# Genome-wide Copy-Number-Variation Study Identified a Susceptibility Gene, *UGT2B17*, for Osteoporosis

Tie-Lin Yang,[1,7] Xiang-Ding Chen,[2,7] Yan Guo,[1] Shu-Feng Lei,[2] Jin-Tang Wang,[4] Qi Zhou,[1] Feng Pan,[1] Yuan Chen,[1] Zhi-Xin Zhang,[1] Shan-Shan Dong,[1] Xiang-Hong Xu,[1] Han Yan,[1] Xiaogang Liu,[1] Chuan Qiu,[1] Xue-Zhen Zhu,[1] Teng Chen,[4] Meng Li,[4] Hong Zhang,[4] Liang Zhang,[5] Betty M. Drees,[3] James J. Hamilton,[3] Christopher J. Papasian,[3] Robert R. Recker,[6] Xiao-Ping Song,[1] Jing Cheng,[5] and Hong-Wen Deng[1,2,3,*]

Osteoporosis, a highly heritable disease, is characterized mainly by low bone-mineral density (BMD), poor bone geometry, and/or osteoporotic fractures (OF). Copy-number variation (CNV) has been shown to be associated with complex human diseases. The contribution of CNV to osteoporosis has not been determined yet. We conducted case-control genome-wide CNV analyses, using the Affymetrix 500K Array Set, in 700 elderly Chinese individuals comprising 350 cases with homogeneous hip OF and 350 matched controls. We constructed a genomic map containing 727 CNV regions in Chinese individuals. We found that CNV 4q13.2 was strongly associated with OF ($p = 2.0 \times 10^{-4}$, Bonferroni-corrected $p = 0.02$, odds ratio = 1.73). Validation experiments using PCR and electrophoresis, as well as real-time PCR, further identified a deletion variant of *UGT2B17* in CNV 4q13.2. Importantly, the association between CNV of *UGT2B17* and OF was successfully replicated in an independent Chinese sample containing 399 cases with hip OF and 400 controls. We further examined this CNV's relevance to major risk factors for OF (i.e., hip BMD and femoral-neck bone geometry) in both Chinese (689 subjects) and white (1000 subjects) samples and found consistently significant results ($p = 5.0 \times 10^{-4}$ –0.021). Because *UGT2B17* encodes an enzyme catabolizing steroid hormones, we measured the concentrations of serum testosterone and estradiol for 236 young Chinese males and assessed their *UGT2B17* copy number. Subjects without *UGT2B17* had significantly higher concentrations of testosterone and estradiol. Our findings suggest the important contribution of CNV of *UGT2B17* to the pathogenesis of osteoporosis.

## Introduction

Copy-number variation (CNV) is a common type of genomic variability, with variations in the size of DNA fragments ranging from 1 kilobase (Kb) to several megabases (Mb). CNVs, which may include duplications or deletions,[1–3] can influence gene expression by disrupting coding sequences, perturbing long-range gene regulation, or altering gene dosage, and these effects could contribute to phenotypic variations or disease risk.[4–6] CNVs can account for nearly 20% of the total detected genetic variation in gene expression.[7]

CNVs have been shown to be associated with several distinct complex human diseases.[8–11] As SNP-genotyping technology has improved, a number of studies using array-based methods have successfully identified CNVs related to specific complex human disorders. For example, Friedman et al.[12] found several copy-number changes associated with mental retardation in children by using Affymetrix Human Mapping 100K Arrays, and Marshall et al.[11] used Affymetrix Human Mapping 500K Arrays to detect several copy-number changes that were associated with increased susceptibility to autism spectrum disorders (MIM 209850).

Osteoporosis (MIM 166710) is the most common metabolic bone disease, generally characterized by low bone-mineral density (BMD), poor bone quality (characterized by bone geometry, etc.), and/or increased susceptibility to low-impact osteoporotic fractures (OF). Although bone geometry correlates with BMD in measurements, it might play an important role independent of BMD in determining bone quality.[13] Therefore, BMD and bone geometry are two major measurable and studied risk factors used to predict susceptibility to OF, the ultimate clinical outcome of osteoporosis. It is widely accepted that variation in BMD and bone geometry, as well as risk for OF, are largely determined by genetic factors, but the overwhelming majority of genetic variation accounting for osteoporosis risk has not been determined.[14–17] Previous investigations, including recent genome-wide association studies,[18–21] have discovered several genes or SNPs that appear to enhance the risk of developing osteoporosis. Collectively, however, all of these implicated genes or SNPs account for no more than 5% of the risk for developing osteoporosis. Consequently, it has become important to explore the possibility that some of the remaining undiscovered genetic factors that influence risk of osteoporosis are due to genomic mechanisms other than individual mutation changes, such as CNVs.

[1]Key Laboratory of Biomedical Information Engineering of the Ministry of Education, Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China; [2]Molecular and Statistical Genetics Lab, College of Life Sciences, Hunan Normal University, Changsha 410081, People's Republic of China; [3]School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA; [4]School of Medicine, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China; [5]National Engineering Research Center for Beijing Biochip Technology, Beijing 102206, People's Republic of China; [6]Osteoporosis Research Center, Creighton University, Omaha, NE 68131, USA
[7]These two authors contributed equally to this work
*Correspondence: dengh@umkc.edu
DOI 10.1016/j.ajhg.2008.10.006. 

In this study, we first performed case-control genome-wide CNV analyses using Affymetrix Human Mapping 500K Arrays, which are effective in identification of genomic CNVs.[2,22] The study population included 700 elderly Han Chinese subjects comprising 350 cases with a history of hip OF and 350 healthy and matched controls. On the basis of the mapping results for genomic CNVs, we conducted association analyses to identify CNVs underlying susceptibility to OF. For those CNVs that were significantly associated with OF, we performed further analysis using PCR and electrophoresis, as well as real-time PCR, to validate and fine map the CNV region for specific genes involved. We identified a CNV of the *UGT2B17* gene (UDP glucuronosyltransferase 2 family, polypeptide B17 [MIM 601903]) that was associated with risk of OF. More importantly, this CNV's association with OF was successfully replicated in an independent Chinese sample consisting of 399 cases with hip OF and 400 controls. We further detected the significant relevance of this CNV to hip BMD and femoral neck (FN) bone geometry using two independent samples of 689 unrelated Chinese subjects and 1000 unrelated white subjects. Our findings strongly support the importance of CNV of *UGT2B17* in the pathogenesis of osteoporosis.

## Subjects and Methods

### Study Subjects

The study was approved by the local institutional review boards or the office of research administration of all participating institutions. After signing an informed consent, all subjects received assistance in completing a structured questionnaire including questions about anthropometric variables, lifestyle, and medical history.

*Chinese Genome-Wide Analyses (GWA) Sample*

This GWA sample consisted of 350 patients with a history of hip OF and 350 healthy age-matched controls. All subjects were unrelated northern Han Chinese adults living in the city of Xi'an and its neighboring areas. Affected individuals with low-impact hip fractures were recruited from affiliated hospitals of Xi'an Jiaotong University. Inclusion criteria for cases were (1) age $< 80$ yrs and age at onset $> 55$ yrs (all women were postmenopausal), (2) minimal trauma fracture, usually due to falls from standing height or less, and (3) fracture site at the femoral neck or intertrochanter regions. Hip fracture was identified through physician diagnosis according to radiological reports or X-ray services; patients with pathological fractures due to known disease conditions or syndromes or with high-trauma fractures (such as those from a motor-vehicle accident) were excluded.

Healthy control subjects were enrolled by use of local advertisements. They were geography- and age-matched to the cases. Inclusion and exclusion criteria for controls included the stipulations that (1) age at exam must be $> 55$ yrs, without any fracture history (all female controls were postmenopausal) and that (2) subjects with chronic diseases and conditions that might potentially affect bone mass, structure, or metabolism were excluded. These exclusions ensure that controls are less likely to suffer OF during the remainder of their life, as compared with the general population. Diseases or conditions resulting in exclusion included chronic disorders involving vital organs (heart, lung, liver, kidney, brain), serious metabolic diseases (diabetes, hypo- and hyperparathyroidism, hyperthyroidism, etc.), other skeletal diseases (Paget's disease, osteogenesis imperfecta, rheumatoid arthritis, etc.), chronic use of drugs affecting bone metabolism (e.g., hormone-replacement therapy, corticosteroid therapy, anticonvulsant drugs), and malnutrition conditions (such as chronic diarrhea, chronic ulcerative colitis). Also included in the inclusion and exclusion criteria was the stipulation that (3) subjects taking anti-bone-resorptive or bone anabolic agents or drugs, such as bisphosphonates, were excluded.

*Chinese Replication Sample*

For replication of the CNV's association with hip OF, an independent Chinese sample containing 799 unrelated Han subjects (399 cases with hip OF and 400 healthy age-matched controls) was used. All of the subjects resided in the same area as the above Chinese GWA sample, and the sample inclusion and exclusion criteria for cases and controls were the same as those adopted in the recruitment of the above GWA sample.

The significance and relevance of the CNVs identified as significant for OF were subsequently evaluated for hip BMD and femoral neck (FN) bone geometry variations in two independent samples (Chinese unrelated sample and white unrelated sample); descriptions are given below.

*Chinese Unrelated Sample*

This sample contained 689 unrelated ethnic Han subjects residing in the same area as the above Chinese samples. Exclusion criteria were the same as those adopted in the recruitment of the control subjects in the above sample and have been detailed in our earlier publication.[23]

*White Unrelated Sample*

This sample consisted of 1000 unrelated subjects, who were all white individuals of Northern European origin living in Omaha, NE and its surrounding regions in midwestern USA. They were defined by the same exclusion criteria as those used in the above Chinese unrelated sample.

*Chinese Young Male Sample*

Because the CNV that we identified consistently significant in the above samples encodes an enzyme catabolizing steroid hormones, we used a sample of young Han Chinese males to examine the relationship between the CNV of interest and serum sex hormone levels. This sample consisted of 236 unrelated, male ethnic Han subjects from ages 22 to 28 who resided in the same area, but were independent of, the above Chinese samples. Sample exclusion criteria were the same as those adopted in the recruitment of control subjects in the above case-control Chinese samples.

### BMD and Bone-Geometry Measurements

The Chinese and white unrelated samples were measured for hip BMD and bone-geometry parameters at the FN. The same model of instrument and protocol were used for both samples. Hip BMD values were measured via dual-energy X-ray absorptiometry (DXA) with Hologic 4500 scanners (Hologic, Bedford, MA, USA) that were calibrated daily. The coefficient of variation (CV) for hip-BMD measurements was less than 1.5% in both populations. Using DXA-derived FN BMD and bone size, we estimated two FN geometric variables. The algorithm was described earlier and has been extensively used.[24–26] The estimated FN geometric variables are buckling ratio (BR), an index for bone structural instability, and cortical thickness (CT), an important biomechanical determinant of bone strength.

## Genome-wide Genotyping and Quality Control

Genomic DNA was extracted from peripheral-blood leukocytes with the use of standard protocols. Affymetrix Human Mapping 500K Array Sets (Affymetrix, Santa Clara, CA, USA), which consisted of two chips (Nsp and Sty) with ~250,000 SNPs each, were used for genotyping each subject from the Chinese GWA sample and the randomly selected white unrelated sample, according to the Affymetrix protocol. In brief, approximately 250 ng of genomic DNA was digested with restriction enzyme NspI or StyI. Digested DNA was adaptor ligated and PCR amplified for each enzyme-digested sample. Fragment PCR products were then labeled with biotin, denatured, and hybridized to the arrays. Arrays were then washed and stained with Phycoerythrin on Affymetrix Fluidics Station FS450 and scanned with the GeneChip Scanner 3000 7G for quantitation of fluorescence intensities. Data management and analyses were conducted with the Affymetrix GeneChip Operating System.

Genotyping calls were determined from the fluorescence intensities with the use of the DM algorithm with a 0.33 p value setting,[27] as well as the BRLMM algorithm.[28] As a result of the efforts of repeat experiments, all of the samples had a call rate of $\geq$ 95% and were thus included in subsequent analyses. The final mean BRLMM call rates reached a high level of 99.02% and 99.14% for the Chinese (n = 700) and white (n = 1000) samples, respectively.

## Assessment of Genetic Background

The STRUCTURE 2.2 program[29] and the method of genomic control[30] were applied to the Chinese GWA sample and the white unrelated sample for the detection of possible population stratification. For structure analysis, in each sample, 2000 SNPs tested to be in Hardy-Weinberg equilibrium were randomly selected genome-wide for clustering of all of the subjects. The program uses a Markov chain Monte Carlo (MCMC) algorithm to cluster individuals into different cryptic subpopulations on the basis of multilocus genotype data. Potential substructure was estimated under an a priori assumption of $K = 2$ discrete subpopulations. For genomic control, we estimated the inflation factor ($\lambda$) on the basis of genome-wide SNP information.

## CNV Calculation

*CNV Analysis of Autosomes in the GWA Sample by CNAT*

DNA CNVs were calculated with the Affymetrix GeneChip Chromosome Copy Number Analysis Tool 4.0 (CNAT 4.0), which implements a Hidden Markov Model (HMM)-based algorithm to identify chromosomal gains and losses by comparing the signal intensity of each SNP probe set for each OF sample against a reference set. As an initial analysis, we used 350 unrelated control subjects as the reference set. We estimated CNVs from cases with OF by comparing the cases to the reference set. In calculating CNVs for controls, when an individual subject was the test sample, he or she was excluded from the reference set, so that the reference set contains all control subjects except the one being tested. CNV was defined when there were at least three consecutive SNPs showing consistent deletion or duplication. Because it was not possible to delineate the exact boundaries of each CNV with genome-wide SNP-genotyping arrays, we used the positions of SNPs to approximate CNV boundaries. After putative variant intervals of CNVs were identified in each individual, we used the following criteria to determine the boundaries of CNV regions (CNVR) (Figure S1, available online). If two individual CNVs overlapped, we merged them as a CNVR, using as boundaries the maximum-interval SNPs selected from these two CNVs. When the interval of the next overlapping individual CNV exceeded this CNVR, the boundaries would extend accordingly.[31] In brief, a CNVR represented a union of overlapping CNVs. We used CNVRs to construct the genomic CNV map in Chinese samples.

*CNV Analysis of Chromosome X in the GWA Sample by CNAT*

CNV of chromosome X was analyzed with CNAT with the use of the corresponding male or female reference set as follows. Each male case subject with OF was compared with a reference set comprising all 173 male controls, and each female OF case subject was compared with a reference set comprising all 177 female controls. In calculating CNVs of controls, when a male or female was the test sample, he or she was excluded from the corresponding reference set, so that the reference set contained all control subjects except the one being tested.

## Analysis of Association between CNV and OF in the GWA Sample

For association analyses, we used the following procedure to redefine the CNVs (those with frequencies exceeding 1%) contained in the CNVRs. We divided complex CNVRs (illustrated in plots C–E in Figure S1), including individual CNVs with discordant boundaries but overlapping regions, into several sub-CNVRs, so that the resulting sub-CNVRs had the same configurations as those in plots A and B in Figure S1. Thus, all CNVRs or sub-CNVRs contained only one kind of CNV with the same boundaries as those of their corresponding CNVR or sub-CNVR. CNVs with frequencies $> 1\%$, defined by the above procedure, were selected for association analyses. For each CNV, we adopted the nonparametric Mann-Whitney U test to evaluate the significance of frequency-distribution differences, between cases and controls, using SPSS 13.0 software.[8,9] Two-tailed p values were estimated by 100,000 Monte Carlo simulations. p values $< 0.05$ in our study were considered nominally significant and were subjected to Bonferroni correction to account for multiple comparisons. The conservative significance threshold for a single test was assessed at a type I error rate of 0.05/N. N was the total number of the tested CNVs. For the significant CNVs, we used a multivariate logistic-regression model to examine their relation to OF risk, taking into account potential covariates such as height, weight, sex, age, and age$^2$.

## CNV-Validation Experiments

As described later, in the Results section, our initial CNAT analyses identified a CNV at 4q13.2 that was significantly associated with risk of OF. Previous studies showed that this region was a recombination "hotspot" region with deletion.[2,31–34] Hence, we performed the following experiments in order to validate this CNV.

*PCR and Electrophoresis Experiments*

PCR and electrophoresis experiments yielded DNA fragments that could be clearly distinguished for CNVs with and without homozygous deletion. Thus, these methods could be used to unambiguously detect and/or validate genotypes with a homozygous deletion at the CNV. Twenty subjects for each copy-number genotype, predicted by CNAT with the initial reference set, were selected for experimental analyses. Reactions were performed on genomic DNA samples with a total of 20 $\mu$l reaction volume containing 10 $\mu$l 2 × Taq PCR Master Mix (TIANGEN Biotech, Beijing, China), 10 pmol forward and reverse primers, and 50 ng of genomic DNA. Primers were designed with Primer Express 5.0 software from Applied Biosystems (Forward primer 5′-GGCAAAACAGAAGCACATT-3′, Reverse primer 5′-AAAGCAAGAAAAGGAAGAATGA-3′). The

incubation was carried out at 94°C for 5 min, followed by 30 cycles at 94°C for 30 s, 60°C for 1 min, and 72°C for 3 min in an Eppendorf Mastercycler (Eppendorf AG, Hamburg, Germany). All PCR products were examined by electrophoresis on 1.5% agarose and TBE gels and photographed with an Alpha Innotech Imager (Alpha Innotech, San Leandro, CA, USA). We found that the locus 4q13.2, inferred to have two copy numbers by CNAT, actually represented a homozygous deletion (see Results).

*Reanalysis of CNV 4q13.2 by CNAT with New Reference Set*

On the basis of the results of the PCR and electrophoresis experiments, we needed to reanalyze CNV 4q13.2 using CNAT for the following reasons. Estimation of raw copy numbers from SNP-mapping array data is based on the ratio of SNP probe-set signal intensity for each test sample versus the average for a reference set, which is assumed to have a copy number of two at all autosomal locations. Within a large reference set, the average copy number for most regions of human chromosomes is likely to be close to two. However, when a large reference set includes dominantly frequent polymorphisms in a region (such as deletions in this study), results of CNAT analyses for CNVs in this region might be skewed.[35] During PCR and electrophoresis experiments, we found a high frequency of homozygous deletions (75.4%) in the reference set. Consequently, it is likely that average intensity values for the reference set were too low, with the result that relatively high inaccuracy rates might be obtained in inferring heterozygous deletions and homozygotes without deletion (i.e., having two copies of the CNV), as recently shown.[35] Therefore, we used all subjects with a nonhomozygous deletion in the CNV at 4q13.2 and with sufficient average signal intensity for a reanalysis with CNAT as the reference set ($n = 216$) to recall CNV genotypes in this region. The newly ascertained genotypes at the CNV were used in all subsequent analyses.

*Real-Time PCR*

According to the results of the above section, "Reanalysis of CNV 4q13.2 by CNAT with the New Reference Set," we selected 20 subjects predicted to have homozygous deletions, 20 subjects predicted to have heterozygous deletions, and all of the subjects ($n = 16$) predicted to be diploid to perform real-time PCR, in order to assess the statistical significance of differences in DNA-amplification rates between groups with different copy numbers. The amplification rate is highly correlated with the copy numbers at the CNV. The primers were the same as those described above. Reactions were conducted in a 384 plate with the ABI 7900HT Sequence Detector system (Applied Biosystems, Foster City, CA, USA). Amplicons were designed against the putative altered locus and a control locus (Actin), which was used for controlling differences in genomic-DNA purity and concentration of different samples. PCR was performed in a 20 μl reaction volume containing 10 μl Power SYBR-Green PCR Master Mix (Applied Biosystems), 10 pmol forward and reverse primers, and 50 ng of genomic DNA. The reaction cycling conditions were 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and at 60°C for 1 min. Sequence Detection Software (SDS) was used for exporting the threshold cycle (Ct) data and further analyzing differences in Ct values (ΔCt) between the test locus and the control locus. For groups predicted to have different copy numbers by CNAT analyses, a t test, with the significant threshold defined by $p < 0.05$, was used for comparing ΔCt values to determine the statistical significance of these predicted copy-number differences.

## Fine Mapping for CNV 4q13.2

With the Affymetrix Human Mapping 500K Array, CNV 4q13.2 contained three SNPs, namely rs4260611 (position: 69,067,201), rs4492065 (position: 69,072,839), and rs4860308 (position: 69,172,267). According to the UCSC Human Genome Browser, the *UGT2B17* gene is located within this region. As a result of a lack of direct DNA-sequence data, the positions of rs4260611 and rs4860308 were not the exact boundaries of CNV 4q13.2. Consequently, for the purpose of fine mapping and identification of the susceptible gene(s) inside the CNV, we extended the CNV boundaries to the neighboring SNPs, i.e., rs1730872 (next to rs4260611) and rs293430 (next to rs4860308), to be inclusive. The interval between these two SNPs is the maximal region that will cover this CNV (physical position: 68,871,643–69,625,838). Four additional protein-coding genes, named YTH domain containing 1 (*YTHDC1*); transmembrane protease, serine 11E (TMPRSS11E [MIM 610399]); transmembrane protease, serine 11E2 (*TMPRSS11E2*); and UDP glucuronosyltransferase 2 family and polypeptide B15 (*UGT2B15* [MIM 600069]), are localized within this region (Figure 2BII). In order to determine whether these genes were involved in this CNV, we used the same genomic-DNA samples as those described in the "Real-Time PCR" section in the Supplemental Material and Methods section of the Supplemental Data to conduct separate real-time PCR experiments for the estimation of copy numbers for the above five genes (primer sequences are presented in Table S1). Because a high level of sequence identity exists between *UGT2B17* and *UGT2B15*, we used gene-specific markers described by Wilson et al.[34] to distinguish *UGT2B17* and *UGT2B15* (marker C: within exon 1 of *UGT2B17*; marker G: within exon 1 of *UGT2B15*).

Subsequently, in order to further validate *UGT2B17* gene deletion, we selected three additional markers described by Wilson et al,[34] which were marker D, located in the 5′ upstream region of *UGT2B17*; marker E, located within exon 6 of *UGT2B17*; and a deletion marker, J, for *UGT2B17* (see Figure S2) for performance of PCR amplification (PCR procedure described in the Supplemental Material and Methods section of the Supplemental Data) on the same subjects used in the above real-time PCR experiments. The ability to amplify markers D and E indicated the presence of *UGT2B17*. The ability to amplify marker J indicated the absence of the genomic region covering *UGT2B17*.

## Replication of CNV's Association with OF

To validate the association between the CNV of *UGT2B17* and OF, we genotyped the *UGT2B17* gene copy number, by real time PCR, using *UGT2B17* gene-specific marker C, in the Chinese replication sample. The reaction conditions were the same as those described in the above "Real-Time PCR" section. Then, we used marker J to perform PCR and electrophoresis experiments in order to validate and differentiate subjects with one or two copy numbers. Association analysis was conducted via the same procedure as that described in the above "Analysis of Association between CNV and OF in the GWA Sample" section. Because this test was for hypothesis-based confirmation, a one-sided p value $< 0.05$ was considered significant.

## Association Analyses between CNV and BMD or Bone Geometry in Chinese Unrelated Sample

To further explore the relationship between *UGT2B17* gene copy number and osteoporosis risk, we performed association analyses between this CNV and BMD or bone geometry in the Chinese unrelated sample. The *UGT2B17* gene copy number was genotyped via the same procedure as that described in the above "Replication of CNV's Association with OF" section. The significant

($p < 0.05$) terms of height, weight, sex, age, age$^2$, and time since menopause (for postmenopausal females) were used as covariates in multiple-regression analyses for the adjustment of derived BMD, BR, and CT values. Association analyses were performed for the adjusted residues by ANOVA in Minitab software (Minitab, State College, PA, USA). A one-sided p value $< 0.05$ was considered significant.

### Population Differentiation of CNV

Previous studies suggested that some CNVs might have elevated population differentiation.[2,6] In order to test potential ethnic differentiation and associated effects of the CNVs that were significantly associated with osteoporosis in Chinese subjects, we calculated copy numbers for these CNVs with CNAT in the 1000 white subjects, using genotype data from the Affymetrix Human Mapping 500K Array.[36] Three hundred out of the successfully genotyped 1000 white subjects were randomly selected as an initial reference set, and all 1000 subjects were screened by CNAT with the use of this reference set. When a subject in the reference set was screened by CNAT for CNV, however, he or she was excluded from the reference set. Using the protocol for CNAT-based CNV analyses that we used in the Chinese GWA sample, we consistently performed reanalysis by using an updated reference set including 230 subjects with two copy numbers. We subsequently performed association analyses between the CNVs and BMD and/or bone geometry parameters (all tested and adjusted for covariates of height, weight, sex, age, age$^2$, and time since menopause [for postmenopausal females only] by multiple regression) in white individuals by ANOVA to examine whether they had consistent effects on risk for osteoporosis. The significance level was defined as a one-sided p value $< 0.05$.

### Relationships between Serum Sex Hormones and *UGT2B17* Gene Copy Number

The *UGT2B17* gene encodes a key enzyme responsible for glucuronidation of androgens and their metabolites in humans. Androgen is a major source for estrogen, and both have direct effects in stimulating bone formation. Therefore, it is reasonable to hypothesize that *UGT2B17* influences serum androgen or estrogen concentrations, specifically through gene-dosage changes via CNVs. To assess this hypothesis, we examined the relationship between serum total testosterone or estradiol levels and *UGT2B17* gene copy number, respectively. Two hundred and thirty six unrelated Chinese young male subjects aged 22–28 yr (mean 23.6 ± 1.8) were recruited. After the subjects fasted overnight, 4 ml blood samples were collected from them at roughly 9 a.m. Serums were prepared from clotted blood and stored at −80°C until use. Serum testosterone and estradiol concentrations were quantified by Iodine [$^{125}$I] Testosterone Radioimmunoassays (RIA) Kit (Tianjing Nine Tripods Medical & Bioengineering, China; limit of detection 2.1 pg/ml, intra-assay CV 7%, inter-assay CV 8%) and Iodine [$^{125}$I] Estradiol RIA kit (Tianjing Nine Tripods Medical & Bioengineering, China; limit of detection 2.1 pg/ml, intra-assay CV 7%, inter-assay CV 9%), respectively. All experimental procedures were conducted according to the manufacturer's instructions.

*UGT2B17* gene copy number was genotyped by the same procedure as described in the section entitled "Replication of CNV's Association with OF." Differences in serum sex-steroid levels between the different genotypes (*UGT2B17* homozygous deletion versus one or two copy numbers) were compared in Minitab software with a Student's t test.

## Results

The basic characteristics of all sample sets are summarized in Table 1. In testing for potential within-population stratification of our Chinese GWA sample and randomly selected white unrelated sample, the STRUCTURE program showed that all subjects within each population were clustered together as one homogeneous sample. The estimated λ values were 1.02 and 1.01 for our Chinese GWA sample and white unrelated sample, respectively. These results indicate that there is no detectable significant population stratification within our Chinese and white samples.

### Genome-wide CNV Distribution in the Chinese GWA Sample

Excellent SNP call rates were obtained for DNA samples with the use of the 500K SNP mapping array. The final average genotyping call rate of the entire sample is 99.02%, which demonstrates high-quality allele-intensity data in this study. We used CNVRs to construct the genomic CNV map. By combining all of the individual CNV data analyzed by CNAT in the Chinese sample, we assembled a genomic map consisting of 727 CNVRs (Table S2) with an average and median size of 349.4 Kb and 223.6 Kb, respectively, covering 8.5% (254 Mb) of the human genome. Figure 1 shows the genomic distribution of these 727 CNVRs, of which 439 were singletons (marked by green lines) and 74 had CNVs with frequencies exceeding 1% (marked by red lines).

### Association Analyses of Common CNVs in the Chinese GWA Sample

In order to detect the distribution difference between cases and controls, we redefined 116 common CNVs (Figure S1) with >1% frequency from these 74 CNVRs for performance of the Mann-Whitney U test. We observed that CNV 4q13.2 had a significant distribution difference. According to the initial CNAT analyses, CNV 4q13.2 had three genotypes (Figure 2AI), namely, individuals with four, three, and two copy numbers. Results of PCR and electrophoresis experiments, however, showed that the CNV locus at 4q13.2 inferred to have two copy numbers by CNAT analyses with the initial reference data set, actually represented a homozygous deletion (Figure 2AII). Thus, we reanalyzed CNV 4q13.2 in the total sample with CNAT (See the Subjects and Methods section for details), and we concluded that there were 484 subjects (69.1%) with a homozygous deletion (Figure 2BIc), 200 subjects (28.6%) with a heterozygous deletion (Figure 2BIb), and 16 subjects (2.3%) with two copy numbers (Figure 2BIa). The newly ascertained genotypes at CNV 4q13.2 were then used in all subsequent analyses. Real-time PCR analyses further confirmed that the average ΔCt value for DNA samples with two copy numbers was significantly lower than that obtained from DNA with a heterozygous deletion at 4q13.2 ($p = 2.8 \times 10^{-5}$). This

**Table 1. Basic Characteristics of the Study Subjects**

| Trait | Sample 1 | | Sample 2 | | Sample 3 | Sample 4 | Sample 5 |
|---|---|---|---|---|---|---|---|
| | Case | Control | Case | Control | | | |
| Number | 350 | 350 | 399 | 400 | 689 | 1000 | 236 |
| Age (yr) | 69.35 (7.41) | 69.54 (6.09) | 67.32 (10.3) | 67.14 (8.34) | 52.45 (16.50) | 50.23 (18.24) | 23.56 (1.83) |
| Weight (kg) | 59.15 (12.05) | 59.61 (10.84) | 59.51 (9.23) | 61.11 (9.81) | 62.74 (12.10) | 80.16 (17.79) | 66.00 (9.54) |
| Height (cm) | 162.84 (8.31) | 159.41 (9.20) | 160.48(7.62) | 159.37 (7.41) | 161.90 (10.74) | 170.83 (9.74) | 172.06 (6.15) |
| Male/Female | 124/226 | 173/177 | 106/293 | 105/295 | 263/426 | 501/499 | 236/0 |
| Hip BMD | – | – | – | – | 0.878 (0.145) | 0.971 (0.154) | – |
| CT | – | – | – | – | 0.142 (0.029) | 0.156 (0.028) | – |
| BR | – | – | – | – | 12.24(2.81) | 12.09(2.35) | – |

Data are shown as mean (standard deviation). Abbreviations are as follows: BMD, bone-mineral density; CT, cortical thickness; BR, buckling ratio. Sample 1 corresponds to the Chinese GWA sample, Sample 2: to the Chinese replication sample, Sample 3 to the Chinese unrelated sample, Sample 4 to the white unrelated sample, and Sample 5 to the Chinese young male sample.

result supports the conclusion that the DNA copy number from samples inferred by CNAT as having two copies was indeed greater than the DNA copy number from samples inferred by CNAT as having one copy. DNA samples with homozygous deletion were observed as nonspecific amplification with weak signal intensity or failed amplification.

Table 2 lists the distribution of CNV 4q13.2 in the GWA sample. The prevalence of the deletion polymorphism in controls was concordant with that expected from the Hardy-Weinberg equilibrium (HWE) test ($p = 0.63$). The p value for the distribution difference between cases and controls was calculated to be $2.0 \times 10^{-4}$ [95% confidence interval (CI): $0–4.8 \times 10^{-4}$] on the basis of CNAT reanalyses with the updated reference set. This p value remained significant even after conservative Bonferroni correction for multiple testing by the number ($n = 116$) of CNVs examined (corrected $p = 0.023$). When all covariates were considered simultaneously, multivariate logistic-regression analysis indicated that the odds ratio (OR) was 1.73 (95% CI: 1.22–2.45, subjects with a heterozygous deletion and two copy numbers versus subjects with a homozygous deletion, $p = 0.001$). This result suggests that CNV in 4q13.2 was a significant predictor for the risk of hip OF, independent of the adjusted covariates (i.e., age, sex, height, and weight).

### Fine Mapping for CNV 4q13.2

Five protein-coding genes, including *UGT2B17*, *YTHDC1*, *TMPRSS11E*, *TMPRSS11E2*, and *UGT2B15*, are located within CNV 4q13.2 and its flanking region (Figure 2BII). In order to determine whether these genes were involved in this CNV, we conducted independent real-time PCR experiments to estimate copy numbers for each of these five genes. Gene copy-number deletion was only observed for *UGT2B17* (Table S1), indicating that *UGT2B17* is the only gene contained in CNV 4q13.2.

In order to further validate *UGT2B17* gene deletion, we selected three additional markers (marker D: within 5′ upstream region of *UGT2B17*, marker E: within exon 6 of *UGT2B17*, marker J: a deletion marker for *UGT2B17*) for the performance of PCR and electrophoresis analysis (Figure S2). The results of these experiments showed that the ability to amplify markers D and E was concordant with the amplification of marker C (within exon 1 of *UGT2B17*). Thus, subjects with two copies of *UGT2B17* had the ability to amplify markers C, D, and E but not marker J (Figure 2BIIIa), whereas subjects with a heterozygous deletion showed the ability to amplify markers C, D, E, and J simultaneously (Figure 2BIIIb). Subjects with a homozygous deletion had the ability to amplify only marker J and not C, D, or E (Figure 2BIIIc). These results indicate that there was a deletion polymorphism in the 150 kb interval spanning the whole *UGT2B17* gene (Figure S2).

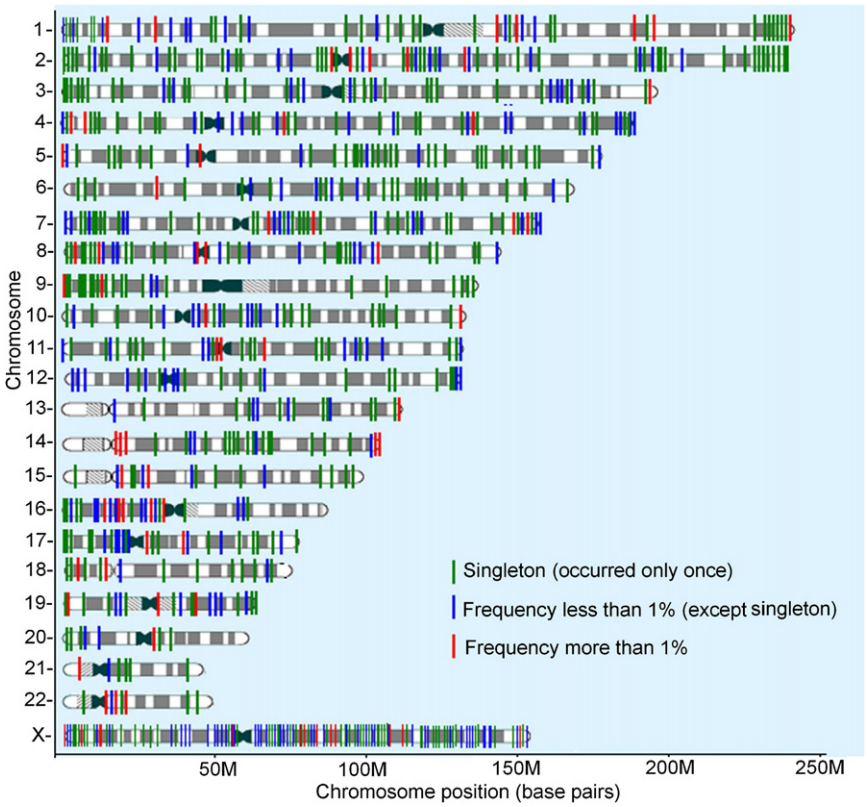


**Figure 1. Genomic Distribution of 727 CNVRs Identified in Chinese Subjects**
The approximate locations of all CNVRs are shown by lines. CNVRs observed in only one subject (singleton) are indicated in green, and CNVRs identified in multiple individuals are indicated in blue (<1% frequency) and red (>1% frequency).

### Replication of *UGT2B17* Gene Copy-Number Association with OF

In the Chinese replication sample, we detected 582 subjects (72.84%) with a homozygous deletion, 209 (26.16%) with a heterozygous deletion, and eight (1.0%) with two copy numbers (HWE test in control: $p = 0.19$; Table 2). Significant association was successfully replicated between *UGT2B17* gene copy number and OF ($p = 6.34 \times 10^{-3}$). A logistic-regression model consistently indicated an increased risk of OF in subjects with an existing *UGT2B17* gene (OR = 1.58, 95% C.I.: 1.12–2.22, $p = 0.01$).

### Association Analyses of *UGT2B17* Gene Copy Number in the Chinese Unrelated Sample

To further explore the relationship between *UGT2B17* CNV and osteoporosis risk, we genotyped *UGT2B17* gene copy number in the Chinese sample of 689 unrelated subjects, and we found 529 subjects (76.8%) with a homozygous deletion, 149 (21.6%) with a heterozygous deletion, and 11 (1.6%) with two copy numbers (HWE test: $p = 0.92$). Because the low frequency of two copy numbers of the *UGT2B17* gene was detected, we combined the subjects with two copy numbers and those with heterozygous deletion as one group for the performance of subsequent association analysis. Figure 3 lists the association results between BMD, FN bone geometry (CT and BR), and the CNV of *UGT2B17*. Compared to subjects with a homozygous deletion, subjects with one or two *UGT2B17* gene copy numbers had significantly lower BMD (3.4% lower, 0.855 g/cm$^2$ versus 0.885 g/cm$^2$, $p = 5.0 \times 10^{-4}$), thinner CT (3.5% lower, 0.138 cm versus 0.143 cm, $p = 0.003$), and a higher BR (5.5% higher, 12.8 versus 12.1, $p = 5.0 \times 10^{-4}$). The contribution of the CNV of *UGT2B17* to variations in BMD, CT, and BR was 1.43%, 1.08%, and 1.46%, respectively.

### Association Analyses of *UGT2B17* Gene Copy Number in the White Unrelated Sample

In the white sample of 1000 unrelated subjects, we detected 119 subjects (11.9%) with a homozygous deletion, 444 subjects (44.4%) with a heterozygous deletion, and 437 subjects (43.7%) with two copy numbers (HWE test: $p = 0.70$). We tested for the independence between genotype and ethnic background and found that CNV of *UGT2B17* had a different frequency distribution in Chinese subjects (unrelated sample: 76.8% had homozygous deletions, 21.6% had heterozygous deletions, and 1.6% had two copy numbers) than in white subjects (2 df, $p < 0.0001$). Consistent with the results obtained in the unrelated Chinese sample, higher *UGT2B17* copy numbers in the white sample were significantly associated with lower BMD (copy numbers 0 versus 1 versus 2: 0.986 g/cm$^2$ versus 0.973 g/cm$^2$ versus 0.964 g/cm$^2$, respectively; $p = 0.021$), thinner CT (0.159 cm versus 0.156 cm versus 0.152 cm; $p = 0.017$), and higher BR (11.71 versus 12.01 versus 12.20; $p = 0.012$) (Figure 3B). In the white sample, the contribution of the CNV of *UGT2B17* to variations in BMD, CT, and BR was 0.67%, 0.71%, and 0.77%, respectively. Compared to those with a homozygous deletion, subjects in whom the *UGT2B17* gene was present (including those with two and one copy numbers) had a 1.77% lower BMD, a 3.14% lower CT, and a 3.37% higher BR.

### Relationships between Serum Sex Hormones and *UGT2B17* Gene Copy Number

In the 236 unrelated Chinese young males studied, we detected 181 subjects (76.7%) with a homozygous deletion, 49 subjects (20.8%) with a heterozygous deletion, and six subjects (2.5%) with two copy numbers (HWE test: $p = 0.27$). Compared with those with one or two *UGT2B17* gene copy numbers, subjects with homozygous

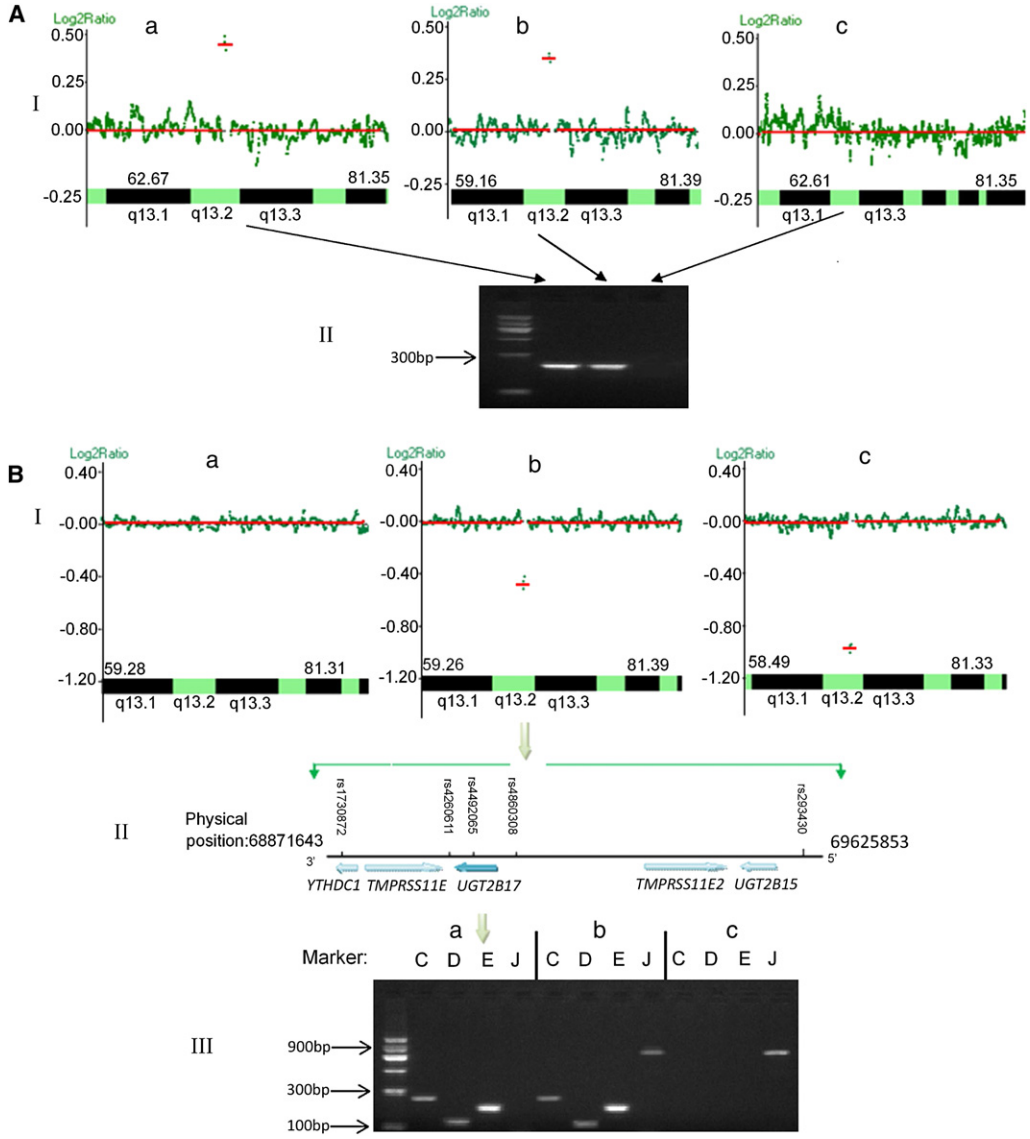


**Figure 2. Examples of Copy-Number Determination at 4q13.2 by CNAT Analyses with Affymetrix Human Mapping 500K Array Data and Electrophoresis Analyses**

A: CNV 4q13.2 calculated by CNAT analyses with the 350 unrelated control subjects used as the reference set.

AI: Three genotypes of the CNV 4q13.2: four copy numbers (a), three copy numbers (b), and two copy numbers (c). The x axis denotes genomic position. The y axis denotes the log2 ratio, which was calculated by comparison of the allele-intensity values of the test sample with those of the reference set.

AII: Electrophoresis analysis for PCR products (with deletion-specific primers) from the three genotypes at the CNV. Results for 4q13.2 inferred to have two copy numbers by CNAT analyses reveal no PCR products, indicating that this sample actually reflects a homozygous deletion.

B: CNV 4q13.2 calculated by CNAT with all of the subjects with nonhomozygous deletion used as the updated reference set.

BI: Three genotypes of CNV 4q13.2: two copy numbers (a), one copy number, (b) and zero copy numbers (c).

BII: On the basis of the UCSC Human Genome Brower, five genes are located in 4q13.2. The heavy arrow above "*UGT2B17*" means that the gene was found in CNV 4q13.2 with SNPs of rs4260611 and rs4860308 used as boundaries. When the boundaries were extended to SNPs of rs1730872 and rs293430 (the interval between these two SNPs was the maximal potential region for this CNV), four more protein-coding genes, named *YTHDC1*, *TMPRSS11E*, *TMPRSS11E2*, and *UGT2B15*, were found to be localized within this region (light arrows). CNV was only observed in *UGT2B17*.

BIII: Electrophoresis analysis for PCR products from gene-specific markers C (within exon 1 of *UGT2B17*), D (within 5′ upstream of *UGT2B17*), and E (within exon 6 of *UGT2B17*) and from deletion marker J of *UGT2B17* . Subjects with two copy numbers of *UGT2B17* had the ability to amplify markers C, D, and E but not marker J (a). Subjects with heterozygous deletion showed the ability to amplify markers C, D, E, and J simultaneously (b). Subjects with homozygous deletion had the ability to amplify only marker J (c). The results revealed that there was a deletion polymorphism in a 150 kb interval spanning the whole *UGT2B17* gene.

**Table 2. Distribution of *UGT2B17* Copy Number or CNV 4q13.2 in Two Chinese Case-Control Samples**

| Sample | Status | Copy Numbers | | | HWE | p Value (95% CI) | OR (95% CI) |
|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | | | |
| GWA sample | case | 220 | 119 | 11 | – | $2.0 \times 10^{-4}$ | 1.73 (1.22–2.45) |
| | control | 264 | 81 | 5 | 0.63 | $(0 - 4.8 \times 10^{-4})$ | |
| Replication sample | case | 275 | 119 | 5 | – | $6.3 \times 10^{-3}$ | 1.58 (1.12–2.22) |
| | control | 307 | 90 | 3 | 0.19 | $(5.9 \times 10^{-3} - 6.8 \times 10^{-3})$ | |

Fine-mapping result indicated that *UGT2B17* is the only gene contained in CNV 4q13.2. Abbreviations are as follows: GWA, genome-wide analysis; HWE, Hardy-Weinberg equilibrium; C.I., confidence interval; OR, odds ratio. p values were calculated via a nonparametric Mann-Whitney U test. Odds ratios were calculated with a logistic-regression model (copy number 0 versus copy numbers 1 and 2), with potential covariates such as height, weight, sex, age, and age$^2$ taken into account.

deletions of *UGT2B17* had significantly higher concentration levels of testosterone ($p = 0.005$) and estradiol ($p = 0.01$) (Table 3).

## Discussion

In this study, we generated a preliminary genome-wide map of CNVRs identified in Chinese subjects and found that nearly 80% of these CNVRs overlapped with previously reported CNVRs, according to the Database of Genomic Variants. Importantly, we discovered that CNV 4q13.2 was significantly associated with the risk of OF. This CNV was one of the CNVs that overlapped with those reported in previous studies, and it was located within a recombination "hotspot" region.[2,31,32,37] CNV 4q13.2 was common in our Chinese study samples; a high frequency of 76.8% was found in the Chinese unrelated subjects, and this frequency is consistent with the findings that Terakura et al. observed in an East Asian sample.[33] Confirmatory real time PCR experiments and replication study lent further support for the concept that the existence of the *UGT2B17* gene in CNV 4q13.2 might increase OF risk compared to a homozygous deletion in this CNV. Specifically, in both Chinese and white samples, the presence of the *UGT2B17* gene was significantly associated with lower BMD, thinner CT, and higher BR, all of which have been consistently identified as major risk factors for the development of OF. Our results consistently supported the important effect of the CNV of *UGT2B17* to the pathogenesis of osteoporosis.

CNVs might be the consequence of recurrent events via nonallelic homologous recombination, mediated by higher-order genomic structural features.[38] Copy-number-deletion variants can be subjected to increased selection pressures.[38] Our results revealed that CNV of *UGT2B17* had an ethnic differentiation, which lends strong support to previous findings.[2,6] Copy-number deletion of *UGT2B17* is much more common in Chinese subjects than in white subjects, and the differential frequency of this deletion in these two populations might contribute to the overall lower risk of OF in Chinese individuals than in white individuals. Despite this differential frequency of deletion, the relationship between CNV of *UGT2B17* and the risk of osteoporosis exhibited the same trend within both Chinese and white populations; i.e., higher gene copy numbers were associated with lower BMD, thinner CT, higher BR, and an increased risk of OF in both populations.

One possible mechanism by which CNVs cause phenotypic diversity is by alteration of the expression of copy-number-variant genes; i.e., dosage effects might underlie

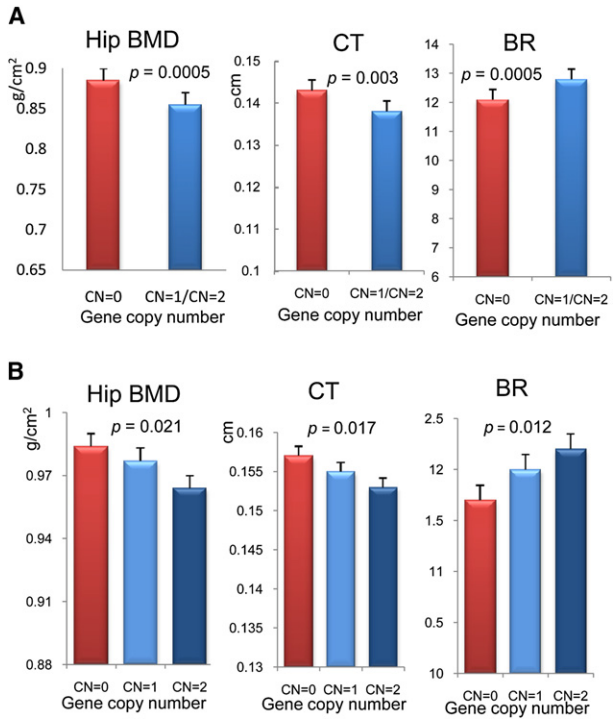


**Figure 3. Comparisons of Hip BMD, CT, and BR Values for CNVs of *UGT2B17* in Chinese and White Unrelated Samples**

(A) Chinese unrelated sample.

(B) White unrelated sample.

p values were estimated with the use of an ANOVA. For the Chinese unrelated sample, because a low frequency (1.6%) of two copy numbers of the *UGT2B17* gene were detected, we combined the subjects with two copy numbers and heterozygous deletions as one group to perform the association analysis. Error bars denote standard error. Abbreviations are as follows: BMD, bone mineral density; CT, cortical thickness; BR, buckling ratio.

**Table 3. Relationships between Serum Testosterone or Estradiol and *UGT2B17* Copy Number**

| | *UGT2B17* Copy Numbers | | |
|---|---|---|---|
| | 0 (181 subjects) | 1 or 2 (49 subjects) | p Value |
| Testosterone (μg/l) | 8.75 (2.93) | 7.62 (2.45) | 0.005 |
| Estradiol (ng/l) | 52.1 (17.0) | 46.4 (12.0) | 0.01 |

Data are shown as mean (standard deviation).

the biological effects of certain CNVs. The *UGT2B17* gene in CNV 4q13.2 encodes a protein belonging to the family of uridine diphospho (UDP)-glucuronosyltransferases enzymes, which play a key role in the homeostasis and metabolism of a number of endogenous molecules, including steroid hormones.[39,40] Chouinard et al.[41] determined that *UGT2B17* encodes a critical enzyme involved in the local inactivation of androgens; i.e., enhanced expression of *UGT2B17* inhibits the androgen-signaling pathway. Our serum-sex-hormones analysis supported this conclusion by showing that subjects with homozygous deletions of *UGT2B17* had significantly higher concentrations of total testosterone and estradiol than did subjects with one or two copies of this gene. Additionally, it has been well established that the steroid hormones, androgen and estrogen, have important roles in maintaining cancellous bone mass and integrity.[42] Androgen is a major source for estrogen and has a direct effect in stimulating bone formation. Consequently, enhanced expression of *UGT2B17*, via its effects on androgen, might increase the risk of osteoporosis by impairing bone formation during growth and/or favoring bone resorption over bone formation. This proposed physiological mechanism is consistent with our current findings that increasing dosage of *UGT2B17* incurred higher risk for OF, lower BMD, and poorer bone geometry.

It is notable that all the subjects in our Chinese sample were of the same Han ethnicity and from the same geographic area and that all of the case subjects experienced the same type of low-impact osteoporotic hip fractures. The homogeneity of our sample minimized or eliminated spurious CNVs and associations due to phenotypic heterogeneity and variation, copy-number polymorphisms in ethnically diverse populations, or other factors caused by population stratification.

It is important to recognize that estimation of raw copy numbers from SNP-mapping array data is based on the ratio of SNP probe-set signal intensity for each test sample versus a reference set. Thus, statistical software uses the average of the reference set to infer changes in copy number by relative duplication or deletion. A larger sample size for the reference set can improve the accuracy of copy-number computation.[35] Similarly, for a specific CNV, exclusion of subjects with homozygous deletions from the reference set can also improve the precision of copy-number inference, as a result, in part, of higher average signal intensities of the reference set. Of course, the results from these statistical inferences need to be validated by actual experimentation, as conducted in the current study with several methodological approaches (PCR and electrophoresis, as well as real-time PCR). Finally, due to concerns about the limitations of association analyses, we did not perform association analyses between OF and rare CNVs (i.e., CNVs with apparent frequencies <1%) in the current study, given that the results, even if significant, might still occur by chance, i.e., by sampling error.[43]

In conclusion, our study revealed 727 CNVs in a Chinese Han population, though this result warrants continual refinement in the same population and ethnicity. We additionally determined that higher *UGT2B17* gene copy number was associated with lower BMD, thinner CT, higher BR, and increased risk of OF at the hip in both Chinese and white populations. However, focused molecular functional studies will be needed for confirmation and clarification of this potential mechanism. Although a genetic basis for osteoporosis has been well established, previous investigations have focused on the association between classical genotypes (such as SNPs) and osteoporosis. To the best of our knowledge, this is the first study implicating gene-dosage changes as a risk factor for osteoporosis.

## Supplemental Data

Supplemental Data include two figures and two tables and can be found with this paper online at http://www.ajhg.org/.

## Acknowledgments

We thank all participants for taking part in this research. Investigators of this project were funded in part by grant 30570875 from the National Natural Science Foundation of China, as well as by Xi'an Jiaotong University, the Ministry of Education of China, Huo Ying Dong Education Foundation, HuNan Province, and Hunan Normal University. H.W.D. was supported by a Dickson and Missouri endowment from University of Missouri-Kansas City and by grants from the NIH (R01 AR050496, R21 AG027110, R01 AG026564, P50 AR055081, and R21 AA015973).

Received: August 6, 2008
Revised: October 1, 2008
Accepted: October 9, 2008
Published online: November 6, 2008

## Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants, http://projects.tcag.ca/variation/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
UCSC Human Genome Browser, http://genome.ucsc.edu/cgi-bin/hgGateway

## References

1. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.

2. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454.
3. Lupski, J.R. (2007). Structural variation in the human genome. N. Engl. J. Med. *356*, 1169–1171.
4. Buckland, P.R. (2003). Polymorphically duplicated genes: their relevance to phenotypic variation in humans. Ann. Med. *35*, 308–315.
5. Nguyen, D.Q., Webber, C., and Ponting, C.P. (2006). Bias of selection on human copy-number variants. PLoS. Genet. *2*, e20.
6. McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J., et al. (2006). Common deletion polymorphisms in the human genome. Nat. Genet. *38*, 86–92.
7. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.
8. Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M., Gough, S.C., de Smith, A., Blakemore, A.I., et al. (2007). FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat. Genet. *39*, 721–723.
9. Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E., et al. (2006). Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature *439*, 851–855.
10. Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science *307*, 1434–1440.
11. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.
12. Friedman, J.M., Baross, A., Delaney, A.D., Ally, A., Arbour, L., Armstrong, L., Asano, J., Bailey, D.K., Barber, S., Birch, P., et al. (2006). Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am. J. Hum. Genet. *79*, 500–513.
13. Nelson, D.A., Barondess, D.A., Hendrix, S.L., and Beck, T.J. (2000). Cross-sectional geometry, bone strength, and bone mass in the proximal femur in black and white postmenopausal women. J. Bone Miner. Res. *15*, 1992–1997.
14. Slemenda, C.W., Turner, C.H., Peacock, M., Christian, J.C., Sorbel, J., Hui, S.L., and Johnston, C.C. (1996). The genetics of proximal femur geometry, distribution of bone mass and bone mineral density. Osteoporos. Int. *6*, 178–182.
15. Liu, Y.J., Shen, H., Xiao, P., Xiong, D.H., Li, L.H., Recker, R.R., and Deng, H.W. (2006). Molecular genetic studies of gene identification for osteoporosis: a 2004 update. J. Bone Miner. Res. *21*, 1511–1535.
16. Deng, H.W., Mahaney, M.C., Williams, J.T., Li, J., Conway, T., Davies, K.M., Li, J.L., Deng, H., and Recker, R.R. (2002). Relevance of the genes for bone mass variation to susceptibility to osteoporotic fractures and its implications to gene search for complex human diseases. Genet. Epidemiol. *22*, 12–25.
17. Michaelsson, K., Melhus, H., Ferm, H., Ahlbom, A., and Pedersen, N.L. (2005). Genetic liability to fractures in the elderly. Arch. Intern. Med. *165*, 1825–1830.
18. Styrkarsdottir, U., Halldorsson, B.V., Gretarsdottir, S., Gudbjartsson, D.F., Walters, G.B., Ingvarsson, T., Jonsdottir, T., Saemundsdottir, J., Center, J.R., Nguyen, T.V., et al. (2008). Multiple Genetic Loci for Bone Mineral Density and Fractures. N. Engl. J. Med. *358*, 2355–2365.
19. Kiel, D.P., Demissie, S., Dupuis, J., Lunetta, K.L., Murabito, J.M., and Karasik, D. (2007). Genome-wide association with bone mass and geometry in the Framingham Heart Study. BMC Med. Genet. *8* (*Suppl 1*), S14.
20. Richards, J.B., Rivadeneira, F., Inouye, M., Pastinen, T.M., Soranzo, N., Wilson, S.G., Andrew, T., Falchi, M., Gwilliam, R., Ahmadi, K.R., et al. (2008). Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet *371*, 1505–1512.
21. Liu, Y.Z., Wilson, S.G., Wang, L., Liu, X.G., Guo, Y.F., Li, J., Yan, H., Deloukas, P., Soranzo, N., Chinnapen-Horsley, U., et al. (2008). Identification of PLCL1 gene for hip bone size variation in females in a genome-wide association study. PLoS. ONE *3*, e3160.
22. Komura, D., Shen, F., Ishikawa, S., Fitch, K.R., Chen, W., Zhang, J., Liu, G., Ihara, S., Nakamura, H., Hurles, M.E., et al. (2006). Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. Genome Res. *16*, 1575–1584.
23. Deng, H.W., Xu, F.H., Huang, Q.Y., Shen, H., Deng, H., Conway, T., Liu, Y.J., Liu, Y.Z., Li, J.L., Zhang, H.T., et al. (2002). A whole-genome linkage scan suggests several genomic regions potentially containing quantitative trait Loci for osteoporosis. J. Clin. Endocrinol. Metab. *87*, 5151–5159.
24. Duan, Y., Beck, T.J., Wang, X.F., and Seeman, E. (2003). Structural and biomechanical basis of sexual dimorphism in femoral neck fragility has its origins in growth and aging. J. Bone Miner. Res. *18*, 1766–1774.
25. Beck, T.J., Stone, K.L., Oreskovic, T.L., Hochberg, M.C., Nevitt, M.C., Genant, H.K., and Cummings, S.R. (2001). Effects of current and discontinued estrogen replacement therapy on hip structural geometry: the study of osteoporotic fractures. J. Bone Miner. Res. *16*, 2103–2110.
26. Xiong, D.H., Shen, H., Xiao, P., Guo, Y.F., Long, J.R., Zhao, L.J., Liu, Y.Z., Deng, H.Y., Li, J.L., Recker, R.R., et al. (2006). Genome-wide scan identified QTLs underlying femoral neck cross-sectional geometry that are novel studied risk factors of osteoporosis. J. Bone Miner. Res. *21*, 424–437.
27. Di, X., Matsuzaki, H., Webster, T.A., Hubbell, E., Liu, G., Dong, S., Bartell, D., Huang, J., Chiles, R., Yang, G., et al. (2005). Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. Bioinformatics *21*, 1958–1963.
28. Rabbee, N., and Speed, T.P. (2006). A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics *22*, 7–12.
29. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
30. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics *55*, 997–1004.
31. de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., et al. (2007). Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males:

implications for association studies of complex diseases. Hum. Mol. Genet. *16*, 2783–2794.
32. Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., et al. (2007). The diploid genome sequence of an individual human. PLoS Biol. *5*, e254.
33. Terakura, S., Murata, M., Nishida, T., Emi, N., Akatsuka, Y., Riddell, S.R., Morishima, Y., Kodera, Y., and Naoe, T. (2005). A UGT2B17-positive donor is a risk factor for higher transplant-related mortality and lower survival after bone marrow transplantation. Br. J. Haematol. *129*, 221–228.
34. Wilson, W. III, Pardo-Manuel de Villena, F., Lyn-Cook, B.D., Chatterjee, P.K., Bell, T.A., Detwiler, D.A., Gilmore, R.C., Valladeras, I.C., Wright, C.C., Threadgill, D.W., et al. (2004). Characterization of a common deletion polymorphism of the UGT2B17 gene linked to UGT2B15. Genomics *84*, 707–714.
35. Baross, A., Delaney, A.D., Li, H.I., Nayar, T., Flibotte, S., Qian, H., Chan, S.Y., Asano, J., Ally, A., Cao, M., et al. (2007). Assessment of algorithms for high throughput detection of genomic copy number variation in oligonucleotide microarray data. BMC Bioinformatics *8*, 368–385.
36. Liu, Y.J., Liu, X.G., Wang, L., Dina, C., Yan, H., Liu, J.F., Levy, S., Papasian, C.J., Drees, B.M., Hamilton, J.J., et al. (2008). Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. Hum. Mol. Genet. *17*, 1803–1813.
37. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet. *38*, 75–81.
38. Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., et al. (2006). Copy number variation: new insights in genome diversity. Genome Res. *16*, 949–961.
39. Chouinard, S., Pelletier, G., Belanger, A., and Barbier, O. (2004). Cellular specific expression of the androgen-conjugating enzymes UGT2B15 and UGT2B17 in the human prostate epithelium. Endocr. Res. *30*, 717–725.
40. Beaulieu, M., Levesque, E., Hum, D.W., and Belanger, A. (1996). Isolation and characterization of a novel cDNA encoding a human UDP-glucuronosyltransferase active on C19 steroids. J. Biol. Chem. *271*, 22855–22862.
41. Chouinard, S., Barbier, O., and Belanger, A. (2007). UDP-glucuronosyltransferase (UGT)2B15 and UGT2B17 enzymes are major determinants of the androgen response in prostate cancer LNCaP cells. J. Biol. Chem. *282*, 33466–33471.
42. Vanderschueren, D., Vandenput, L., Boonen, S., Lindberg, M.K., Bouillon, R., and Ohlsson, C. (2004). Androgens and bone. Endocr. Rev. *25*, 389–425.
43. McCarroll, S.A., and Altshuler, D.M. (2007). Copy-number variation and association studies of human disease. Nat. Genet. *39*, S37–S42.