sizes of the identified genetic risk factors for osteoporosis is modest at best, with effects of 0.1 SD in BMD and 20-30% increases in risk for osteoporotic fracture.
- The results of GENOMOS activities have not yet lead to commercially interesting activities with any economic impact. This is due to the candidate gene approach it has taken so far, and the small effect sizes of individual genetic markers.

Although successful, drawbacks of the GENOMOS consortium include the fact that only well known candidate genes were an-

alyzed, so we could not expect to generate much new biology with the exception of WP4. GENOMOS only studied Caucasians so the generalizability of the findings is unknown for other ethnicities. The endpoints were limited to the classical osteoporosis endpoints BMD and fracture risk. Both of these are cumbersome to interpret in biological and clinical ways, and not telling the complete story about osteoporosis. Finally, no risk modeling was performed to assess the contribution of the genetic risk factors we identified in GENOMOS in relation to well-established osteoporosis risk factors such as age, gender, BMI, use of a walking aid, etc.

Therefore, to address these shortcomings of the GENOMOS project we went on to compose a follow-up proposal. We were lucky enough to see that under the FP7 program of the European Commission, a call for proposals was launched in which complementary wording to our plans could be read. We therefore submitted a new proposal, called GEFOS, under the FP7 program and were indeed selected and funded: the son of GENOMOS was born.

The GEFOS proposal ([www.gefos.org](http://www.gefos.org)) has started in march 2008 will use the several GWAS on osteoporosis as its starting point, will increase the sample size of the GENOMOS consortium for replication purpose of hits coming from the GWAS, will include additional bone phenotypes (such as CT and ultrasound) to study as endpoints in the meta-analysis and/or GWAS, and will introduce risk modeling of the identified genetic risk factors in relation to other osteoporosis risk factors.

### Genome-wide association studies

Due to technological developments the association study design has regained popularity, but now on a genome-wide scale with an unprecedented density of genetic markers: the genome-wide association study or GWAS (18). This renaissance has mostly been driven by the discovery of millions of Single Nucleotide Polymorphisms or SNPs throughout the human genome and the development of so-called micro-array technology to type such SNPs accurately on a massive parallel scale.

Genome Wide Association Studies (GWAS; see Figure 4) have only recently become available but already have had considerable success in identifying genetic susceptibility alleles. GWAS consists of screening the genome of many hundreds to thousands of subjects in a case-control study or population base cohort study, with > 500.000 Single Nucleotide Polymorphisms (SNPs), followed by a simple association analysis between a phenotype and all the genetic markers. Such a GWAS then identifies genetic markers associated to the phenotype of inter-

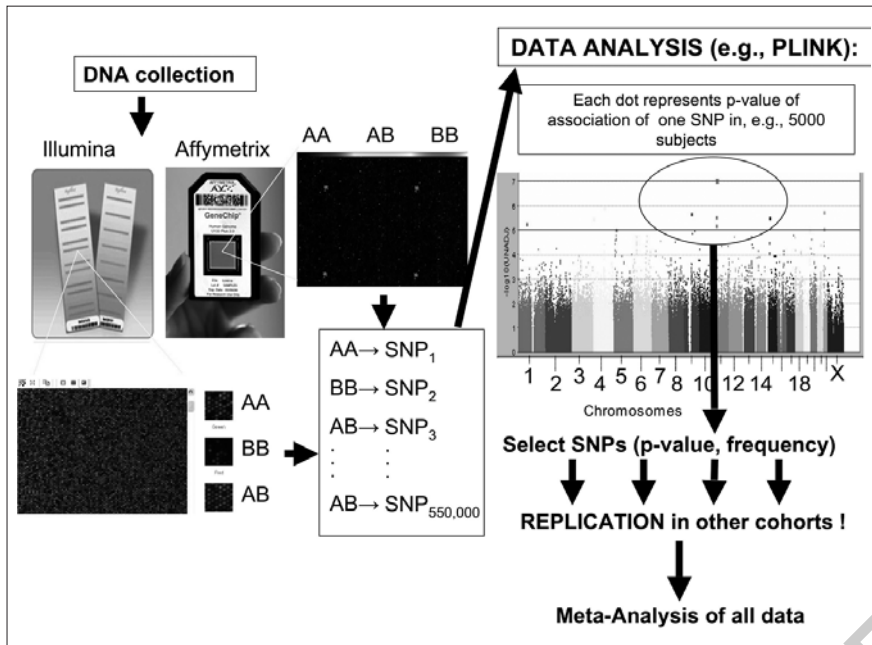


Figure 4 - A schematic overview of the process of performing a Genome Wide Association study. Each of the DNA samples from, for example, a typical case-control collection for a disease consisting of 5,000 DNA samples, are subjected to genome wide SNP genotyping using any of two platforms (Illumina and Affymetrix). These will generate ~0.5 million genotypes per DNA sample which combine to 2.5 billion genotypes across the 5,000 samples which are then analyzed for association to the disease by standard chi-square tests in programs such as PLINK. The output of such programs is usually depicted as so-called “Manhattan plots” whereby each dot represents the p-value for association of one SNP (out of the 0.5 million). Selected SNPs are then identified (“the low hanging fruit”) to be subjected to replication efforts in additional DNA collections with the phenotype of interest. A meta-analysis on the combined evidence will establish consistency of the association, effect-size and statistical significance.

Table II - Comparison of two GWAS on osteoporosis.

	TWINS UK/ ERGO NL	deCODE/DK/AUS
Author	Richards B, Rivadeneira F, et al. (23)	Styrkarsdottir U, et al. (24)
Journal	Lancet, april 29, 2008	NEJM, april 30, 2008
GWA discovery	Twins UK (n = 1586)	deCODE (n = 5861)
Platform	Illumina 317K	Illumina 317K
Replication cohorts	Rotterdam Study NL (n=4877) Chingford UK (n=718)	Iceland (n = 4165) Danish (n = 2269) Australian (n = 1491)
Total n	8,557	13,786
gws hits: # BMD loci	2: LRP5, OPG	5: OPG, RANKL, ESR1, ZFBTB40, MHC
Also fx risk?	1 (LRP5)	3 (+ 3 post hoc)
Explained variance BMD	~ 1 %	~ 3 %

est with a certain statistical significance. Because of the multiplicity of testing with so many markers certain thresholds have been considered to declare an association “genome wide significant” (gws). For example, when analyzing 500.000 markers this threshold is  $1.10^{-7}$  ( $0.05/500.000$ ) but given more recent imputation approaches to exploit the linkage relationships between the millions of SNPs in the HapMap database,  $5.10^{-8}$  is now a more widely used gws level in GWAS. Naturally, the chance of seeing such gws associations is depending on factors such as the effect size of the markers on the phenotype, and the size of the study sample. Usually, a typical GWAS consist of a discovery sample with GWA data and a replication sample with GWA data. From the discovery GWAS the “top-hits” (for example, all genetic markers that reach a significance of  $1.10^{-4}$  or less will be analyzed in the replication GWA dataset to see which markers will reach the gws level after combining the two datasets). Following such a GWAS, the remaining top hits (for example, any marker that has  $p < 1.10^{-5}$ ) will be analyzed in subsequent replication cohorts which do not necessarily have GWA data but which can be genotyped for the particular genetic markers identified.

The explained genetic variance in diabetes mellitus by the re-

cently uncovered common variants in risk genes by GWAS is still limited to – at most – 10% (18), and similar figures are observed for typical quantitative traits such as height (19-21). Yet, we must realize these are still early days in complex genetics because we have just uncovered the “low hanging fruit” in this first round of GWAS. The complex genetics research community has now organized themselves into still growing consortia of research groups who join (GWAS) datasets and are providing excellent forums for harmonizing phenotype definitions, such as GENOMOS ([www.genomos.eu](http://www.genomos.eu)). Such consortia will provide a very good testing ground to uncover more genetic variants in various disease areas.

### GWAS of osteoporosis

The very first attempt to identify BMD loci through GWAS is presented by investigators from the Framingham study using the 100K Affymetrix platform and a limited sample size of n=1141 men and women (22). This effort did not result in BMD loci that reached the so-called genome wide significance and made it clear that larger samples sizes were to be used and al-

so genotyping platforms with a higher genome coverage such as the Illumina 317K or 550K platforms. Recently, the first two GWAS on BMD have been published that identified several loci contributing to BMD that reached genome wide significance. One study was coming from analysis of the TwinsUK cohort and the Rotterdam Study (23), and the other study was from deCODE (24) (see Table II for an overview of the two studies). These two studies represent the first large scale high density GWAS efforts reporting BMD loci from a hypothesis free approach. One locus was overlapping between the two efforts, i.e., the OPG locus on chromosome 8, while the TwinsUK./ERGO effort reported on additional locus, i.e., LRP5, and the deCODE GWAS reported 4 additional loci. LRP5 was interesting because this gene and indeed this particular SNP (the exon 13 one) was just one month earlier reported by the GENOMOS consortium to be associated with very high confidence to BMD in the study by van Meurs et al.(17). The explained variance of the genetic factors for BMD that was reached by both studies was low (1-3%). This indicates that, very similar to the GWAS results for height (19-21), BMD is a truly complex trait with many loci (hundreds?) of small effect. This also indicates that even larger samples sizes than the one used by the deCODE study (n=13,000) are necessary to identify these additional common factors. Taken together, this puts the GEFOS effort now in the spot light because this consortium is able to eventually accumulate > 40,000 samples with GWAS data and therefore well-powered to identify the second wave of BMD loci in combination with the expanded GENOMOS consortium to include >100.000 replication samples.

#### Rare variants

Thus, we have experienced a plethora of these GWAS that have produced dozens of common variants that confer modest risk for a variety of common disorders and traits. Yet, the current round of GWAS tend to focus on this low hanging fruit while we know that there are many more such common and less common variants to be discovered with less impressive p-values in the discovery phase due to even smaller effect size and/or even smaller population frequency. Identifying such small effects is possible but will require even larger sample sizes to detect them in a statistically robust and convincing way.

While all GWAS currently focus on the common variants (say, >5% population frequency) we also suspect that there are less common variants (0.5-5% population frequency) and even rare variants (<0.5% population frequency) that will contribute to explain risk of disease or variance of a trait of interest. Examples highlighting the existence of such less common variant was demonstrated for sequence variations in cholesterol-related genes, in relation to cholesterol levels. An example of this was described by Cohen et al. (25) who tested whether rare DNA sequence variants collectively contribute to variation in plasma levels of high density lipoprotein cholesterol (HDL-C). They sequenced three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) that cause Mendelian forms of low HDL-C levels in individuals from a population-based study. Non synonymous sequence variants were significantly more common (16% versus 2%) in individuals with low HDL-C (<fifth percentile) than in those with high HDL-C (>95<sup>th</sup> percentile). Similar findings were obtained in an independent population, and biochemical studies indicated that most sequence variants in the low HDL-C group were functionally important. Thus, rare alleles with major phenotypic effects contribute significantly to low plasma HDL-C levels in the general population. Similarly, such rare alleles of bone genes might contribute to variation in BMD and other bone parameters, and even fracture risk in the general population.

The challenge for the future will therefore be to identify also such more rare variants through deep sequencing approaches of the many genes identified through GWAS. Combinations of such rare and the more common genetic variants in these particular genes can then be scrutinized for their diagnostic potential in large well phenotyped cohorts, also in relation to the more classical risk factors. Such combinations of genetic risk factors are expected to explain more of the genetic risk for particular common diseases than just the common variants or just the (very) rare variants.

#### Acknowledgements

The work described was sponsored by the European Commission under contract Number QLK6-CT-2002-02629 (GENOMOS) and FP7-HEALTH-2007-201865 (GEFOS). I would like to thank my colleagues Dr Joyce van Meurs and Dr Fernando Rivadeneira for their invaluable help with the projects and their inspiring discussions and criticism. I also thank Pascal Arp, Mila Jhamai, Lisette Stolk, and Yue Fang who have all contributed their help and expertise in the course of the GENOMOS project. Finally, I would like to thank all of the many national and international co-workers within the GENOMOS consortium, for their inspiring help and contributions.

#### References

1. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, Cauley J, Black D, Vogt TM. Risk factors for hip fracture in white women. *New Engl J Med.* 1995;332:767-773.
2. Smith DM, Nance WE, Kang KW, Christian JC, Johnston CC. Genetic factors in determining bone mass. *J Clin Invest.* 1973; 52:2800-2808.
3. Pocock NA, Eisman JA, Hopper JL, Yeates GM, Sambrook PN, Ebert S. Genetic determinants of bone mass in adults: a twin study. *J Clin Invest.* 1987;80:706-710.
4. Flicker L, Hopper JL, Rodgers L, Kaymakci B, Green RM, Wark JD. Bone density determinants in elderly women: a twin study. *J Bone Miner Res.* 1995;10:1607-1613.
5. Andrew T, Antoniadou 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.
6. Michaelsson K, Melhus H, Ferm H, Ahlborn A, Pedersen NL. Genetic liability to fractures in the elderly. *Arch Intern Med.* 2005;165: 1825-1830.
7. Kannus P, Palvanen M, Kaprio J, Parkkari J, Koskenvuo M. Genetic factors and osteoporotic fractures in elderly people: prospective 25 year follow up of a nationwide cohort of elderly Finnish twins. *BMJ.* 1999;319:1334-1337.
8. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet.* 2001;29(3):306-309.
9. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet.* 2003;361(9357):567-571.
10. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet.* 2003;33(2):177-182.
11. Ioannidis JP, Bernstein J, Boffetta P, Danesh J, Dolan S, Hartge P, Hunter D, Inskip P, Jarvelin MR, Little J, Maraganore DM, Bishop JA, O'Brien TR, Petersen G, Riboli E, Seminara D, Taioli E, Uitterlinden AG, Vineis P, Winn DM, Salanti G, Higgins JP, Khoury MJ. A network of investigator networks in human genome epidemiology. *Am J Epidemiol.* 2005;162(4):302-304.
12. Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffet-



- ta P, Bondy M, Bray MS, Brenchly PE, Buffler PA, Casa JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ. Human Genome Epidemiology Network and the Network of Investigator Networks. A roadmap for efficient and reliable human genome epidemiology. *Nat Genet.* 2006;38(1): 3-5.
13. Ioannidis JPA, Ralston SH, Bennett ST, Brandi ML, Grinberg D, Karassa FB, Langdahl BL, van Meurs JBJ, Mosekilde L, Scollen S, Albagha OME, Bustamante M, Carey AH, Dunning AM, Enjuanes A, van Leeuwen JPTM, Mavilia C, Masi L, McGuigan FEA, Noguez X, Pols HAP, Reid DM, Schuit SCE, Sherlock RE, Uitterlinden AG for the GENOMOS Study. Large-scale evidence for differential genetic effects of ESR1 polymorphisms on osteoporosis outcomes: The GENOMOS Study. *JAMA.* 2004;292(17):2105-2114.
  14. Ralston SH, Uitterlinden AG, Brandi ML, Balcells S, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Scollen S, Bustamante M, Husted LB, Carey AH, Diez-Perez A, Dunning AM, Falchetti A, Karczmarewicz E, Kruk M, van Leeuwen JPTM, van Meurs JBJ, Mangion J, McGuigan FEA, Mellibovsky L, Monte FD, Pols HAP, Reeve J, Reid DM, Renner W, Rivadeneira F, van Schoor NM, Sherlock RE, Ioannidis JPA. Large-scale evidence for the effect of the COL1A1 Sp1 polymorphism on osteoporosis outcomes: the GENOMOS study. *PLoS Med.* 2006;EPub 3(4):e90.
  15. Uitterlinden AG, Ralston SH, Brandi ML, Carey AH, Grinberg D, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Reeve J, Reid DM, Amidei A, Bassiti A, Bustamante M, Husted LB, Diez-Perez A, Dobnig H, Dunning AM, Enjuanes A, Fahrleitner-Pammer A, Fang Y, Karczmarewicz E, Kruk M, van Leeuwen JPTM, Mavilia C, van Meurs JBJ, Mangion J, McGuigan FEA, Pols HAP, Renner W, Rivadeneira F, van Schoor NM, Scollen S, Sherlock RE, Ioannidis JPA, APOSS investigators, EPOS investigators, EPOLOS investigators, FAMOS investigators, LASA investigators, and Rotterdam Study investigators; for the GENOMOS Study. Large-scale analysis of association between common vitamin D receptor gene variations and osteoporosis: the GENOMOS study. *Ann Int Medicine.* 2006;145(4):255-64.
  16. Langdahl BL, Uitterlinden AG, Ralston SH, Trikalinos TA, Balcells S, Brandi ML, Scollen S, Lips P, Lorenc R, Obermayer-Pietsch B, Reid DM, Armas JB, Arp PP, Bassiti A, Bustamante M, Husted LB, Carey AH, Pérez Cano R, Dobnig H, Dunning AM, Fahrleitner-Pammer A, Falchetti A, Karczmarewicz E, Kruk M, van Leeuwen JP, Masi L, van Meurs JB, Mangion J, McGuigan FE, Mellibovsky L, Mosekilde L, Nogués X, Pols HA, Reeve J, Renner W, Rivadeneira F, van Schoor NM, Ioannidis JP; APOSS investigators; DOPS investigators; EPOS investigators; EPOLOS investigators; FAMOS investigators; LASA investigators; ERGO investigators; for the GENOMOS Study. Large-scale analysis of association between polymorphisms in the transforming growth factor beta 1 gene (TGFB1) and osteoporosis: The GENOMOS study. *Bone.* 2008 May;42(5):969-981. Epub 2007 Dec 3.
  17. van Meurs JBJ, Trikalinos T, Ralston SH, Brandi ML, Brixen K, Diez-Perez A, Jonell O, Kiel D, Langdahl BL, Lips P, Lorenc R, Obermayer-Pietsch B, Ohlson C, Pettersson U, Reeve J, Reid DM, Rousseau F, Akeson K, Arp PP, Balcells S, Bassiti A, Bustamante M, Husted LB, Dobnig H, Dunning AM, Fahrleitner-Pammer A, Falchetti A, Grinberg D, Karczmarewicz E, Kruk M, Ljunggren O, Kaptoge S, Karasik D, Masi L, McGuigan FEA, Mellibovsky L, Mosekilde L, Nogués X, Pols HAP, Renner W, Rivadeneira F, van Schoor NM, Ioannidis JPA, Uitterlinden AG, APOSS investigators, DOPS investigators, EPOS investigators, EPOLOS investigators, ERGO investigators, FAMOS investigators, FOS/Framingham investigators, LASA investigators, Mr OS investigators; for the GENOMOS Study. Large-scale analysis of association between polymorphisms in the LRP-5 and -6 genes and osteoporosis: The GENOMOS Study. *JAMA.* 2008;299(11): 1277-90.
  18. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9:356-69.
  19. Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet.* 2008;40:575-83.
  20. Lettre G, Jackson AU, Gieger C, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet.* 2008;40:584-91.
  21. Gudbjartsson DF, Walters GB, Thorleifsson G, et al. Many sequence variants affecting diversity of adult human height. *Nat Genet.* 2008;40:609-15.
  22. Kiel DP, Demissie S, Dupuis J, Lunetta KL, Murabito JM, Karasik D. Genome-wide association with bone mass and geometry in the Framingham Heart Study. *BMC Med Genet.* 2007 Sep 19;8 Suppl 1:S14.
  23. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, Andrew T, Falchi M, Gwilliam R, Ahmadi KR, Valdes AM, Arp P, Whittaker P, Verlaan DJ, Jhamai M, Kumanduri V, Moorhouse M, van Meurs JB, Hofman A, Pols HA, Hart D, Zhai G, Kato BS, Mullin BH, Zhang F, Deloukas P, Uitterlinden AG, Specator TD. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet.* 2008 May 3; 371(9623):1505-12.
  24. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med.* 2008;358:2355-65.
  25. Cohen JC, Kiss RS, Pertsemlidis A, et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science.* 2004; 305:869-72.