

ORIGINAL ARTICLE

A survey of haplotype variants at several disease candidate genes: the importance of rare variants for complex diseases

P-Y Liu, Y-Y Zhang, Y Lu, J-R Long, H Shen, Lan-J Zhao, F-H Xu, P Xiao, D-H Xiong, Y-J Liu, R R Recker, H-W Deng

J Med Genet 2005;42:221–227. doi: 10.1136/jmg.2004.024752

See end of article for authors' affiliations

Correspondence to:
Dr H-W Deng,
Osteoporosis Research
Center, Creighton
University Medical Center,
601 N. 30th St., Suite
6787, Omaha, NE 68131,
USA; deng@creighton.eduReceived 30 June 2004
Revised version received
17 September 2004
Accepted for publication
4 October 2004**Background:** The haplotype based association method offers a powerful approach to complex disease gene mapping. In this method, a few common haplotypes that account for the vast majority of chromosomes in the populations are usually examined for association with disease phenotypes. This brings us to a critical question of whether rare haplotypes play an important role in influencing disease susceptibility and thus should not be ignored in the design and execution of association studies.**Methods:** To address this question we surveyed, in a large sample of 1873 white subjects, six candidate genes for osteoporosis (a common late onset bone disorder), which had 29 SNPs, an average marker density of 13 kb, and covered a total of 377 kb of the DNA sequence.**Results:** Our empirical data demonstrated that two rare haplotypes of the parathyroid hormone (PTH)/PTH related peptide receptor type 1 and vitamin D receptor genes (*PTH1R1* and *VDR*) with frequencies of 1.1% and 2.9%, respectively, had significant effects on osteoporosis phenotypes ($p=4.2 \times 10^{-6}$ and $p=1.6 \times 10^{-4}$, respectively). Large phenotypic differences (4.0–5.0%) were observed between carriers of these rare haplotypes and non-carriers. Carriers of the two rare haplotypes showed quantitatively continuous variation in the population and were derived from a wide spectrum rather than from one extreme tail of the population phenotype distribution.**Conclusions:** These findings indicate that rare haplotypes/variants are important for disease susceptibility and cannot be ignored in genetics studies of complex diseases. The study has profound implications for association studies and applications of the HapMap project.

Haplotype analyses have become increasingly important in genetic studies of human diseases. When multiple markers, often in linkage disequilibrium (LD), in a chromosomal region are studied to assess the association between this region and the study traits of interest, a statistical analysis based on haplotypes may often be more efficient than separate analyses of individual markers. This has been demonstrated both through empirical^{1,2} and simulation^{3–5} studies. Firstly, haplotype analyses take into account of a number of tightly linked markers, which are much more informative than individual markers.^{6,7} Secondly, haplotype analyses can identify unique chromosomal segments likely to harbour disease predisposing genes. The phenotypic effect of several mutations at different sites within a gene can depend on whether the mutations occur on the same chromosome (in *cis*, as a haplotype) or on opposite homologous chromosomes (in *trans*).^{1,8,9} These findings emphasise an important aspect of examining candidate genes by SNP haplotyping.

The human genome has been portrayed as a series of high LD regions with limited haplotype diversity.^{10,11} Several common haplotypes that can be captured by a few tagged SNPs usually account for a majority of genetic variation in the genomic regions or candidate genes.^{10–12} Such haplotype patterns observed in empirical studies have triggered the development of the International HapMap Project (<http://www.hapmap.org>), which aims to determine the common patterns of DNA sequence variation in the human genome.¹³ Focusing on these common haplotypes greatly facilitates LD based mapping analyses.¹² By comparing the frequency of haplotype variants in unrelated cases and controls and/or the disease phenotypic distribution among haplotype variants in the cohort samples of moderate size, genetic association

studies can identify specific disease predisposing common haplotypes. However, a critical question is whether rare haplotypes play an important role in influencing disease susceptibility and thus should not be ignored in design and execution of association studies.

Given the heightened interest in association studies using haplotypes, we aimed to assess the potential role of rare haplotypes influencing disease susceptibility. In the present study, we surveyed six candidate genes with 29 SNPs, which had an average marker density of 13 kb and covered a total of 377 kb DNA sequence, for osteoporosis (a common late onset bone disorder) in a large sample of 1873 white subjects. We systemically investigated haplotype variations at these candidate genes, haplotype effects on disease phenotypes, and phenotypic distribution of haplotype carriers in the sample.

METHODS

Study subjects

The subjects came from study to search for genes underlying the risk of osteoporosis being carried out in the Osteoporosis Research Center of Creighton University Medical Center. We recruited 405 nuclear families totaling 1873 subjects, including 740 parents, 744 daughters, and 389 sons with a mean (SD) family size of 4.62 (1.78). All the subjects were white, of European origin. Only healthy people were included, with the exclusion criteria being as detailed earlier.¹⁴ Briefly, patients with chronic diseases and conditions that might potentially

Abbreviations: BMD, bone mineral density; CD-CV, common diseases common variants; LD, linkage disequilibrium; PTH, parathyroid hormone; *PTH1R1*, parathyroid hormone receptor 1; *VDR*, vitamin D receptor

affect bone mass were excluded from the study. These diseases/conditions included chronic disorders involving vital organs (heart, lung, liver, kidney, and brain), serious metabolic diseases (including diabetes, hypoparathyroidism and hyperparathyroidism, hyperthyroidism), other skeletal diseases (including Paget's disease, osteogenesis imperfecta, and rheumatoid arthritis), chronic use of drugs affecting bone metabolism (corticosteroid therapy and anti-convulsant drugs), and malnutrition conditions (including chronic diarrhoea and chronic ulcerative colitis). For each study subject, we obtained the information on age, sex, medical, family and reproductive history, physical activity, alcohol use, and dietary and smoking habits. The study was approved by the institutional review board of Creighton University and informed consent documents were obtained for each subject.

Candidate genes

The chosen study candidate genes were apolipoprotein E (*APOE*), type I collagen $\alpha 1$ (*COL1A1*), oestrogen receptor- α (*ER- α*), parathyroid hormone (PTH)/PTH-related peptide receptor type 1 (*PTHRI*), transforming growth factor- $\beta 1$ (*TGF- $\beta 1$*) and vitamin D receptor (*VDR*). They are significant in terms of their functional roles in bone metabolisms, and/or their prominence in the genetic studies of osteoporosis.¹⁵ A total of 29 SNPs for these candidate genes were identified from the database dbSNP (<http://www.ncbi.nih.gov/SNP/>). These selected SNPs for the study were based on a comprehensive consideration of the criteria of: (a) functional relevance and importance (missense mutation, frameshift mutation etc.), (b) level of heterozygosity, (c) position in or around the gene, and (d) their use in previous genetic epidemiology studies. Detailed information about SNPs analyzed in this study is presented in table 1. These SNPs spanned a total of about 377 kb; the average physical distance between neighbouring markers was 13.0 kb. However, the pairwise LD (D') is highly variable among these candidate genes, ranging from 0.02 to 1.0 with an average of 0.48.¹⁶

SNP genotyping

Genomic DNA was extracted from whole blood using a commercial isolation kit (Gentra Systems, Minneapolis, MN, USA) following the procedure detailed in the kit. The genotyping procedure for all SNPs was similar, involving PCR and Invader assay (Third Wave Technology, Madison, WI, USA). PCR was performed in 10 μ l reaction volume with 35 ng genomic DNA, 0.2 mmol/l each of dCTP, dATP, dGTP and dTTP, 1 \times PCR buffer and 1.5 mmol/l $MgCl_2$, 0.4 μ mol/l each of the primers, and 0.35 U of *Taq* polymerase (ABI, Applied Biosystems, Foster City, CA, USA). The sequences of the PCR primers for all SNPs are presented in table 1. The following procedure was used on an ABI 9700 thermal cycler: 95°C for 5 minutes, 30 cycles of 94°C for 1 minute, 50°C for 1 minute, 72°C for 1 minute, and then 72°C for 5 minutes. After amplification, the product was diluted 1:20 in nuclease free water. Invader reaction was performed in a 7.5 μ l reaction volume, with 3.75 μ l diluted PCR product, 1.5 μ l probe mix, 1.75 μ l Cleavase FRET mix, and 0.5 μ l Cleavase enzyme/ $MgCl_2$ solution (Third Wave Technology). The reaction mix was overlaid by 15 μ l mineral oil and denatured at 95°C for 5 minutes, and then incubated at 63°C for 20 minutes in an ABI 9700 thermal cycler. After incubation, the fluorescence intensity for both colours (FAM and Red dyes) was measured using Cytofluor 4000 (ABI). The data were then used with Invader Analyzer software (Third Wave Technology), and the genotype for every sample was identified according to the ratio of the fluorescence intensity of the two dyes. PedCheck software¹⁷ was used to verify Mendelian inheritance of the alleles within each family and

the family relationships (<http://watson.hgen.pitt.edu/register/docs/pedcheck.html>).

Measurement

Bone mineral density (BMD) is the most important surrogate phenotype for osteoporosis, which is mainly characterised by low BMD (34). Femoral neck and total hip BMDs (g/cm^2) were measured by a Hologic 2000+ or a 4500 dual energy x ray absorptiometry (DXA) scanner (Hologic Inc., Bedford, MA, USA). Both machines were calibrated daily, and the coefficient of variability (CV) values of the DXA measurements at the femoral neck and total hip were 1.87% and 1.0% on the Hologic 2000+, and 1.98% and 1.4% on the Hologic 4500. Of the subjects, 92% were measured on the Hologic 4500. Data obtained from different machines were transformed to a compatible measurement,¹⁸ which has been shown to be highly reliable and accurate.¹⁹ Members of the same nuclear family were measured on the same scanner. At the same visit of the BMD scan, weight was measured using a calibrated balance beam scale, and height was measured using a calibrated stadiometer.

Statistical analyses

Haplotype pairs carried by each individual were inferred in nuclear families using Genehunter (version 2.1; <http://www.fhcr.org/labs/kruglyak/Downloads/index.html>).²⁰ Subjects with ambiguous haplotypes were excluded for further analyses. Specifically, we excluded 87, 59, 49, 70, 47 and 46 such subjects for haplotype analyses at the *APOE*, *COL1A1*, *ER- α* , *PTHRI*, *TGF- $\beta 1$* and *VDR* genes, respectively. To avoid reporting results based on very few individuals (~ 20), only those haplotypes with frequencies greater than 0.8% were analysed. In genetic analyses, the phenotypic values were tested for measured potentially important covariates and adjusted for those significant ones (age, sex, weight and height). These adjustments, in consideration of the correlation structures among subjects,²¹ were performed in the regression model described by George and Elston²² and implemented in SOLAR (<http://www.sfbr.org/sfbr/public/software/solar>).²³ The model residuals were calculated by subtracting the fitted values for covariate effects from the original phenotypic values and were used as phenotypic variables in haplotype analyses. This adjustment procedure is similar to that using the regular multiple regression model with BMD as a dependent variable and with age, sex, weight, and height as independent variables, apart from considering kinship of subjects within pedigrees. The heritabilities of BMD phenotypes and their standard errors were also estimated in the above covariate analysis. The normality of the phenotype data was examined by the Kolmogorov-Smirnov test implemented in SPSS 10.0 (SPSS Inc., Chicago, IL, USA). The mean difference of the studied phenotypes was performed by comparing individuals carrying a specific haplotype with non-carriers, using two sample two sided *t* tests. The empirical *p* values of the *t* tests were obtained by permutation tests. During the permutation procedure, the original phenotypic and genotypic data were reshuffled 10^7 times. The *t* test statistics were then computed on each dataset generated by reshuffling. Using Bonferroni correction for multiple tests, we obtained an empirical threshold $p \leq 5.4 \times 10^{-4}$ for single test, which achieves a global significance level of 0.05 for our study. The power for association studies using variance component models for TDT for sibship data²⁴ was obtained by the Genetic Power Calculator (<http://statgen.iop.kcl.ac.uk/gpc/>).²⁵

We also calculated the effective number of haplotypes and expected haplotype heterozygosity for each gene. The effective number of haplotypes, analogous to the effective number of alleles,²⁶ was calculated as:

Table 1 Information about the 29 SNPs in the six candidate genes for osteoporosis

Gene	dbSNP accession no.	Nucleotide*	Domain	Distance†	Frequency (%‡)	
APOE	ss12568587	G-C	5' UTR		35.7	F:TCCCCAGGAGCCGGTGA R:CCCCAAGCCCCGACCCC
	ss12568609	G-A	Intron 2	1277	39.9	F:CCTCAGGTGATCTGCCCGTTTC R:ACTCTGGGCTCAAGTGATCCTC
	ss12568607	T-C	Exon 4	1497	14.9	F:CGGGCACGGCTGTCCAA R:CGAGCATGGCCTGCACCTC
	ss12568612	C-T	Exon 4	138	8.7	F:GCTGCGTAAGCGGCTCC R:GCGGCCCTGTTCCACC
COL1A1	ss12568606	G-T	5' UTR		15.4	F:GCACCCTGCCCTAGACCAC R:CTAGTGCCAGCGACTGCA
	ss12568597	G-T	Intron 1	3543	18.8	F:CCAATCAGCCGCTCCCATTC R:CATCGGGAGGGCAGGCTC
	ss12568598	G-T	Exon 8	2365	0.0	F:GGAAGACTGGGATGAGGGCA R:GGCTCGCCAGGCTCACC
	ss12568584	G-A	Exon 45	9862	1.9	F:CTCAGCCTTCCCTGGCCAA R:AGGCGGAAGTCCATTGGCATC
	ss12568579	A-G	Exon 1		48.4	F:TTGAGCTGCGGACGGTTCA R:CGCCGTTTTCTGAGCCTC
ER- α	ss12568596	T-C	Intron 1	34 258	44.9	F:TGGGATTCCAGGCATGAACCAC R:TGGCGTCGATTATCTGAATTTGGCC
	ss12568619	G-A	Intron 3	66 110	25.7	F:CCAGAAACAAGTCATCTGCTATTGACA R:GTAACAAAAGGTTAACAATGGTTAGCCC
	ss12568618	G-C	Exon 4	36 077	21.8	F:ACAGCCTGGCCTTGTCCTC R:CAGGTTGGTCAGTAAGCCCATCA
	ss12568585	G-A	Intron 4	39 074	9.8	F:GATCAATGAAGTGGTCTTGAAAAACCAA R:GGTGACAAGCTGGAAATCTAAGCTTCA
	ss12568605	G-A	Intron 6	83 068	11.9	F:GGAACGGCCCTTGGAAATTTGAAA R:CTGCCTACAGAATACAGTCAGCCA
	ss12568617	G-A	Exon 8	32 431	20.3	F:TGCGATTCTGCAAAAGTATTACATCAC R:CAAGCAAATGAATGGCCACTCATAGAAA
	ss12568589	G-A	Intron 1		0.1	F:GACTTACATTAGGATTCAAGGTTACTGCCA R:GGGACGCAAGCCTGAGTCC
	ss12568592	A-G	Intron 2	4250	39.8	F:GCAGAACCCCTAAGGCTTGTC R:GGCGGACCCAGGATACA
	ss12568591	C-T	Intron 8	5436	37.4	F:CGAGCCTCAATTCAGGTGAATCTAACC R:CCCCCCCCAAGTGGAACA
	ss12568588	G-A	Intron 10	1912	39.7	F:CCTGAGCCCTTGGTTTTCTTTC R:GCTCCGGGAACAAAAGTGGATCA
TGF- β 1	ss12568613	C-T	5' UTR		31.1	F:GGGCCAGTTTCCCTATCTGTAAA
	ss12568603	+C/-C	Intron 4	12 354	2.1	F:CCACGCCCACTTATCTATCCC
	ss12568593	C-T	Exon 5	83	0.8	F:CAGGCTACAAGGCTCACCTGAA R:GGTTCACCTACCGGCCG
	ss12568602	C-T	Intron 5	9644	27.4	F:GGCTGTCTTAAGCATTGCGTGAATTA R:GTACAGCTGCCGACGC
VDR	ss12568583	G-A	5' UTR		28.1	F:CAGCATGCCTGTCTCAGC R:CCAGTACTGCCAGCTCCA
	ss12568581	C-T	Exon 2	4818	37.3	F:TGGCCCTGGCACTGACTC R:GGCACGTTCCGGTCAAAGTC
	ss12568582	C-T	Exon 4	21 590	0.0	F:GGACAGTCTGCGGCCA R:CCCTACTCCCTGGGCC
	ss12568610	G-A	Intron 8	11 470	41.9	F:GTGCCCTCACTGCCCTTA R:CCTCAAATAACAGGAATGTTGAGCCCA
	ss12568608	T-C	Exon 9	1078	40.8	F:GGGCCAGGCAGTGGTATCAC R:AGGTCGGCTAGCTTCTGGATCA

*Minor alleles are given in **bold**. †Distance from the previous SNP within each gene in the unit of bp. ‡Minor allele frequency, and those SNP with minor allele frequency >0.8% were used to construct haplotypes.

$$n_e = 1 / \sum p_i^2$$

where p_i is the frequency of the i th haplotype and n_e is the effective number of haplotypes. Expected haplotype heterozygosity was calculated using the equation $1 - (1/n_e)$.

RESULTS

Characteristics of the study subjects

Descriptive characteristics of the study subjects stratified by sex are presented in table 2. Men were generally older, taller, and heavier than women in our sample. The mean BMD unadjusted for any covariates (age, sex, height, and weight) at the femoral neck and hip were significantly different between men and women ($p < 0.001$); men had 6.0% and 10.8% higher femoral neck and hip BMD, respectively.

Covariates accounted for 38.3% and 38.7% of phenotypic variations in femoral neck and hip BMD, respectively. After adjusting for covariate effects, the heritabilities (SE) for femoral neck and hip BMD were estimated to be 0.61 (0.04) and 0.65 (0.04), respectively, which fall into the range of the heritability estimates for BMD reported elsewhere in whites.¹⁵

Haplotype variations

For the six candidate genes studied here, the effective number of haplotypes varied from 2.3 to 7.4 (mean 5.0) and the haplotype heterozygosity varied from 0.57 to 0.91 (mean 0.73). The number of common haplotypes with frequencies $\geq 5\%$ ranged from two to seven, with an average of 4.3 per gene. These common haplotypes accounted for, on average, 87% (ranging from 56 to 98%) of all chromosomes in our white sample (fig 1). We observed a large number of rare haplotypes with frequencies as low as 0.2%, owing to the large sample size used in our study.

Table 2 Basic characteristics of the study subjects

Characteristic	Men (n=749)	Women (n=1124)
Age (year)	48.93 (0.63)	46.10 (0.47)
Height (m)	1.78 (0.003)	1.64 (0.002)
Weight (kg)	88.95 (0.60)	71.61 (0.50)
Femoral neck BMD (cm ²)	0.862 (0.005)	0.810 (0.004)
Hip BMD (cm ²)	1.050 (0.005)	0.937 (0.004)

Note: BMD data are presented as mean (SE) of the raw phenotypic values without adjustment for covariates. Femoral neck BMD data were not available for 123 subjects, and hip BMD data were not available for 25 subjects in our sample.

Haplotype effects on osteoporosis phenotypes

We compared individuals carrying a specific haplotype with non-carriers in our sample. Two rare haplotypes of the *PTHRI* and *VDR* genes showed significant effects on femoral neck and hip BMD, respectively (fig 2). The haplotype H5 of the *PTHRI* gene (that is, the *GATG* haplotype) accounted for only 1.1% of chromosomes in the population (fig 2A). However, individuals carrying H5 had 5.5% significantly higher femoral neck BMD than non-carriers ($p = 4.2 \times 10^{-6}$). This significant result remained unchanged by permutation tests ($p < 1.0 \times 10^{-7}$), suggesting that the association was very unlikely to be a false positive. Similarly, the H10 of the *VDR* gene (that is, the *ATAC* haplotype) accounted for 2.9% of chromosomes in the population, and subjects with this haplotype had an average of 4.0% lower hip BMD than those without ($p = 1.6 \times 10^{-4}$, fig 2B). The other unspecified rare haplotypes did not show significant differences between carriers and non-carriers. Among common haplotypes, the haplotype H4 (that is, the *C+C* haplotype) of the *TGF-β1* gene accounted for 47.7% of chromosomes in the population, and showed significant differences (~1.4%) in femoral neck BMD between its carriers and non-carriers ($p = 4.1 \times 10^{-4}$).

Phenotypic distribution of haplotype carriers

Based on the interesting results for rare haplotypes found above, we further examined the distribution of femoral neck BMD among various *PTHRI* haplotype carriers and attempted to determine if the haplotype effect of the rare H5 is an artefact due to outliers and/or anomalous data distribution (fig 3A). The femoral neck BMD data had kurtosis and skewness coefficients of 0.54 and 0.22, respectively, and fitted well to a normal distribution ($p = 0.34$). Importantly, the H5 haplotype carriers showed quantitatively continuous variation in the sample. They were obviously derived from a wide spectrum rather than only from one extreme tail of the

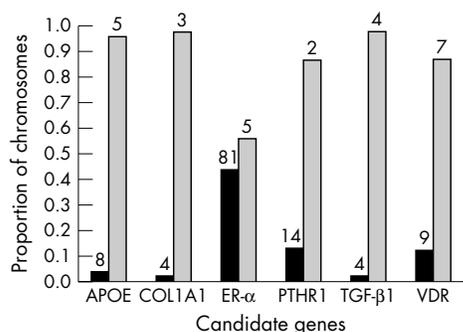


Figure 1 Proportion of chromosomes represented by rare haplotypes with frequencies <5% (black bars) and common haplotypes with frequencies $\geq 5\%$ (grey bars) for different candidate genes. The numbers at the top of bars indicate the total amount of rare or common haplotypes observed in the sample.

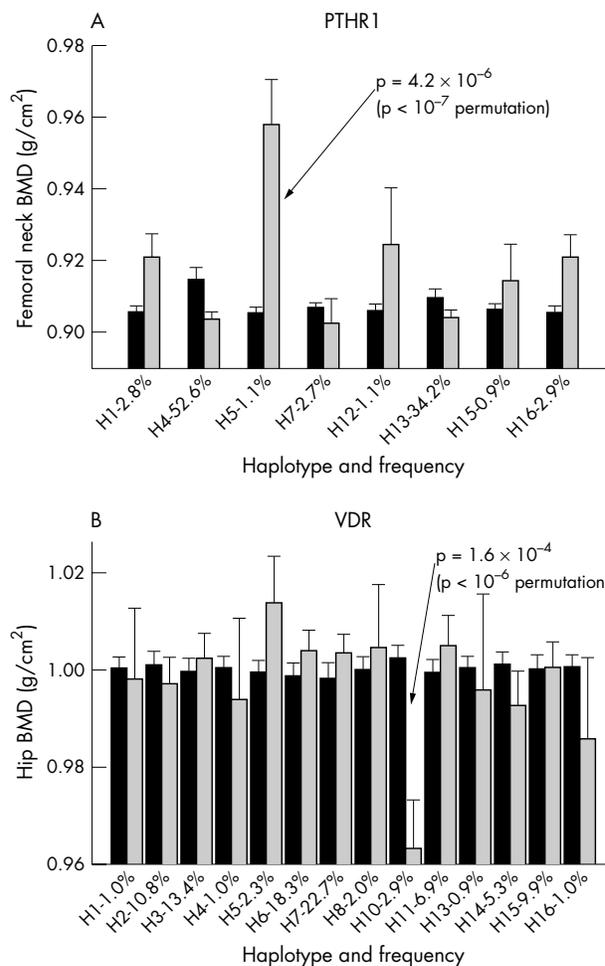


Figure 2 Mean BMD (g/cm²) of carriers (black bars) and non-carriers (grey bars) of specific haplotypes. Only haplotypes with frequencies $>0.8\%$ are considered. (A) Femoral neck BMD and *PTHRI* haplotypes. (B) Hip BMD and *VDR* haplotypes.

population phenotype distribution. In our sample, only two individuals carrying the H5 had femoral neck BMD phenotype within the range $\mu \pm 4\sigma$, while other carriers within the range $\mu \pm 3\sigma$. Similar distribution was observed in the H10-haplotype carriers at the *VDR* gene for the hip BMD data (fig 3B). Therefore, the effects of the rare haplotypes observed are not artefacts due to outliers and/or anomalous data distribution.

DISCUSSION

In our study, we assessed haplotype effects on osteoporosis phenotypes in a large white sample, with a particular emphasis on the potential role of rare haplotypes influencing disease susceptibility. We demonstrated that two rare haplotypes had significant effects on osteoporosis phenotypes. Large phenotypic differences were observed between carriers of these rare haplotypes and non-carriers. These findings indicate that rare haplotypes/variants are important for disease susceptibility and cannot be ignored in genetics studies of complex diseases.

Our results have particular relevance for association studies, particularly using haplotypes, for identification of complex disease genes. The major attraction of haplotype methods is that common haplotypes explain most of the genetic variation in the genomic regions or candidate genes, and that these haplotypes can be captured by a small number

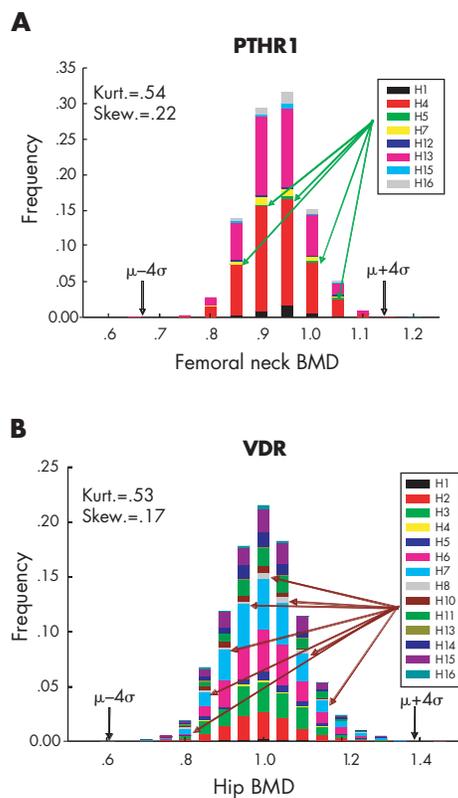


Figure 3 BMD distributions among various haplotype carriers. Carriers of different haplotypes are indicated by different colours. (A) Femoral neck BMD among *PTHRI* haplotype carriers. (B) Hip BMD among *VDR* haplotype carriers.

of tag SNPs. Focusing on these common haplotypes greatly facilitates experimental design and execution of association studies.¹² In our empirical data, on average, four common haplotypes per gene represented 87% of all chromosomes in the sample. We found two rare haplotypes showing large phenotypic difference between carriers and non-carriers. For example, two common *PTHRI* haplotypes accounted for 87% of total genetic diversity in the sample, while the remaining diversity was explained by the remaining 14 rare haplotypes, of which the haplotype H5 (with a frequency of 1.1%) had significant effects on femoral neck BMD. However, we did not observe any significant evidence for the two common haplotypes of *PTHRI*. These results imply that analysis based exclusively on common haplotypes may be inadequate in association studies.

Our results provide indirect evidence for better understanding the genetic architecture of complex diseases, which is ultimately important for the success of association mapping. The “common diseases common variants” (CD-CV) hypothesis proposes that the genetic risk for common disease will often be due to disease predisposing variants found relatively commonly in susceptible populations.^{27–29} Accordingly, a systematic association analysis of common variants in the human genome should reveal the major causative genetic contributions to diseases with considerably greater statistical power than linkage approach. Several favourite examples from the CD-CV proponents include *APOE e4* in Alzheimer’s disease,³⁰ *PPAR γ* Pro12Ala in type 2 diabetes,³¹ factor V Leiden in deep vein thrombosis,³² and *CCR5* in protection against HIV.³³ On the other hand, increasing evidence of allelic complexity at the loci predisposing to complex diseases has been observed,³⁴ in contrast to the CD-CV hypothesis. A recent modelling study showed that

most genetic variance underlying complex diseases probably attributes to loci where susceptibility mutations are mildly deleterious and where the overall mutations rate (and allelic heterogeneity) is relatively high.³⁵ Allelic complexity may be even greater for late onset chronic diseases (such as osteoporosis), as negative selection does not act strongly on phenotypes that typically afflict individuals later in life, after reproduction has taken place.³⁶ Our empirical data demonstrated that two haplotypes of the *PTHRI* and *VDR* genes, which were significantly associated with the variation in osteoporosis phenotypes ($p = 4.2 \times 10^{-6}$ and $p = 1.6 \times 10^{-4}$, respectively), were rare, with frequencies of 1.1% and 2.9%, respectively. The two rare haplotypes conferred large phenotypic differences (4.0–5.0%) for carriers versus non-carriers in the sample. It should be noted that similar significance was also found at a common *TGF- β 1* haplotype that had a frequency of 47.7% and accounted for about 1.4% phenotypic difference in the sample. These findings should increase knowledge on the genetic architecture of complex diseases.

Our results unambiguously showed that rare variants may play an important role in influencing the susceptibility to common diseases. Identification of such rare variants in whole genome association studies will pose a daunting challenge. For example, assuming additive models and using variance component approaches for TDT for sibship data,²⁴ 3940 random sibling pairs would be required to have 80% power to detect a QTL underlying 1% phenotypic variation at a genome-wide significant level of $p < 5.0 \times 10^{-8}$ if the tested marker was a functional mutation variant and had a minor allele frequency of 1%.³⁷ However, only 415 sibling pairs would be required in the same situation if the frequency of the functional mutation allele was increased to 10%. The sample size required to detect rare variants with sufficient statistical power becomes prohibitive in association studies when there is incomplete LD and/or the frequencies of the tested marker and functional mutation allele are not matched. The currently employed sample size in association studies is generally on the order of a few hundred individuals. This sample size may be suitable for identifying common variants, but will probably miss important rare variants without a sample as large as or larger than ours. This attests to the necessity of large sample association studies with collaborative efforts for identifying these rare variants. This also implies that the results from current association studies with small to moderate sample sizes may tend to favour the CD-CV hypothesis unduly.

Data are limited at present to anticipate how frequently the hypothesis of common variants or rare variants is correct at different susceptibility loci.^{38–40} Common variants may contribute to a large extent of phenotype variations in general populations; however, rare variants may also be important in human health and cannot be ignored in genetic studies of complex diseases, because they may confer large phenotype difference for carriers versus non-carriers, as shown here. Certainly, association methods will work well in some susceptibility loci for some common diseases. It should be kept in mind, however, that gene mapping studies should not overlook those rare variants that exert a large effect size on common diseases. Classical linkage analysis and positional cloning still remain the method of choice for identifying rare and high risk disease associated variants, owing to the clear inheritance patterns that they display in large and affected pedigrees.^{38–41}

Several caveats for our findings should be acknowledged. Firstly, the main strength of our study is the large sample with 1873 subjects used in the analyses, which allows assessment of the potential role of rare haplotypes. However, it makes the studied candidate genes unavailable for molecular determination of haplotypes and comprehensive

genotyping, owing to limited resources. In our sample, haplotypes were inferred using family data. Information from relatives and the use of a large family dataset can help resolve haplotype ambiguity and greatly increase the precision of haplotype inference.⁴² Empirical studies also indicate advantages to using family data, including detection of genotyping errors and integration with meiotic maps.⁴³ Therefore, our analyses based on statistically inferred haplotypes are reliable and robust. Secondly, the chosen SNPs were distributed with an average density of 13 kb, which should capture most of the genetic variation in the studied candidate genes. Patterns of haplotype variation observed in our sample were largely consistent with previous studies.^{10–12} Thus, our conclusions drawn from such haplotype data should be generally applicable to genetic studies of complex diseases. Thirdly, although TDT methods are robust to population stratification, they can only use offspring information from informative families. This leads to a potentially large reduction in power to detect allelic associations. In our study, we used whole family data in the analyses, and the results were validated by robust permutation tests. In a previous study, Long *et al* examined population stratification in the same sample as ours by testing the equality of within and between family genetic components.⁴⁴ They did not find any evidence for population stratification in the sample. Therefore, the results obtained from our analyses are convincing.

ACKNOWLEDGEMENTS

Investigators of this work were partially supported by grants from Health Future Foundation, NIH, State of Nebraska, US DOE. The study was also benefited by grants from CNSF, the Huo Ying Dong Education Foundation, and the Ministry of Education of China.

Authors' affiliations

P-Y Liu, Y-Y Zhang, Y Lu, J-R Long, H Shen, L-Juan Zhao, F-H Xu, P Xiao, D-H Xiong, Y-J Liu, R R Recker, H-W Deng, Osteoporosis Research Center, Creighton University, Omaha, NE 68131, USA
H Shen, L-Juan Zhao, F-H Xu, P Xiao, D-H Xiong, Y-J Liu, H-W Deng, Department of Biomedical Sciences, Creighton University, Omaha, NE 68131, USA

H-W Deng, Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan 410081, China

Competing interests: none declared

REFERENCES

- Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* 2000;**97**:10483–8.
- Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. *Am J Hum Genet* 2000;**67**:383–94.
- Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet Epidemiol* 2002;**23**:221–33.
- Zhang K, Calabrese P, Nordborg M, Sun F. Haplotype block structure and its applications to association studies: power and study designs. *Am J Hum Genet* 2002;**71**:1386–94.
- Zhang S, Sha Q, Chen HS, Dong J, Jiang R. Transmission/disequilibrium test based on haplotype sharing for tightly linked markers. *Am J Hum Genet* 2003;**73**:566–79.
- Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 2001;**293**:489–93.
- Zhao H, Pfeiffer R, Gail MH. Haplotype analysis in population genetics and association studies. *Pharmacogenomics* 2003;**4**:171–8.
- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, Bosque-Plata L, Horikawa Y, Oda Y,

- Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hais CL, Bell GI. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000;**26**:163–75.
- Joosten PH, Toepoel M, Mariman EC, Van Zoelen EJ. Promoter haplotype combinations of the platelet-derived growth factor alpha-receptor gene predispose to human neural tube defects. *Nat Genet* 2001;**27**:215–17.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002;**296**:2225–9.
- Patil N, Bero AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001;**294**:1719–23.
- Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001;**29**:233–7.
- Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Chang LY, Huang W, Liu B, Shen Y, Tam PK, Tsui LC, Wayne MM, Wong JZ, Cheng C, Zhang Q, Chee MS, Galver LM, Kruglyak S, Murray SS, Oliphant AR, Montpetit A, Hudson TJ, Chagnon F, Ferretti V, Leboeuf M, Phillips MS, Verner A, Kwok PY, Duan S, Lind DL, Miller RD, Rice JP, Saccone NL, Taillon-Miller P, Xiao M, Nakamura Y, Sekine A, Sorimachi K, Tanaka T, Tanaka Y, Tsunoda T, Yoshino E, Bentley DR, Deloukas P, Hunt S, Powell D, Altshuler D, Gabriel SB, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Aniagwu T, Marshall PA, Matthew O, Nkwodimma C, Royal CD, Leppert MF, Dixon M, Stein LD, Cunningham F, Kanani A, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Donnelly P, Marchini J, McVean GA, Myers SR, Cardon LR, Abecasis GR, Morris A, Weir BS, Mullikin JC, Sherry ST, Feolo M, Altshuler D, Daly MJ, Schaffner SF, Qiu R, Kent A, Dunston GM, Kato K, Niikawa N, Knoppers BM, Foster MW, Clayton EW, Wang VO, Watkin J, Gibbs RA, Belmont JW, Sodergergen E, Weinstock GM, Wilson RK, Fulton LL, Rogers J, Birren BW, Han H, Wang H, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Todani K, Fujita T, Tanaka S, Holden AL, Lai EH, Collins FS, Brooks LD, McEwen JE, Guyer MG, Jordan E, Peterson JL, Spiegel J, Sung LM, Zacharia LF, Kennedy K, Dunn MG, Seabrook R, Shillito M, Skene B, Stewart JG, Valle DL, Jorde LB, Belmont JW, Chakravarti A, Cho MK, Duster T, Foster MW, Jasperse M, Knoppers BM, Kwok PY, Licinio J, Long JC, Marshall PA, Ossorio PN, Wang VO, Rotimi CN, Royal CD, Spallone P, Terry SF, Lander ES, Lai EH, Nickerson DA, Altshuler D, Bentley DR, Boehnke M, Cardon LR, Daly MJ, Deloukas P, Douglas JA, Gabriel SB, Hudson RR, Hudson TJ, Kruglyak L, Kwok PY, Nakamura Y, Nussbaum RL, Royal CD, Schaffner SF, Sherry ST, Stein LD, Tanaka T. The International HapMap Project. *Nature* 2003;**426**:789–96.
- Deng HW, Shen H, Xu FH, Deng HY, Conway T, Zhang HT, Recker RR. Tests of linkage and/or association of genes for vitamin D receptor, osteocalcin, and parathyroid hormone with bone mineral density. *J Bone Miner Res* 2002;**17**:678–86.
- Liu YZ, Liu YJ, Recker RR, Deng HW. Molecular studies of identification of genes for osteoporosis: the 2002 update. *J Endocrinol* 2003;**177**:147–96.
- Long JR, Zhao LJ, Liu PY, Lu Y, Dvornyk V, Shen H, Liu YJ, Zhang YY, Xiong DH, Xiao P, Deng HW. Patterns of linkage disequilibrium and haplotype distribution in disease candidate genes. *BMC Genet* 2004;**5**:11.
- O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.
- Genant HK, Grampp S, Gluer CC, Faulkner KG, Jergas M, Engelke K, Hagiwara S, Van Kuijk C. Universal standardization for dual x-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res* 1994;**9**:1503–14.
- Recker R, Lappe J, Davies K, Heaney R. Characterization of perimenopausal bone loss: a prospective study. *J Bone Miner Res* 2000;**15**:1965–73.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996;**58**:1347–63.
- Elston RC, George VT, Severtson F. The Elston-Stewart algorithm for continuous genotypes and environmental factors. *Hum Hered* 1992;**42**:16–27.
- George VT, Elston RC. Testing the association between polymorphic markers and quantitative traits in pedigrees. *Genet Epidemiol* 1987;**4**:193–201.
- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;**62**:1198–211.
- Sham PC, Cherny SS, Purcell S, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000;**66**:1616–30.
- Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;**19**:149–50.
- Hartl DL, Clark AG. *Principles of population genetics*. Sunderland, MA: Sinauer Associates, 1997.
- Chakravarti A. Population genetics--making sense out of sequence. *Nat Genet* 1999;**21**:56–60.
- Lander ES. The new genomics: global views of biology. *Science* 1996;**274**:536–9.
- Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001;**17**:502–10.

- 30 **Corder EH**, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;**261**:921-3.
- 31 **Altshuler D**, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;**26**:76-80.
- 32 **Bertina RM**, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* 1994;**369**:64-7.
- 33 **Maayan S**, Zhang L, Shinar E, Ho J, He T, Manni N, Kostrikis LG, Neumann AU. Evidence for recent selection of the CCR5-delta 32 deletion from differences in its frequency between Ashkenazi and Sephardi Jews. *Genes Immun* 2000;**1**:358-61.
- 34 **Terwilliger JD**, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 1998;**9**:578-94.
- 35 **Pritchard JK**. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;**69**:124-37.
- 36 **Weiss KM**, Terwilliger JD. How many diseases does it take to map a gene with SNPs? *Nat Genet* 2000;**26**:151-7.
- 37 **Risch N**, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;**273**:1516-17.
- 38 **Clark AG**. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. *Curr Opin Genet Dev* 2003;**13**:296-302.
- 39 **Pritchard JK**, Cox NJ. The allelic architecture of human disease genes: common disease-common variant ...or not? *Hum Mol Genet* 2002;**11**:2417-23.
- 40 **Weiss KM**, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002;**18**:19-24.
- 41 **Botstein D**, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;**33**(suppl):228-37.
- 42 **Chen HS**, Zhang SL. Haplotype inference for multiple tightly linked marker phenotypes including nuclear family information. 2003:165-71.
- 43 **Dawson E**, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaar E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I. A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 2002;**418**:544-8.
- 44 **Long JR**, Liu PY, Liu YJ, Lu Y, Xiong DH, Elze L, Recker RR, Deng HW. APOE and TGF-beta1 genes are associated with obesity phenotypes. *J Med Genet* 2003;**40**:918-24.

Register now!

10th European Forum on Quality Improvement in Health Care
 13-15 April 2005, ExCel Conference Centre, London
 For further information on how to register please go to:
<http://www.quality.bmjpg.com>