Hylonome

Review Article

Identifying genetic risk factors for osteoporosis

A.G. Uitterlinden[1,2,3], J.B.J. van Meurs[1], F. Rivadeneira[1,2], H.A.P. Pols[1,2]

[1]Department of Internal Medicine, [2]Department of Epidemiology and Biostatistics,
[3]Department of Clinical Chemistry, Erasmus Medical Centre, Rotterdam, The Netherlands

Abstract

Over the past decades epidemiological research of so-called "complex" diseases, i.e., common age-related disorders such as cancer, cardiovascular disease, diabetes, and osteoporosis, has identified anthropometric, behavioural, and serum parameters as risk factors. Recently, genetic polymorphisms have gained considerable interest, propelled by the Human Genome Project and its sequela that have identified most genes and uncovered a plethora of polymorphic variants, some of which embody the genetic risk factors. 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 small effect size for each individual risk polymorphism, this is mostly due to low statistical power and limitations of analytical methods. Genome-wide scanning approaches can be used to find the responsible genes. It is by now clear that linkage analysis is not suitable for this, but genome-wide association analysis has much better possibilities, as is illustrated by successful identification of risk alleles for several complex diseases. Candidate gene 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, we need large (global) collaborative studies using standardized methodology and definitions, to quantify by meta-analysis the subtle effects of the responsible gene variants.

Keywords: Genetic, Haplotypes, Polymorphism, Meta-analysis

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)[2-4]. For BMD the heritability has been estimated to be high: 50-80%[2-4]. 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 widespread availability of devices to measure BMD. 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 end point 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 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 years) this would require control subjects of

Type of approach

"Top-down"

* **Genome search by** <u>linkage</u> **analysis**
- Pedigrees
- Sib-pairs
- Human, mouse

* **Whole-Genome-Association**

Genotype haplotype tagging SNPs

"Bottom-up"

* <u>Association</u> **analyses of candidate**
  gene polymorphisms

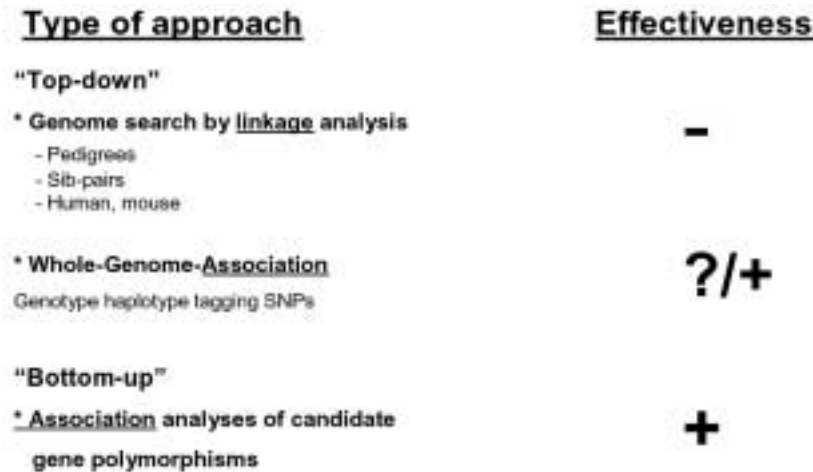## Effectiveness

–

?/+

+

Figure 1. Methods of dissection of a complex disease/trait to identify the "risk" alleles of candidate genes that explain the genetic contribution to the trait.

90-100 years. It is questionable whether such healthy survivors are proper controls for fracture cases.

Andrew et al.[5] recently studied 6,570 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.

While it might be difficult to demonstrate that fracture risk is heritable, one can also argue that it follows from simple logical 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 osteo-

porosis 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 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 analyze the contribution of environmental factors and their interplay with genetic factors.

Genome-wide approaches to find the genes: linkage vs. association

To identify and disentangle the genetic factors underlying the risk for osteoporosis basically two approaches can be applied: the "Genome-wide analysis" approach, which is free of any hypothesis regarding which gene(s) are involved, and the "Candidate Gene" approach, which requires an *a priori* hypothesis as to which gene is involved (see Figure 1).

The idea is that genome-wide approaches are preferred because they will identify true and major genetic effects (the "low hanging fruit") while candidate gene approaches are prone to heavy bias and thus not to be pursued with great priority. While genome-wide approaches have of course great

appeal the results obtained so far have been (very) disappointing. This is mainly due to methodological limitations of linkage analysis as indicated below. However, novel and much better possibilities are now offered by the so-called genome-wide association strategies, which are also discussed below.

In the meanwhile, several of the more classical candidate genes have already been analyzed in the past to establish what the role of polymorphisms in these genes is in conferring osteoporosis risk. Even in the absence of such genome-wide scans, this is a valid approach to determine their contribution to the genetic risk for osteoporosis. Indeed, candidate gene analyses have identified genetic risk factors for osteoporosis, albeit of modest effect size. In addition, the outcome of any genome-wide analysis is the subsequent study of a particular candidate gene so this approach will be among us for the coming years in any case.

A long-time paradigm in human genetics is to scan the whole genome, free of any hypothesis and to look for co-segregation of the disease gene(s) with a DNA marker, so-called linkage. With the advent of the Human Genome Project a multitude of single nucleotide polymorphism (SNP) markers has now become available which coupled with the discovery of so-called "haplotype blocks" in the genome, has led to a new genome-wide approach based on the old-fashioned association analysis in cases and controls: the genome-wide association analysis.

Thus, two genome-wide approaches can now be distinguished: those based on linkage analysis and those based on association analysis.

Genome-wide linkage analysis

Finding the responsible gene for <u>monogenic</u> disorders (caused by rare mutations in a single gene) is a straightforward routine exercise for specialized laboratories. This is based on linkage analysis in pedigrees (in which the disease is segregating according to Mendelian laws) whereby a standardized set of hundreds of well-characterised DNA markers are analyzed for co-segregation with a phenotypic end point.

An example of such an approach is the identification of LRP5 gene mutations being responsible for Osteoporosis Pseudoglioma as well as for a trait called High Bone Mass[6,7]. Many other examples in the bone field exist and so, for single gene diseases this approach works very well to identify "bone genes". However, the complex (and non-Mendelian) character of osteoporosis makes it quite resistant to the methods of analysis which in the past decades have worked so well for the monogenic diseases. Therefore, different and often more cumbersome approaches have to be applied which can be broadly defined as top-down and bottom-up approaches (see, e.g.,[8]).

In *top-down* approaches whole genome searches are performed which indicate which chromosomal areas might contain osteoporosis genes. For linkage analysis hundreds of relatives (sibs, pedigrees, etc.) are genotyped for hundreds of DNA markers (mostly micro-satellites but now also thou-

sands of SNPs can be used, e.g., the Affymetrix 10K SNP chip) evenly spread over the genome. Most genome searches focus on humans, although several mice genome searches have also been performed. Such genome searches are based on the assumption that relatives who share a certain phenotype will also share one or more chromosomal areas identical-by-descent containing one or more gene variants causing (to a certain extent) the phenotype of interest (e.g., low BMD). The gene is then said to be linked with the DNA marker used to "flag" a certain chromosomal region, but this area is usually several million base pairs big. Upon positive linkage, subsequent research will then have to analyze dozens of genes in the chromosomal area to determine which one is (are) the one(s) involved in bone metabolism, and then identify the particular sequence variant in that (those) gene(s) giving rise to (aspects of) osteoporosis.

Up to the stage of finding linkage to one or more chromosomal regions this approach has been quite successful in the past and linkage results from several genome searches have indeed been published (e.g.,[9]). However, so far the subsequent step to identify the gene variant causing the linkage peak has been difficult and has not resulted in many – if any – "osteoporosis risk genes". Why is that?

1. **Weak statistical linkage evidence.** It has proven difficult to find statistically significant linkage with LOD scores above the 3.7 for genome-wide significance. Typically only "suggestive" linkage is found with LOD scores of 1-3. This indicates that there are not a few major genes for osteoporosis but rather many subtle genes.
2. **Lack of replication of linkage peaks.** There is hardly any single region which has been identified convincingly (and statistically significant) by more than one genome search. Replication has also proven difficult because the families/pedigrees/sib pairs between different genome searches have differences in ethnicity, environmental factors, gender, age, etc.
3. **Lack of power to find a gene.** The effects per polymorphism are too weak to be detected with the typical number of sib pairs available. It has therefore been difficult to go beyond "linkage" and to demonstrate that a certain gene variant is <u>causing</u> the linkage peak observed in the genome search. Chromosomal regions showing linkage are typically 1-10 million base pairs wide containing dozens of candidate genes. In these candidate genes hundreds of polymorphisms occur organized in linkage disequilibrium blocks of 10-20 kb, making it virtually impossible (by statistical genetics alone) to pinpoint the causative variant using the linkage design.
4. **Choice of end point and case-control design.** Most genome searches have focussed on BMD, as an end point. BMD however, explains only a part of osteoporotic fracture risk. It is difficult to "switch" between major outcomes during or after the study because the families/sib-pairs are selected on the basis of such an end point and – thus – all inherent flaws of the "case-control study design" apply here.

It is also noteworthy that all linkage scans have identified few linkage peaks, suggesting few genes to explain the genetics of osteoporosis. It is by now widely assumed however, that many, maybe hundreds of gene variants are implicated. Given the very limited population attributable risk of the claimed genes identified so far by this approach, this probably also reflects the very limited power of genome-wide linkage scans.

Recently, an example of the identification of an allegedly "major osteoporosis gene" through a genome search was published. It was the identification of BMP2 (20p12.3) as a risk factor for osteoporotic fracture by analysis of Icelandic pedigrees and a Danish cohort by the company Decode from Iceland[10]. Although one might interpret this as proof of the success of the genome search approach, several notions would preclude that:

1. BMP2 was already well-known as an important gene for bone metabolism for several decades and as such represents a good candidate gene. So far, however, nobody had looked for polymorphisms in this gene in relation to osteoporosis.
2. The effect size of the BMP2 gene variants on fracture risk in the two samples (Icelandic and Danish) is modest and in line with what has been found for other candidate genes. It is rather premature to call this a "major" risk gene for osteoporosis and, especially given the low population frequency of a risk allele (37Ser, f=10%) the evidence presented would not support this view.
3. Only one "major" linkage peak was observed. Does this imply this is the only major osteoporosis gene? Clearly not. Perhaps then only so in Iceland and/or in Denmark?
4. Enormous effort (in terms of money and people) went into this research and the cost-benefit balance is very unfavourable.
5. The study identified a low frequent amino acid variant (Ser37Ala) in the gene as being responsible for the effect but did not provide functional evidence. In addition, haplotypes were constructed that associated with osteoporosis (defined in many different ways). The very large haplotypes (up to 200 kb!), however, are ill-defined and encompass dozens of so far unknown polymorphisms and this haplotype association could not be replicated in another, Danish sample.

While BMP2 was heralded as the first osteoporosis gene to be identified by way of a genome linkage search, more research is needed including multiple association studies in different populations and meta-analysis, to establish what the contribution of variants in this gene are to osteoporosis risk. So, the final proof of its involvement must come from candidate gene association studies and it is far from certain this will be a major important risk gene.

Genome-wide association analysis

Since we now know that the human genome has a haplotype block structure, this has opened a novel approach to search the genome for genetic markers of disease: the genome-wide association (GWA) analysis (reviewed in[11,12]). In this approach many hundreds of thousand SNPs are analyzed in sets of (usually) a few hundred unrelated cases and unrelated controls. The exact set of SNPs depends somewhat on the techniques used: Affymetrix (www.affymetrix.com) currently offers chips with SNPs based on XbaI/HindIII RFLPs that are more or less evenly spread across the genome. Densities have increased over time from 10.000 (10k), to 100K and now 500K. Perlegen (www.perlegen.com) offers as a service their in-house developed Affymetrix chip technology containing > 1 million haplotype tagging SNPs[13]. Illumina (www.illumina.com) is using glass arrays which are spotted at high density with very selected SNPs, such as coding SNPs (100k), and will shortly offer such arrays with haplotype tagging SNPs.

The first successful use of a genome-wide association analysis using such high numbers of SNPs, was reported by Ozaki et al.[14] who, by means of a large-scale, case (n=94)-control (n=658) association study using 92,788 gene-based SNPs, identified significant associations between myocardial infarction and 2 SNPs in LTA (encoding lymphotoxin-alpha): one SNP changed an amino-acid residue from threonine to asparagine (Thr26Asn) while another SNP in intron 1 influenced transcription level of LTA.

More recently, Klein et al.[15] reported a GWA study of 96 cases and 50 controls for polymorphisms associated with age-related macular degeneration (AMD), a major cause of blindness in the elderly. Among 116,204 single-nucleotide polymorphisms genotyped (using Affymetrix chips), a tyrosine-histidine change at amino acid position 402 (T402H) in the complement factor H gene (CFH) was strongly associated with AMD. This polymorphism is in a region of CFH that binds heparin and C-reactive protein. The CFH gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

The relatively low density of SNPs used in these studies in combination with the limited genetic complexity (of AMD in particular) could explain why only a few associated regions were observed in these studies, while one would expect (many) more gene regions to show up. Indeed, very recent press-releases from Affymetrix (www.microarraybulletin.com; April 2005) indicated similar successes for multiple sclerosis (MS; Cohen et al.), Graft vs. Host disease (GVHD; Ogawa et al.), and cardiovascular disease (CVD; Salonen et al.), where more cases/controls were analyzed (1,800 for MS, 2,000 for GVHD, and with higher density SNP arrays (100-500K) and more gene regions were found to be associated with the disease (80 for multiple sclerosis, 400 for CVD). However, we will have to await the detailed report in the scientific literature of these studies to know the exact methods used and results found. Nevertheless, these results from GWA studies show the great potential they have for elucidating complex diseases and it will only be a matter of time before similar approaches will be reported for osteoporosis.

The only current limitation might be that they are quite expensive: they cost roughly 1,000 euros per DNA sample.

Yet, the same statistical requirements apply as in a given case-control or population-based study of a candidate gene polymorphism. This means that for truly complex traits a minimal of several hundred cases/controls have to be studied, thus requiring up to a million euros per GWA study. For less complex traits, e.g., those with a few major genes such as might be the case for AMD, this might be less.

It should be noted that the size of the haplotype blocks/chromosomal regions identified through the GWA approach is much much smaller (10-50 kb) than that which is usually found in genome-wide linkage analyses (which is 1-10 million (!) base pairs). This offers major advantages for subsequent research. Yet, even when one or more such haplotype blocks are found associated, these blocks need further scrutiny to identify the one or more polymorphisms driving the association and functionality has to be established. So, this approach (also) ends up with a candidate gene analysis. After that, the associations have to be replicated in other populations and finally, meta-analysis has to be used to quantify the effect size (see later in article).

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 (see Figure 2). 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.

Subsequently, in the candidate gene frequently occurring sequences variants (polymorphisms) have to be identified which supposedly lead to subtle differences in the 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.

Sequence analysis of a "candidate" osteoporosis gene in a number of different individuals will identify sequence variants, but also several databases are now available which contain this information (e.g., NCBI, Celera, HapMap, and several more specialized databases). Some DNA sequence variations will be just polymorphic (anonymous polymorphisms) while others will have consequences for the level and/or activity of the protein encoded (functional polymorphisms).
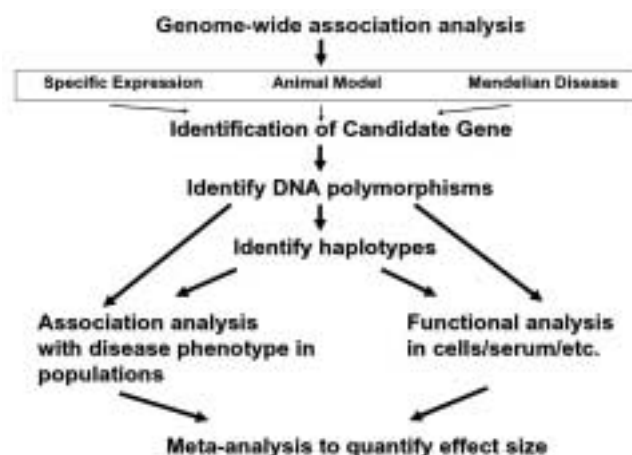

Figure 2. A schematic flow-diagram depicts the different steps in a candidate gene polymorphism analysis. On top genome-wide association analysis is indicated that will identify multiple areas across the genome as LD blocks within candidate genes. This is used in concordance with biological evidence based on 3 independent sources, to implicate a gene in the disease of interest.
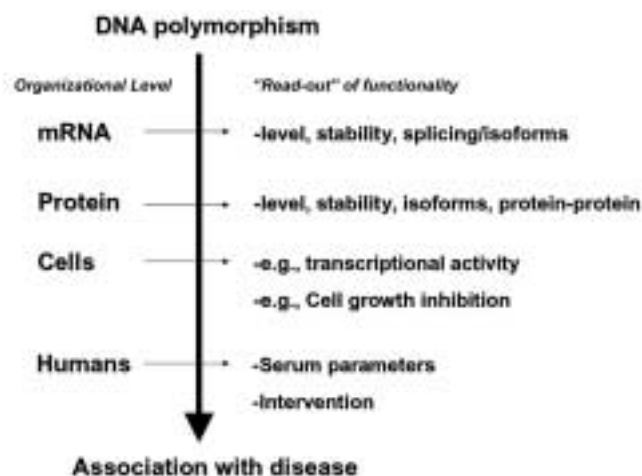

Figure 3. Depiction of how "functional" DNA polymorphisms might affect physiological processes at different levels of organization, that ultimately result in an association that is seen after many years (for age-related disorders this can be 80 years) of "exposure" to the risk factor.

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.

Clearly, it depends on the gene as to how many and what kind of polymorphisms will be present in the population. Some genes will have been, for example, under more evolutionary pressure and will not display much variation. Other

GENE NAME	GENE SYMBOL	CHROMOSOMAL LOCATION
Matrix protein molecules		
Osteocalcin	BGLAP	1q25-q31
α2HS Glycoprotein	AHSG	3q27
Osteopontin	SPP1	4q21-q25
Osteonectin	SPOCK	5q31.3-q32
Collagen type Iα1*	COLIA1	17q21.3-q22.1
Collagen type Iα2	COLIA2	7q22.1
Matrix associated enzymes		
Cathepsin K	CTSK	1q21
Alkaline Phosphatase	ALPL	1p36.1-p34
Carbonic Anhydrase II	CA2	8q22
Matrix metalloproteinase 3	MMP3	11q22.3
Lysyl Oxidase	LOX	5q23.3-q31.2
Lysine Hydroxylase	PLOD	1p36
Lysine Hydroxylase 2	PLOD2	3q23-q24
Lysine Hydroxylase 3	PLOD3	7q36
Calciotropic (Steroid) Hormone/Receptors/Enzymes		
Estrogen Receptor a*	ESR1	6q25.1
Estrogen Receptor β	ESR2	14q23
Aromatase	CYP19	15q21.1
Androgen Receptor	AR	Xq11
Glucocorticoid Receptor	GR/NR3C1	5q31
Vitamin D Receptor*	VDR	12q13
Vitamin D binding protein	DBP/GC	19q13.3
B3-adrenergic Receptor	ADRB3	8p12-p11.2
Peroxisome proliferator-activated receptor-gamma	PPARG	3p25
Calcium Sensing Receptor	CASR	3q21-q24
Calcitonin receptor	CALCR	7q21.3
Parathyroid Hormone	PTH	11p15.3-p15.1
PTH receptor	PTHR1	3p22-p21.1
Epidermal Growth Factor	EGF	4q25
Gonadotropin releasing hormone 1	GNRH1	8p21-p11.2
Gonadotropin releasing hormone receptor	GNRHR	4q21.2
Luteinizing Hormone beta peptide	LHB	19q13.32
LH-choriogonadotropin receptor	LHCGR	2p21
Growth Factors/Cytokines/Receptors		
Interleukin-1β Receptor Antagonist	IL-1RN	2q14.2
Interleukin-4	IL-4	5q31.3
Interleukin-6	IL-6	7p21
Transforming Growth Factor β1*	TGFB1	19q13.2
Transforming Growth Factor β2	TGFB2	1q41
Growth Hormone	GH1	17q22-q24
Growth Hormone Receptor	GHR	5p13-p12
Insulin-like Growth Factor I	IGFI	12q22-q23
Insulin-like Growth Factor I Receptor	IGF1R	15q25-q26
Insulin-like Growth Factor 2	IGF2	11p15.5
Insulin-like Growth Factor II Receptor	IGF2R	6q26
IGF-binding protein 3	IGFBP3	7p14-p12
Tumour Necrosis Factor α	TNF	6p21.3
TNF receptor superfamily/1β	TNFRG5	1p36.3-p36.2
Bone Morphogenetic Protein 2	BMP2	20p12
Bone Morphogenetic Protein 3	BMP3	4p14-q21
Sclerostin	SOST	17q12-q21
Osteoprotegerin	OPG/TNFRSF11B	8q24
RANK 18q22.1	RANK/TNFRSF11A	
RANL-ligand/OPG-ligand	RANKL/TNFSF11	13q14
Wnt-signalling pathway		
Low density lipoprotein receptor-related protein 5*	LRP5	11q12
Homocystein pathway		
Methylene TetraHydroFolate Reductase	MTHFR	1p36.3
Cystathionine beta-synthase	CBS	21q22.3
Methionine synthase reductase	MTRR	5p15.3-p15.2
Methyltetrahydrofolate-homocysteine s-methyltransferase	MTR	1q43
Thymidylate synthetase	TYMS	18p11.32
Miscellaneous		
Major Histocompatibility Complex	MHC/HLA	6p21.3
Apolipoprotein E	APOE	19q13.2

*Genes, polymorphisms which have and will be analysed within the GENOMOS consortium

Table 1. Selected osteoporosis candidate genes.

genes, however, might be part of a pathway with sufficient redundancy to allow for more genetic variation to occur.

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. Yet it is also important to realize in this respect that it is of uncertain value to test func-tionality of a certain polymorphism in the absence of an association at the population level.

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 (see Figure 3) and depends on the type of protein analyzed, 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 end point 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[16].

Out of the three lines of evidence mentioned above, numerous candidate genes have emerged and Table 1 lists only a few of these. These include "classical" candidate genes such as collagen type I, the vitamin D receptor, and the estrogen receptors. Yet, also recently identified "bone" genes, such as LRP5, can become candidate genes because their involvement in bone biology has only recently become known. These studies have identified LRP5 as a candidate gene, but of course not established its role as a genetic risk factor for osteoporosis.

With this plethora of candidate genes it is difficult to decide where to start. Initially this happened somewhat randomly but current choices are guided by increasing insights in the metabolic pathways in which the genes play a pivotal role. For example, the identification of LRP5 as a candidate gene has put the complete Wnt-signalling pathway on the map as a target. It can therefore now be expected that multiple genes from this pathway will be tested as candidate genes for osteoporosis.

The current focus in genetic studies of osteoporosis is quite strongly on common variants which are expected to explain a substantial portion of population variance, simply due to their frequency in the population (10-50%). However, also more rare variants (1-10% or even less frequent) can contribute to population variance with stronger effects, and perhaps can play an important role in certain populations but not in others.

An example of this was described by Cohen et al.[17] 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 (>95th 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 param-

eters, and even fracture risk in the general population.

Thus, when we compare the genome-wide linkage approach to the candidate gene approach, the latter approach is now clearly the more promising one[11-13,18]. Genome-wide linkage searches are not designed and not statistically powered to detect the many subtle gene effects which underlie osteoporosis following the "common variant-common disease" hypothesis. The genome-wide association analysis seems to be a better alternative, but has not been used so far in osteoporosis and has not identified any risk genes yet. Thus, the current way forward is to simply test individual candidate genes to establish what their contribution is to osteoporosis risk. Once that has been established, the interaction or multiplicative effects of several genes will be analyzed and, finally, gene-environment interactions can be studied.

Haplotypes

From re-sequencing studies for the HapMap project (www.hapmap.org) it has become evident that, on average, 1 out of every 300 base pairs is varying in the population. Given an average size of 100 kb of a gene this means there are hundreds of polymorphisms in a given gene. Thus, candidate gene analyses will have to focus on which of the many variant nucleotides are the ones that actually matter. That is, which sequence variation is functionally relevant by changing expression levels, changing codons, etc. Given the average size of a gene and the relatively young age of human populations it can be predicted that several sequence variations "that matter" will co-exist in a gene in a given number of subjects from a study population. A major challenge of fundamental research will therefore be to unravel the functionality of these variations and how they interact with each other within the gene.

More recently, it has become clear that these neighbouring polymorphisms are not independent from each other in genetic terms, that is to say they tend to "travel together" in so-called haplotypes. Haplotypes are strings of coupled or linked variants which occur, on average, over a distance of 10-30 kb in the human genome. With polymorphisms occurring roughly in 1 out of 300 base pairs this means there will be dozens of polymorphisms within these "haplotype blocks".

An important aspect of association analyses in this respect is then to establish which common haplotype alleles are occurring in the candidate gene, which has two important practical consequences:

1. If association is found of a particular allele of an individual polymorphism with a certain phenotype/disease, this can also be explained by an adjacent polymorphism within the haplotype block. Thus, one can never be sure what causes the association until the haplotype structure at that position within the gene has been resolved.
2. When, for example, 20 polymorphisms are located within a haplotype block only a fraction (typically only 30%) has to be genotyped to identify the haplotype alleles. This saves
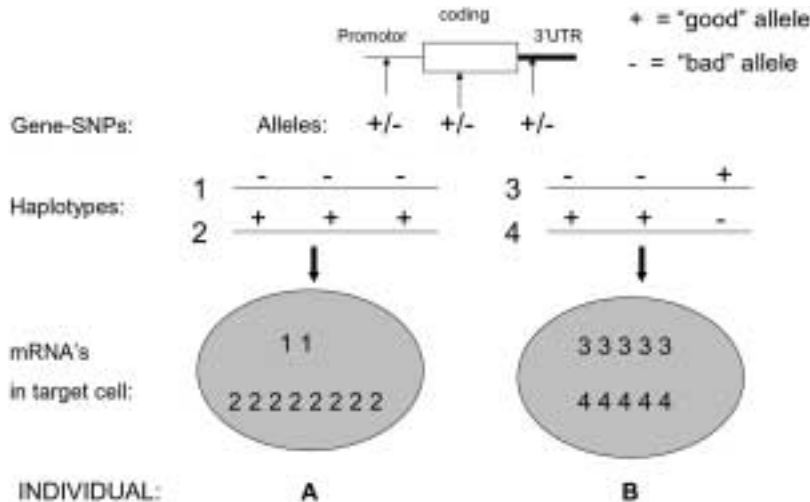
Figure 4. Hypothetical example of the importance of gene-wide genotype combinations. Three adjacent SNPs in different parts of a gene are shown for two individuals (A and B indicated at the bottom). The subjects A and B have identical genotypes, i.e., they are both heterozygous for all three SNPs. However, they have different allele combinations on the same chromosome (numbered 1-4): 1+2 for subject A and 3+4 for subject B.

The promoter area regulates production of mRNA while the 3'UTR is involved in degradation of mRNA and their interaction/combined effects regulates the net availability of the mRNA for translation into the protein.

In this case the example is shown for a promoter polymorphism which has two alleles + and -, of which the + allele is the high producer variant in certain target cells. Of the two different 3'UTR variants + and -, the + is the more stable 3'UTR resulting in more mRNA being maintained. Hence, a "good" promoter allele and a "good" 3'UTR allele on the same chromosome, result in more protein being produced. The protein itself can occur in two variants: a less active "risk" form (-) and a more active form (+), and both A and B are again heterozygous for this polymorphism.

The combined result of the particular allele combinations is that individual A has less of the "risk" protein than individual B in the target cell. This could not have been predicted by analyzing single SNPs and/or only looking at genotypes of individual SNPs, but is only evident upon analysis of the gene-wide genotype combinations.

on time and money to perform the association analyses while obtaining maximal information relevant for point 1.

Once such haplotype blocks have been identified, it becomes important how they are organized in the gene of interest. A typical gene can have one or several haplotype blocks covering the promoter region, another block covering the coding region and yet another block covering regulatory regions 3' of the gene. For the functioning of a complete gene in a given cell of a given subject, it is then important to know which combination of haplotype alleles is present in that subject. In Figure 4 a hypothetical example is given of the functional relevance of gene-wide combinations of genotypes (based on single SNPs or on haplotypes). The figure describes the situation when two subjects have identical genotypes for 3 adjacent polymorphic sites when analyzed independently. Yet, they differ in their combination of alleles on one chromosome, and this will result in different effects at the cellular level. This example illustrates that the effects of single polymorphisms might be difficult to interpret when ignoring the polymorphisms in the rest of the haplotype block and the other haplotype blocks in the gene.

Meta-analyses

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[19] and that large studies are more conservative in these estimates and should preferably be used[20]. 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[21].

In the field of osteoporosis, 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, and design details are described in the first meta-analysis of individual-level data on the ESR1 gene[22].

The GENOMOS meta-analysis of three polymorphisms in the ESR1 gene (intron 1 polymorphisms XbaI [dbSNP: rs9340799] and PvuII [dbSNP: rs2234693] 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[22].

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 cannot, for example, assess all genetic diversity across a gene.

While fracture has been debated as an end point in genetics of osteoporosis studies, this was chosen in the GENOMOS study because it is clinically the most relevant end point. Statistical power of the GENOMOS study to detect genetic effects on fracture risk is high with >5,000 fractures.

With such a diverse set of populations included in the GENOMOS study, possible population stratification could be a problem. This is not likely, however, because GENOMOS involves almost exclusively white Caucasians, who in addition come from very stable populations (with little immigration/emigration). Indeed, so far, the tested allele frequencies for ESR1[22] and COLIA1 (Ralston et al., manuscript submitted) are remarkably similar between populations supporting the absence of major population stratification.

Importantly, some functional SNPs can show similar effects across different ethnic groups in spite of different genetic backgrounds of the ethnic groups. In this respect it has recently been demonstrated that genetic markers for proposed gene-disease associations can vary in frequency across populations, but their biological impact on the risk for common diseases may usually be consistent across traditional 'racial' boundaries[16]. Yet, it is also conceivable that some genetic variants will have particular "local" effects, either due to particular environmental factors and/or due to differences in genetic background. Such factors could mask or enhance the effect of the particular polymorphism of interest.

Thus, such a meta-analysis approach will identify individual genetic risk factors but it will probably also be instrumental in estimating presence and effect size of genetic interactions (gene-gene) and gene-environment interactions. This approach will be followed for genes in a certain pathway, for which we know that interaction is likely, and can be extended to explore unexpected interactions. However, even with large studies of, e.g., 20,000 subjects, this might be difficult to convincingly demonstrate. This stresses the need for even larger studies.

In any case, performing meta-analyses and establishing functionality of polymorphisms should therefore be a major requirement before genetic polymorphisms can be considered for use in clinical practice.

Humans or mice

Several of the approaches discussed above have also been attempted in mice as a model for humans. Especially, transgenic and knock-out mice have provided very interesting clues regarding bone biology and, thus, have been a source of candidate genes to pursue in human studies of genetic variation contributing to risk of osteoporosis. Yet, the obvious drawback of this approach is that humans are not mice and, thus, biology can be very different, and that – in the end – we always have to turn to analyzing humans. Indeed, there are examples of knock-out mouse models that did not result in a clear bone phenotype whereas the human Mendelian counterpart did result in clear bone phenotypes. For example, the carbonic anhydrase II (CAII) null mice do not show the prominent osteopetrotic phenotype that is seen in human Mendelian CAII mutants[23]. In addition, such mice approaches show us what genes are important in bone biology, but they do not tell us what genes have relevant functional genetic variation in the human population that contribute to osteoporosis.

In this respect, it could be more informative to analyze different mouse strains for genetic differences that contribute to variation in bone phenotypes in mice. With respect to genome-wide linkage analysis, again, many examples of linkage peaks have been reported for the linkage approaches in mice, but – again – so far very few actual genes have been identified. Only recently, in an elegant combination of approaches Klein et al.[24] identified the lipoxygenase gene Alox15 as a negative regulator of peak bone mineral density in mice. Cross-breeding experiments with Alox15 knock-out mice confirmed that 12/15-lipoxygenase plays a role in skeletal development whereas pharmacologic inhibitors of Alox15 improved bone density and strength in two rodent models of osteoporosis. In humans, however, the situation is somewhat unclear to which of the three Alox genes is important in bone metabolism.

Taken together and given the amount of effort and time involved, and the substantial progress in knowledge of the human genome and its variation, it remains questionable whether this approach will deliver many osteoporosis risk gene variants.

Summary

So, in summary, if people were to embark on an association study of a candidate gene to identify genetic markers for

osteoporosis, what would be the crucial issues to address? A few suggestions:

Take a large population. Bigger is better to make your initial observations statistically robust.

Identify proper end points upfront. Fractures are clinically the most relevant but you need substantial numbers to make your findings statistically robust. BMD is only one of the risk factors but it is a continuous trait and gives more statistical power. Population-based studies have the advantage of being able to switch phenotypes during analysis very easily, for case-controls this possibility is very limited.

Cover all relevant genetic variation within the gene. Focus on functionally relevant variants within a gene. A clear-cut functional variant can be analyzed in isolation, ignoring the rest of the genetic variation in the gene. However, determine the haplotype structure to understand how the complete gene is functioning.

P-values: rather seek replication of your finding. Simple adjustment for multiple testing is regarded as not appropriate (where to start and stop counting?). Rather, formulate a proper *a priori* hypothesis and seek replication(s) of the observed association in similar populations.

Perform a meta-analysis to quantify effect size and assess heterogeneity. Join consortia with your population and datasets to standardize genotype and phenotype definition and estimate effect size of polymorphisms, preferably by prospective meta-analysis rather than meta-analysis of published data.

Although still in its infancy with respect to having clinical implications, the field of genetics of osteoporosis (or any complex disease for that matter) is expected to eventually find applications in two main areas:

1. Prediction of response-to-treatment. Polymorphisms in, e.g., drug-metabolizing enzymes will result in different efficiencies with which drugs can exert their effect. The same holds true for receptors of hormones and growth factors, analogues of which are currently prescribed as treatment. Genotype analysis can identify those subjects expected to profit most from a particular treatment or exclude those subjects which will suffer more from side-effects (personalised medicine).
2. Identification of subjects-at-risk. Subjects carrying risk alleles are more likely to develop osteoporosis. Genotype analysis will allow taking preventive measures, targeted at the individual at an early stage.

So far only one polymorphism is currently being considered as an osteoporosis risk factor (the COLIA1 Sp1 polymorphism) and commercial parties have taken up interest in this genetic marker. However, its utility in clinical practice has to be considered with considerable caution. For example, analyses in different ethnic populations have shown it to be present mostly in Caucasian subjects[25]. Furthermore, interaction of this variant with another polymorphism (the VDR 3' variants) has been demonstrated[26]. This latter study

also highlights the complex and multi-genic nature of osteoporosis. It underlines the need to identify additional osteoporosis risk alleles to better understand how particular genetic markers are expressed and result in a phenotype.

Another spin-off of genetic research of osteoporosis is the discovery of new and/or unexpected genes and pathways to be involved in determining, e.g., BMD. A good example is the identification of the Wnt-signalling pathway to be involved in bone metabolism, through the analysis of LRP5. Such discoveries lead to new possibilities to develop drugs to treat osteoporosis. In addition, such genes become candidate osteoporosis risk genes and will be searched for polymorphisms. Risk alleles resulting from such analyses can then be added to the still growing list of osteoporosis gene variants.

Thus, in spite of complicating factors, genetic research will contribute to a further understanding of complex diseases, including osteoporosis. The identification of new genes or new roles of already known genes, will allow insights in mechanistic pathways which might help in designing therapeutic protocols. Finally, the description of genetic variation underlying phenotypic variation can be used, in concert with existing easy-to-assess risk factors, in prediction of risk for aspects of osteoporosis. In this respect, novel therapeutic protocols but also insights into gene-environment interactions allow for ways to further improve treatment of patients.

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. N 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, 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; 20:67-74.
6. Gong Y, Slee RB, Fukai N, Rawadi G, Roman-Roman S, Reginato AM, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. Cell 2001; 107:513-523.
7. Little RD, Carulli JP, Del Mastro RG, Dupuis J, Osborne M, Folz C, et al. A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high bone mass trait. Am J Hum Genet 2002; 70:11-19.

8. Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994; 265:2037-2048.

9. Ralston SH, Galwey N, MacKay I, Albagha OM, Cardon L, Compston JE, et al. Loci for regulation of bone mineral density in men and women identified by genome-wide linkage scan: the FAMOS study. Hum Mol Genet 2005; 14:943-951.

10. Styrkarsdottir U, Cazier JB, Kong A, Rolfsson O, Larsen H, Bjarnadottir E, et al. Linkage of osteoporosis to chromosome 20p12 and association to BMP2. PLoS Biol 2003; 1(3):E69.

11. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005; 6:95-108.

12. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005; 6:109-118.

13. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science 2005; 307:1072-1079.

14. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 2002; 32:650-654.

15. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, Sangiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science 2005; 308:385-389.

16. Ioannidis JP, Ntzani EE, Trikalinos TA. 'Racial' differences in genetic effects for complex diseases. Nat Genet 2004; 36:1312-1318.

17. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 2004; 305:869-872.

18. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273:1516-1517.

19. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29:306-309.

20. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. Lancet 2003; 361:567-571.

21. 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; 33:177-182.

22. Ioannidis JP, Ralston SH, Bennett ST, Brandi ML, Grinberg D, Karassa FB, et al.; GENOMOS Study. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. JAMA 2004; 292:2105-2114.

23. Lewis SE, Erickson RP, Barnett LB, Venta PJ, Tashian RE. N-ethyl-N-nitrosourea-induced null mutation at the mouse Car-2 locus: an animal model for human carbonic anhydrase II deficiency syndrome. Proc Natl Acad Sci USA 1988; 85:1962-1966.

24. Klein RF, Allard J, Avnur Z, Nikolcheva T, Rotstein D, Carlos AS, Shea M, Waters RV, Belknap JK, Peltz G, Orwoll ES. Regulation of bone mass in mice by the lipoxygenase gene Alox15. Science 2004; 303:229-232.

25. Beavan S, Prentice A, Bakary D, Yan L, Cooper C, Ralston SH. Polymorphism of the collagen type Iα1 gene and ethnic differences in hip fractures rates. N Engl J Med 1998; 339:351-352.

26. Uitterlinden AG, Weel AEAM, Burger H, Fang Y, van Duijn CM, Hofman A, van Leeuwen JPTM, Pols HAP. Interaction between the vitamin D receptor gene and collagen type Iα1 gene in susceptibility for fracture. J Bone Miner Res 2001; 16:379-385.