Common variants at the PCOL2 and Sp1 binding sites of the COL1A1 gene and their interactive effect influence bone mineral density in Caucasians

P-Y Liu, Y Lu, J-R Long, F-H Xu, H Shen, R R Recker, H-W Deng

Background: Osteoporosis, mainly characterised by low bone mineral density (BMD), is a serious public health problem. The collagen type I alpha 1 (COL1A1) gene is a prominent candidate gene for osteoporosis. Here, we examined whether genetic variants at the COL1A1 gene can influence BMD variation.

Methods: BMD was measured at nine skeletal sites in 313 Caucasian males and 308 Caucasian females. We screened four single nucleotide polymorphisms (SNPs) at the COL1A1 gene: PCOL2 (-1997 G/T) in the promoter, Sp1 (1546 G/T) in the intron 1, Gly19Cys (3911 G/A) in exon 8, and Ala897Thr (13 773 G/A) in exon 45. Univariate and multivariate association approaches were used in the analyses.

Results: In multivariate analyses, we found a strong association between the PCOL2 SNP and BMD (p = 0.007 to 0.024) and a suggestive association between the Sp1 SNP and BMD (p = 0.023 to 0.048) in elderly Caucasian females. Interestingly, the interaction of these two SNPs was highly significantly associated with BMD variation (p = 0.001 to 0.003). The haplotype GG at the two SNPs had, on average, 2.7% higher BMD than non-carriers (p = 0.006 to 0.026).

Conclusions: Our data suggested that the common genetic variants at the PCOL2 and Sp1 sites, and importantly, their interactive effects, may contribute to BMD variation in elderly Caucasian females. Further studies are necessary to delineate the mechanisms underlying the effects of these common variants on BMD variation and to test their clinical relevance for general populations. In addition, our study highlighted the importance of multivariate analyses when multiple correlated phenotypes are under study.

METHODS

Subjects and measurement
A cohort of 313 males and 308 females were selected from 405 nuclear families that were recruited for genetic studies aimed at searching for genes underlying osteoporosis. The selected subjects for association analyses were unrelated. The study was approved by the Creighton University institutional review board. All subjects were Caucasians of European origin. Only healthy people were included with the exclusion criteria detailed earlier by ourselves. Briefly, patients with chronic diseases and conditions which may potentially affect

Abbreviations: BMD, bone mineral density; DXA, dual energy X-ray absorptiometry; FS, factor scores; MANOVA, multivariate ANOVA; PCA, principal component analysis; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium tests; UANOVA, univariate analysis of variance
bone mass were excluded from the study. These diseases/conditions include chronic disorders involving vital organs (heart, lung, liver, kidney, and brain), serious metabolic diseases (diabetes, hypo- and hyperparathyroidism, and hyperthyroidism, etc), other skeletal diseases (Paget disease, osteogenesis imperfecta, and rheumatoid arthritis, etc), chronic use of drugs affecting bone metabolism (cortico-steroid therapy and anti-convulsant drugs), and malnutrition conditions (chronic diarrhoea and chronic ulcerative colitis, etc). All the study subjects signed informed consent documents before entering the study. For each study subject, we obtained information on age, sex, medical history, family history, female history, physical activity, alcohol use, diet habits, and smoking history.

BMDs (g/cm²) of the lumbar spine; femoral neck, trochanter, and intertrochanter regions at the hip; ultradistal, mid-distal, and one-third regions at the wrist; and total body were measured by a Hologic 2000+ or a 4500 DXA scanner (Hologic, Bedford, MA, USA). For the spine, the quantitative phenotype was combined BMD of L1–L4. For the hip, the quantitative phenotype was combined BMD at femoral neck, trochanter, and intertrochanter regions. Both of the machines were calibrated daily, and the coefficient of variability values of the DXA measurements at the spine, hip, femoral neck, trochanter, intertrochanter, ultradistal, mid-distal, one-third, and total body regions were 0.8, 1.0, 1.87, 1.22, 1.25, 2.55, 1.06, 1.13, and 1.13% on the Hologic 2000+, and 0.9, 1.4, 1.98, 1.18, 1.72, 2.28, 1.32, 1.32, and 0.98% on the Hologic 4500, respectively. In total 92% of the subjects were measured on the Hologic 4500 scanner. Data obtained from different machines were transformed to a compatible measurement, which has been shown to be highly reliable and accurate. Members of the same nuclear family were measured on the same scanner. At the same visit as the BMD scan, weight was measured using a calibrated balance beam scale; height was measured using a calibrated stadiometer (table 1).

SNPs genotyping
After searching the dbSNP database (www.ncbi.nlm.nih.gov/SNP/) and reviewing previously published studies, four SNPs were selected in the COLIA1 gene in the study (table 2). The selected SNPs for the study were based on a comprehensive consideration of: (i) functional relevance and importance, (ii) position in or around the gene, and (iii) their use in previous genetic epidemiology studies.

Genomic DNA was extracted from whole blood using a commercial isolation kit (Genta Systems, Minneapolis, MN, USA) following the procedure detailed in the kit. The genotyping procedure for all SNPs was similar, involving polymerase chain reaction (PCR) and Invader assay (Third Wave Technology, Madison, WI, USA). PCR was performed in 10 μl reaction volume with 35 ng genomic DNA, 0.2 mM each of dCTP, dATP, dGTP, and dTTP, 1×PCR buffer and 1.5 mM MgCl₂, 0.4 μM each of the primers, and 0.35 U Taq polymerase (ABI, Applied Biosystems, Foster City, CA, USA).

The sequences of the PCR primers for the four SNPs are presented in table 2. The following procedure was used on 9700 Thermal Cyclers: 95 °C for 5 min, 30 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min, and then 72 °C for 5 min. After amplification, the product was diluted 1:20 in nuclease-free water. The Invader reaction was performed in 7.5 μl reaction volume, with 3.75 μl diluted PCR product, 1.5 μl probe mix, 1.75 μl Cleavage FRET mix, and 0.5 μl Cleavage enzyme/MgCl₂ solution (Third Wave Technology). The reaction mix was overlaid by 15 μl mineral oil and denatured at 95 °C for 5 min, and then incubated at 63 °C for 20 min on 9700 Thermal Cyclers. After incubation, the fluorescence intensity for both colours (FAM and Red dyes) was read out using Cytofluor 4000 (ABI). The data were then loaded to the software Invader Analyzer (Third Wave Technology), and the genotype for every sample was called according to the ratio of the fluorescence intensity of the two dyes. The PedCheck program (http://watson.hgen.pitt.edu/register/soft_doc.html) was used to verify Mendelian inheritance of the alleles within each family and the family relationships.

Statistical analyses
Univariate analysis
Association of the COLIA1 gene with BMD can be tested by using univariate analysis of variance (UANOVA), with different genotypes as main effects and BMD of different subjects within each genotype as random effects. UANOVA analyses were performed for each of the nine BMD phenotypes separately. The test statistic for UANOVA is the $F$ statistic, which measures the ratio of between-group variability to within-group variability. Significant results may indicate genotypic difference in BMD variation (that is, declaring an association of the COLIA1 gene with BMD).

Multivariate analysis
Unlike UANOVA, multivariate ANOVA (MANOVA) can be used to test joint genotypic differences in multiple BMD phenotypes. Two commonly used test statistics for MANOVA were adopted in our analyses: Wilks’ lambda and Roy’s largest root. The former is an exact $F$ statistic and the later is an upper bound $F$ statistic. In the MANOVA analyses, partial correlation coefficients between different BMD phenotypes were estimated, and their significance was tested. The partial correlation measures the strength of relationship between two variables, controlling for the effects of other variables.

Principal component analysis (PCA)
BMD variation at various skeletal sites may share common genetic sources. The correlation analysis showed that BMD phenotypes are significantly correlated among different skeletal sites in our sample (table 3). Therefore, a stabilised multivariate analysis based on linear scores of multiple phenotypes obtained from PCA was performed on our data as follows. First, PCA of the covariance matrix of the nine BMD phenotypes was conducted and factor scores (FS) for each individual were calculated. Then, UANOVA analyses were performed for the three principal FS separately (see Results). The aforementioned MANOVA analysis of these FS was also performed.

Each of the common SNPs and their haplotypes at the COLIA1 gene were analysed separately. The haplotypes were reconstructed according to all subjects from the nuclear
families using the program Genehunter version 2.1 (http://www.hgmp.mrc.ac.uk/Registered/Option/genehunter.html).27 The linkage disequilibrium between SNPs was calculated by Lewontin’s measure.28 Before the genetic analyses, the data were tested for measured potentially important covariates and adjusted for the significant ones (age, age ², weight, and height) by a multiple stepwise regression procedure (carried out separately for each sex), using a significance level of 0.1 for both inclusion and retention in the model. All the data analyses were performed for male and female groups separately and implemented by using PROC REG, GLM, MANOVA, and PRINCOMP in SAS 8.0.e. The experiment-wise significance level was set at 0.05.

RESULTS

Table 1 presents the basic characteristics of the male and female study subjects separately. The males had significant higher mean BMD at the nine skeletal sites than the females (p<0.001), indicating gender effects on BMD variation. Table 2 gives minor allele frequencies of the four SNPs at the COL1A1 gene. The T allele of the Gly19Cys was not observed in our sample. The minor allele frequencies of the SNPs were 0.152, 0.187, and 0.019, respectively. Multivariate analysis of the univariate analyses, only a weak association of the PCOL2 with total body BMD (p<0.056) was observed. This association was obviously not significant after Bonferroni correction. In the multivariate analysis, an enhanced association of the PCOL2 with composite BMD phenotypes was found with p values of 0.034 (Wilk’s lambda) and 0.016 (Roy’s largest root). Similar results were observed at the Sp1 with p values of 0.048 and 0.023, respectively. In the subsequent principal component analyses (PCA), the first three principal factor scores (FSs) accounted for 74.4%, 10.1%, and 6.30% of the total BMD variation at the nine skeletal sites, respectively. p Values for univariate association analyses of the three factor scores FS1, FS2, and FS3 with the PCOL2 were 0.474, 0.029, and 0.074, respectively. Multivariate analysis of the

| Table 2 | Information and the primer sequences for the studied SNPs in the COL1A1 gene |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| SNP             | dbSNP* or reference | Polymorphism† | Frequency‡ | Location in gene§ | Primers (5’–3’) |
| PCOL2           | Garcia-Giralt et al* | -1997 G/T | 0.152 | Promoter | F: GCACCTGCTCCCTAGCCAC |
|                 | Grant et al†     | 1546 G/T | 0.187 | Intron 1 (3543) | R: CCTAGTCCGCGCGACTGCA |
| Gly19Cys        | Lund et al‡     | 3911 G/T | – | Exon 8 (2365) | F: CACATGCCGGCTCAGGCTCAC |
| Ala897Thr       | rs1800215     | 13 773 G/A | 0.019 | Exon 45 (9862) | F: GCAGACCTGGATGAGGCCA |

*SNP ID in the database dbSNP, www.ncbi.nlm.nih.gov/SNP; †Bold-faced letters are the minor alleles; ‡Allele frequencies of the minor alleles; §The T allele of SNP3 was not observed in our sample; †The numbers in parentheses are physical distances from the previous SNP in bp units; ‡The Sp1 polymorphism lies at position 1546 on the sequence AF017178 (GenBank accession number).

| Table 3 | Partial correlation coefficients between BMD at different skeletal sites |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| BMD             | Lumbar spine   | Total hip       | Femoral neck    | Trochanter      | Intertrochanter | Ultradistal     | Mid-distal      | One-third       | Total body      |
| Lumbar spine    | 0.695          | 0.598           | 0.678           | 0.635           | 0.571          | 0.452           | 0.411           | 0.733           |
| Total hip       | 0.684          | 0.831           | 0.888           | 0.962           | 0.640          | 0.487           | 0.453           | 0.745           |
| Femoral neck    | 0.682          | 0.845           | 0.725           | 0.738           | 0.544          | 0.457           | 0.438           | 0.669           |
| Trochanter      | 0.671          | 0.915           | 0.784           | 0.878           | 0.592          | 0.427           | 0.386           | 0.674           |
| Intertrochanter | 0.623          | 0.974           | 0.825           | 0.825           | 0.562          | 0.605           | 0.417           | 0.696           |
| Ultradistal     | 0.542          | 0.590           | 0.557           | 0.562           | 0.552          | 0.790           | 0.627           | 0.673           |
| Mid-distal      | 0.494          | 0.580           | 0.554           | 0.519           | 0.566          | 0.838           | 0.865           | 0.600           |
| One-third       | 0.496          | 0.577           | 0.543           | 0.513           | 0.559          | 0.856           | 0.865           | 0.562           |
| Total body      | 0.747          | 0.746           | 0.692           | 0.721           | 0.702          | 0.620           | 0.616           | 0.644           |

Values above and below the diagonal indicate correlation coefficients for male and female subjects, respectively. All correlation coefficients are significantly larger than zero (p<0.001).

| Table 4 | p Values of association tests for each of the two SNPs with BMD in the female group |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| PCOL2           | 0.056           | 0.805           |                  |                  |
| Sp1             | 0.048           | 0.023           |                  |                  |

*In the UANOVA analyses of each of the nine BMD phenotypes, only the body BMD showed a p value of 0.056 for the PCOL2; †The number in the parenthesis is the proportion that the FS (factor score) accounts for of the total BMD variation at the nine skeletal sites.
three factor scores produced stronger evidence for the association with the PCOL2 with p values of 0.024 (Wilks’ lambda) and 0.007 (Roy’s largest root). However, no increasing evidence for association was observed at the Sp1 in the PCA analyses.

Based on the interesting results for single SNPs found above, we further tested associations of the interaction of the PCOL2 and Sp1 and their haplotypes with BMD (table 5). The two SNPs were in significant linkage disequilibrium (D’ = 0.918). The three common haplotypes TG, GG and GT had the frequencies of 0.172, 0.616, and 0.210, respectively, while the other haplotype TT was rare with a frequency of 0.002, and thus was excluded from further analyses. The interaction of the PCOL2 and Sp1 showed highly significant evidence for the association of BMD phenotypes (p = 0.003 for Wilk’s lambda and p = 0.001 for Roy’s largest root). Significant evidence was only found at the haplotype GG (p = 0.026 for Wilk’s lambda and p = 0.006 for Roy’s largest root) and carriers of the GG haplotype had, on average, 2.7% higher BMD than non-carriers.

**DISCUSSION**

In the present study, we analysed the association of the common SNPs at the COLIA1 gene and their haplotypes with BMD phenotypes in elderly Caucasians. We found a strong association between the PCOL2 and BMD phenotypes and a suggestive association between the Sp1 and BMD phenotypes in the elderly females. Interestingly, the interaction of the PCOL2 and Sp1 was shown to be most significantly associated with BMD variation. Of their four haplotypes, strong evidence was also found for the haplotype GG in relation to BMD variation.

**Association of the COLIA1 gene with BMD and its putative molecular mechanisms**

To date, a number of studies have been performed to test whether allelic variants of the COLIA1 gene can be associated with variations in osteoporosis phenotypes. The tested PCOL2 (-1997 G/T), named by Garcia-Giralt et al., was recently found within the COLIA1 promoter region known to harbour cis elements important for in vivo expression. This polymorphism site may bind primary osteoblast nuclear proteins and was reported to be significantly associated with BMD values at the lumbar spine and femoral neck in postmenopausal women of Spanish origin. However, no further study on this SNP has been documented. Our results confirmed this positive association in a cohort of elderly Caucasian women. Similar analyses did not reveal such an association in elderly Caucasian men.

The tested Sp1 was one of the most frequently studied polymorphisms affecting a binding site for the transcription factor Sp1 in the COLIA1 gene. The polymorphism Sp1 site lies within the first intron of the COLIA1 gene and was first reported to be associated with BMD and osteoporosis fractures by Grant et al. Our results provide suggestive evidence for an association of the Sp1 polymorphism and BMD phenotypes in elderly Caucasian women. Several other studies have also confirmed the positive association with bone loss in elderly populations. Recently, a meta-analysis of 26 studies confirmed the association for the T allele of the Sp1 polymorphism with a modest reduction in BMD and a significant risk of osteoporotic fracture. However, conflicting results were also reported in younger populations, such as Norwegian young adults, American premenopausal women, Irish young men and women, and French premenopausal women. In our analyses, we did not find any association of the Sp1 polymorphism with BMD phenotypes in elderly men. This is consistent with most previous studies, suggesting that the COLIA1 gene may exert a greater influence on BMD variation in females than in males or there may be gender heterogeneity of the COLIA1 gene on BMD.

The most remarkable finding here is that the interaction of the PCOL2 and Sp1 may play an important role in contributing to BMD variation besides their respective effects. In our female sample, after adjusting for age, age², height, and weight, subjects with the PCOL2 G allele had an average of 1.5% higher BMD values than those without this allele. Subsets carrying the Sp1 G allele had 0.5% higher mean BMD values than non-carriers. However, haplotype GG carriers had, on average, 2.7% higher BMD values than non-carriers, indicating a larger haplotype joint effect than their accumulated individual effects on BMD variation. Interestingly, the PCOL2 was found to interact with another SNP in the promoter of COLIA1, namely PCOL1 (-1663 indelT) that was in strong linkage disequilibrium with the Sp1. Bornstein et al. proposed that normal regulation of COLIA1 gene transcription may result from the interplay of positive and negative elements present in the promoter region and within the first intron. Recently, Mann et al. found that the T allele (that is, “s” allele) of the Sp1 polymorphism has higher binding affinity for Sp1 protein compared with the G allele (that is, “a” allele). Higher affinity for Sp1 protein at the T allele increases the relative abundance of collagen α1(I) to s2(1) protein chains in “TG” heterozygotes that is accompanied by an increase in the relative abundance of COL1A1 to COL1A2 mRNAs. The abnormal relative level of COL1A1 to COL1A2 mRNAs may reduce bone quality and quantity. In addition, gel retardation assays demonstrated that the G allele of the PCOL2 polymorphism has a higher binding capacity of single-stranded binding proteins. The single-stranded binding proteins may repress transcription either by preventing access of dsDNA binding factors (such as Sp1 protein mentioned above) or by stabilising a secondary structure, which would prevent their binding. It seems that the allelic variants at the PCOL2 and Sp1 binding sites may interact in transcription regulation of the COLIA1 gene. This may partially explain the larger haplotype difference between the haplotype GG carriers and non-carriers and the most significant association of the interaction of the PCOL2 and Sp1 polymorphisms with BMD phenotypes in our study.

**Statistical issues involved in analyses and their implications**

The results from our comparative analyses also have important implications for association studies in bone fields. The multivariate analyses revealed strong and/or suggestive evidence for association of the COLIA1 gene with BMD phenotypes, while the univariate analyses largely failed to detect such an association (table 4). BMD variation at various skeletal sites may share common genetic sources.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>p Values of association tests for the interaction of the PCOL2 and Sp1 and their haplotypes with BMD in the female group*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP interaction</td>
<td>Haplotypes†</td>
</tr>
<tr>
<td>PCOL2-Sp1</td>
<td>TG (0.172)</td>
</tr>
<tr>
<td>Wilk’s lambda</td>
<td>0.003</td>
</tr>
<tr>
<td>Roy’s largest root</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*MANOVA of the nine BMD phenotypes was performed for the interaction of the PCOL2 and Sp1 and their haplotypes; †The rare haplotype TT with a frequency of 0.003 was not tested in the analyses (see text); ‡The number in the parenthesis is the haplotype frequency.
present sample, we observed highly significant correlations between BMD at various skeletal sites (table 3). From a statistical point of view, multivariate analysis has several advantages over its simpler univariate counterpart.34 First, multivariate analysis exploits important information regarding the relationship between dependent variables (such as different BMD measurements here) and thus may reveal differences not shown in the multi-fold univariate analyses. Second, multivariate analysis obviates multiple testing problems that arise in univariate analyses and thus may increase statistical power. From a biological point of view, multivariate analyses of BMD phenotypes jointly may decrease the effect of developmental randomness during ontogenesis of various skeletal sites. Therefore, multivariate association analysis may be more powerful than multi-fold univariate association analyses as shown in our results. It should be noted, however, that univariate analysis is warranted in some situations, for example, when there is genetic heterogeneity among surrogate phenotypes of complex diseases.35

The SNPs Gly19Cys in exon 8 and Ala897Thr in exon 45 are very rare with frequencies of <2% in our sample. The two rare variants may also play a role in influencing BMD variation. However, identification of such rare variants using association studies will pose a daunting challenge. We simulated a functional mutation allele with minor frequency of 2% and responsible for 1% phenotypic variation under additive inheritance in the general population. Simulations were performed in 10 000 replicates using a significance level of 0.05. To achieve 80% power for association analysis, the sample size required for finding such a rare variant was approximately 4200 unrelated subjects, while in the same situation only about 160 subjects would be required if the frequency of the functional allele was increased to 15%. Thus, larger sample sizes are required for future investigation of such rare variants.

In addition, our sample was from 405 nuclear families with a total of 1873 subjects. We also performed transmission disequilibrium tests (TDT) in these nuclear families and did not find any significant/suggestive association between the common SNPs and BMD phenotypes (Liu et al, unpublished). Since the TDT analyses only utilise phenotype information of informative offspring, it may lose considerable statistical power compared with association analyses. However, association studies may yield spurious results due to population substructure. Several studies showed that population substructure may not be a problem for association studies in US Caucasian populations,32 including our recent data for our study population.33 The study subjects were recruited from a population of relatively homogenous ethnic background and population stratification was not observed.34 Therefore, the results obtained in the present study should be robust and reliable.

CONCLUSIONS

Our data suggest that the common genetic variants at the PCOL2 and Sp1 sites of the COLIA1 gene, and importantly, their interactive effects, may contribute to BMD variation in elderly Caucasian females. Further studies are necessary to delineate the mechanisms underlying the effects of these common genetic variants on BMD variation and to test their clinical relevance for general populations. In addition, our study highlighted the importance of multivariate analyses when multiple correlated phenotypes are under study.

Authors’ affiliations

P-Y Liu, Y Lu, J-R Long, F-H Xu, H Shen, R R Recker, H-W Deng. Osteoporosis Research Center, Creighton University, Omaha, NE 68131, USA

F-H Xu, H Shen, 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

The study was partially supported by grants from the Health Future Foundation of NH and the State of Nebraska. The study was also benefited by grants from CNSF, the Huo Ying Dong Education Foundation, and the Ministry of Education of China.

Conflict of interest: none declared.

REFERENCES


