Identification of a Copy Number Variation on Chromosome 20q13.12 Associated with Osteoporotic Fractures in the Korean Population

Tae-Joon Park§, Mi Yeong Hwang§, Sanghoon Moon, Joo-Yeon Hwang, Min Jin Go, Bong-Jo Kim*

Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Cheongju 28159, Korea

Osteoporotic fractures (OFs) are critical hard outcomes of osteoporosis and are characterized by decreased bone strength induced by low bone density and microarchitectural deterioration in bone tissue. Most OFs cause acute pain, hospitalization, immobilization, and slow recovery in patients and are associated with increased mortality. A variety of genetic studies have suggested associations of genetic variants with the risk of OF. Genome-wide association studies have reported various single-nucleotide polymorphisms and copy number variations (CNVs) in European and Asian populations. To identify CNV regions associated with OF risk, we conducted a genome-wide CNV study in a Korean population. We performed logistic regression analyses in 1,537 Korean subjects (299 OF cases and 1,238 healthy controls) and identified a total of 8 CNV regions significantly associated with OF (p < 0.05). Then, one CNV region located on chromosome 20q13.12 was selected for experimental validation. The selected CNV region was experimentally validated by quantitative polymerase chain reaction.

The CNV region of chromosome 20q13.12 is positioned upstream of a family of long non-coding RNAs, LINC01260. Our findings could provide new information on the genetic factors associated with the risk of OF.

Keywords: DNA copy number variations, genome-wide association study, osteoporotic fracture, real-time polymerase chain reaction

Introduction

Osteoporotic fractures (OFs) result from a decrease in bone strength, which can be induced by low bone mass and microarchitectural deterioration in bone tissue [1]. OFs are the critical hard outcome of osteoporosis, a disease that affects more than 75 million people in the United States, Europe, and Japan. With a lifetime fracture risk of 30% to 40% (vertebral or non-vertebral fractures), osteoporosis has an incidence rate similar to that of coronary artery disease. Furthermore, with the exception of forearm fractures, OFs are associated with increased mortality. Most fractures cause acute pain and lead to patient hospitalization, immobilization, and slow recovery [2-4].

Genetic studies have suggested a correlation of genetic variations with OF and its related traits. First, a candidate gene study has reported that common non-synonymous variants in low-density lipoprotein receptor-related protein 5 (LRP5) are consistently associated with bone mineral density (BMD) and fracture risk [5]. Furthermore, meta-analyses of genome-wide association studies (GWASs) have identified single-nucleotide polymorphisms (SNPs) located in more than 56 loci independently associated with BMD, and some of these studies have also found associations with fracture risk [6-13].

Copy number variations (CNVs) also have shown associations with OF. A genome-wide CNV study performed in a Chinese population identified a deletion variant of UDP glucuronosyltransferase 2 family, polypeptide B17 (UGT2B17) in chromosome 4q13.2 associated with OF [14]; however, this variant was not replicated in a study of Caucasian
women [15], possibly showing ethnic specificity. Additionally, a rare deletion variant located on chromosome 6p25.1 showed an association with the risk of OF in a Dutch population [16].

Particularly, chromosome 20q13.12 is known as an enriched region of histone modifications. Many genes located on chromosome 20q13.12 have enriched levels of histone modifications, such as di-methylated lysine 4 of histone 3 (H3K4me2) and tri-methylated lysine 4 of histone 3 (H3K4me3), in their promoter regions [17]. H3K4me2 and H3K4me3 regulate the expression of the Runx2 gene, which encodes a transcription factor controlling bone development and osteoblast differentiation [18]. Although there is no previous study on an association between chromosome 20q13.12 and bone-related traits, an SNP in another 20q region (rs4811196 in CTNBL1, located on 20q11.23) has been reported to be associated with BMD [19].

In this study, we conducted an association study in the Korean population to identify new susceptibility loci for the risk of OF. We identified 8 CNV regions associated with the risk of OF through a statistical analysis. Among the 8 regions, we selected and validated the existence of one CNV by quantitative PCR.

Methods

Study subjects and diagnostic criteria for OF

A total of 10,038 participants from the Korea Association Resource (KARE) were recruited. Among the recruited individuals, 1,537 subjects, consisting of 299 OF cases and 1,238 normal controls, who agreed to an X-ray examination forOF were recruited. The clinical characteristics of study subjects are summarized in Table 1. We used written informed consent was provided to all participants, and the study was approved by the Institutional Review Board of the Korea Centers for Disease Control and Prevention, and written informed consent was provided to all participants, including cases and controls. The clinical characteristics of the study participants are summarized in Table 1. We used

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Case</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subjects</td>
<td>299</td>
<td>1,238</td>
<td>1,537</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>83/216</td>
<td>332/906</td>
<td>415/1,122</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>59.7 ± 7.1</td>
<td>60.5 ± 6.8</td>
<td>60.3 ± 6.8</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>59.6 ± 10.1</td>
<td>59.3 ± 9.2</td>
<td>59.4 ± 9.4</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>156 ± 8.9</td>
<td>155.4 ± 8.1</td>
<td>155.5 ± 8.3</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.5 ± 3.4</td>
<td>24.6 ± 3.2</td>
<td>24.5 ± 3.3</td>
</tr>
</tbody>
</table>

Values are presented as number or mean ± SE. BMI, body mass index; SE, standard error.

Reference:

CNVtools. CNV genotypes consist of three genotype groups: “single-class,” “multiple-class,” and “ambiguous” (Supplementary Fig. 1) [24, 25]. We performed an association study with only well-clustered “multiple-class” CNV regions.

**Statistical analysis**

Logistic regression analysis was performed using R package version 3.0.2 to find significant association signals between CNV loci and OF risk. Age, height, and weight were adjusted as covariates to calibrate the statistical significance. Genotypes of each CNV were coded as 0 (homozygous deletion) and 1 (heterozygous deletion) for two-class CNV loci. In the case of three-class CNVs, their genotypes were coded as 0 (homozygous deletion), 1 (heterozygous deletion), and 2 (normal).

**Validation of CNVs**

We carried out quantitative PCR using the TaqMan Copy Number Assay (Life Technologies, Foster City, CA, USA) according to the manufacturer’s protocols to verify estimated CNV regions. In total, seven pre-designed and one custom designed probe was used to validate the existence of the CNV. Validation samples, including cases and controls, were randomly selected from each estimated genotype cluster. All validation experiments were replicated 3 times to increase the accuracy of the validation. The copy number in each individual was calculated with Copy Caller v2.0 using the comparative threshold cycle (CT) method according to the manufacturer’s protocols.

**Results**

**Characteristics of study subjects**

A total of 1,537 Korean individuals, consisting of 299 OF patients and 1,238 normal controls, participated in this study (Table 1). There were 415 males and 1,122 females in the overall population. The ratio of females in controls was slightly higher than in the cases (70% in cases vs. 72% in controls, data not shown). Mean age, height, weight, and body mass index showed no significant differences between cases and controls.

**Selection of candidate CNV regions**

To select candidate CNV regions, we manually performed a visual inspection to classify well-clustered “multiple-class” regions into genotype groups. In total, 3,660 multiple-class CNV regions were selected for the association study. Of the 3,660 CNV regions, 518 CNV regions were two-class regions (one and two copies) and 3,142 were three-class regions (zero, one, and two copies) (Supplementary Fig. 1).

**Association analysis of CNV regions with OF risk**

We performed logistic regression analysis to evaluate associations between identified CNV regions and OF risk. As a result, we identified a total of eight CNV loci associated with OF ($p < 0.05$) (Table 2)—distributed on chromosome 4q13.1, 4q13.2, 9q34.2, 13q12.2, 14q24.3, 14q31.3, 20p13, and 20q13.12. The OF-associated loci were located in intergenic regions near SRIP1; TMPRSS11E; POLR1D; LOC100421611; a family of long non-coding RNAs (IncRNA), LINC01260; and genic regions of ABO, ELMSANI, and SIRPA. The base-pair lengths of the eight regions ranged from 1.9 kb to 18.6 kb. The CNVs in ABO and SIRPA included both exonic and intronic regions of each gene, whereas the CNV locus in ELMSANI included only intronic regions.

**Validation of a CNV region on chr20: 42739446–42741539**

We selected one CNV region located on chromosome 20q13.12—chr20: 42739446–42741539—for validation based on position and expected biological function. The allele frequencies of chr20: 42739446–42741539 are indicated in Supplementary Table 1 and Supplementary Fig. 2. First, we

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start*</th>
<th>End*</th>
<th>Length (bp)</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>59662401</td>
<td>59670658</td>
<td>8,258</td>
<td>Intergenic (nearby SRIP1)</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>69056562</td>
<td>69075975</td>
<td>19,414</td>
<td>Intergenic (nearby TMPRSS11E)</td>
<td>0.0084</td>
</tr>
<tr>
<td>9</td>
<td>135120477</td>
<td>135122444</td>
<td>1,968</td>
<td>ABO (intron and exon)</td>
<td>0.003</td>
</tr>
<tr>
<td>13</td>
<td>27153540</td>
<td>27157992</td>
<td>4,453</td>
<td>Intergenic (nearby POLR1D)</td>
<td>0.0054</td>
</tr>
<tr>
<td>14</td>
<td>73310052</td>
<td>73320816</td>
<td>10,765</td>
<td>ELMSANI (intron)</td>
<td>0.0078</td>
</tr>
<tr>
<td>14</td>
<td>84366904</td>
<td>84378906</td>
<td>12,003</td>
<td>Intergenic (nearby LOC100421611)</td>
<td>0.0058</td>
</tr>
<tr>
<td>20</td>
<td>1843096</td>
<td>1861664</td>
<td>18,569</td>
<td>SIRPA (intron and exon)</td>
<td>0.0016</td>
</tr>
<tr>
<td>20</td>
<td>42739446</td>
<td>42741539</td>
<td>2,094</td>
<td>Intergenic (nearby LINC01260)</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

CNV, copy number variation; OF, osteoporotic fracture; Chr, chromosome.

*Positions of each CNV region were defined based on hg18/NCBI build 36.
conducted in silico verification for the existence of the CNV region using web-based public data. Many previously reported CNVs, including the chr20: 42739446-42741539 region, were identified using the UCSC Genome Browser (http://genome.uscs.edu/cgi-bin/hgGateway) (Supplementary Fig. 3). Also, frequencies of reported CNVs evaluated in various ethnic groups were presented in the Database of Genomic Variants (DGV, http://dgv.tcg.ca/dgv/app/home) (Supplementary Table 1).

To experimentally verify the accuracy of the CNV, we performed quantitative PCR. The position of the CNV probe was defined based on hg19/NCBI build 37. The probe ID used for the TaqMan genotyping assay of the validated CNVs is listed in Table 3. Applications of a positive predictive value as a measurement standard of accuracy indicated the proportion of positive results in the validation of the CNV region. The positive predictive value of the verified CNV region evaluated in our validation was 0.82. Fig. 1 shows the genotype of the identified CNV region, verified by quantitative PCR.

**Discussion**

In this study, we performed a GWAS to investigate the associations between identified CNV regions and the risk of OF. A total of eight CNV loci were identified to be nominally associated with OF. For validation, we selected one CNV region based on position and expected biological meaning among the eight loci. The CNV, which is located on chromosome 20q13.12 (intergenic region near LINC01260) was experimentally validated through quantitative PCR.

An lncRNA is defined as a non-coding transcript having a length of more than 200 nucleotides [26] and is considered “junk” that is non-functional initially [27]. However, the functional roles of lncRNAs have now been identified, such as key regulators of transcriptional and translational products that affect cell functions [28-31]. IncRNAs play a role in epigenetic effects, such as histone modification (e.g., methylation), which regulates gene expression [32]. A significant histone modification, monomethylation of lysine 4 on histone H3 (H3K4me1) in human mammary epithelial cells, was predicted in our identified CNV region near an lncRNA, LINC01260, using University of California Santa Cruz (UCSC) Genome Browser (http://genome.uscs.edu/cgi-bin/hgGateway) (Supplementary Fig. 4). A previous genome-wide study has identified several runs of homozygosity (ROHs) associated with BMD and observed the enrichment of H3K4me1 in two ROHs, ROH15q22.3 and ROH1p31.1, by an analysis of histone modification marks in the Gm12878 cell line [33]. Taken together, our identified CNV region on 20q13.12 might have a putative effect on the risk of OF by histone modification of the

---

**Table 3. Results of CNV validation through TaqMan copy number assays**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Probe ID</th>
<th>CNV genotype</th>
<th>TP</th>
<th>FP</th>
<th>PPV&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43306032</td>
<td>43308125</td>
<td>Hs07172967_cn</td>
<td>0/1/2</td>
<td>84</td>
<td>18</td>
<td>0.82</td>
</tr>
</tbody>
</table>

CNV, copy number variation; TP, true-positive; FP, false-positive; PPV, positive predictive value.

<sup>a</sup>PPV is defined as number of TPs/(number of TPs + number of FPs).

---

**Fig. 1.** Validated genotypes of a 20q13.12 copy number variation (CNV) region in this study. Genotypes of a CNV region on chromosome 20q13.12. The copy number state of cases (left) and control samples (right). The higher bar, lower bar, and no bar in each figure represent a normal number of copies, heterozygous deletion, and homozygous deletion, respectively. The blue bar means the copy number state of the NA10851 sample, which was used as the reference sample.
promoter region of LINC01260.

In addition, a CNV on chromosome 20q13.12 is located in a region about 35 kb apart from the WISP2 gene. The WISP2 gene has some evidence of functions associated with bone-related traits. WISP2, also known as CCN family member 5 (CCN5) or connective tissue growth factor like (CTGF-L), was identified as a member of the CCN family, downstream of the WNT-β-catenin signaling pathway, using subtractive hybridization between WNT1-transformed mammary epithelial cells and parental cells [34]. The WNT-β-catenin signaling pathway has been suggested to be a major component in the accrual, regulation, and maintenance of bone mass [35]. WISP2 consists of 3 domains: an insulin-like growth factor binding domain, a von Willebrand Factor type C motif, and a thrombospondin type I repeat. The mRNA of WISP2 is expressed in primary human osteoblasts, fibroblasts, ovary, testes, and heart. WISP2 protein is secreted from primary human osteoblasts and fibroblasts [36]. A previous study has shown the relatedness of increased mRNA expression of WISP2 with rheumatoid arthritis and osteoarthritis [34]. Furthermore, the WISP2 gene has been reported as a candidate gene that is associated with an OF-related trait, BMD [37].

WISP2 expression is affected by the action of LRP5. LRP5 has been known to play a role in signaling through WNT-β-catenin and acts as a co-receptor with the frizzled family of receptors [38]. Previous research has reported that the LRP5 mutation (G171V), inducing the enhancement of structural strength and material properties of skeleton and bone mass, is correlated with increased expression of WISP2 [38]. Also, WISP2 is suggested to be involved in the frizzled receptors/LRP5/6 pathway by phosphorylation of LRP5 [39]. Considering the previously reported functions and interactions with LRP5 in the WNT-β-catenin signaling pathway, WISP2 could be an important component for bone formation or strength, which might affect the risk of OF.

In conclusion, we identified a total of 8 CNV regions associated with OF in the Korean population. Among the 8 CNVs, one locus, located on chromosome 20q13.12, was selected for validation and verified by quantitative PCR. Although further study in a larger number of study subjects is needed, to our knowledge, the current report is the first to investigate the association with the risk of OF. Our findings from this study could provide new insights into the genetic factors associated with OF risk.

Supplementary materials

Supplementary data including one table and four figures can be found with this article online at http://www.genominfo.org/src/sm/gni-14-216-s001.pdf.

Acknowledgments

This work was supported by intramural grants from the Korea National Institute of Health (2012-N73004-00 & 2012-N73005-00) and was provided with biospecimen data from the Korea Genome Analysis Project (4845-301), the Korean Genome and Epidemiology Study (4851-302), and the Korea Biobank Project (4851-307).

References

12. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson


