International Journal of Obesity (2017) **41,** 660–663

www.nature.com/ijo

SHORT COMMUNICATION

Whole-exome sequencing study reveals common copy number variants in protocadherin genes associated with childhood obesity in Koreans

S Moon[1], MY Hwang[1], HB Jang[2], S Han[1], YJ Kim[1], J-Y Hwang[1], H-J Lee[2], SI Park[2], J Song[2] and B-J Kim[1]

Recently, the prevalence of childhood obesity has significantly increased in industrialized countries, including Korea, and now controlling obesity is becoming an economic burden. However, knowledge of the risk factors associated with obesity is still limited. In this study, we aimed to discover additional obesity-associated loci in children. To achieve this, we conducted an exome-wide association analysis of copy number variation (CNV) using whole-exome sequencing (WES) data from a total of 102 cases and 86 controls. We newly identified a CNV locus that overlapped two protocadherin genes, PCDHB7 and PCDHB8, which are brain function-related genes (P-value $= 6.40 \times 10^{-4}$, odds ratio = 2.2189). A subsequent replication analysis using WES data from 203 obese and 291 normal weight children showed that this CNV region satisfied the genome-wide significance standard (Fisher's combined P-value $= 3.76 \times 10^{-5}$). Moreover, correlation test using 199 additional samples supported significant association between CNV and increased body mass index. This region also showed a meaningful association with 273 cases and 2596 controls in adult samples. Our findings suggest that differences in the common CNV region at 5q31.3 may have an impact on the pathophysiology of obesity.

International Journal of Obesity (2017) **41,** 660–663; doi:10.1038/ijo.2017.12

INTRODUCTION

Obesity is a medical condition that is defined as excessive fat accumulation. It is the result of combined contributions of environmental, behavioral and genetic factors.[1,2] Obesity leads to a reduced life expectancy and/or an increased risk of chronic diseases.[2] Moreover, about 70% of obese adolescents become obese adults.[3] In South Korea, the prevalence of childhood obesity also increased twofold, from 5.4% in 1998 to 10.8% in 2008, because of the fast-growing economy and rapid industrialization of this country.[4]

Recently, with great advances in genome-wide association studies, many genetic variants associated with obesity-related traits have been identified.[5,6] Although previous studies estimated that common variation accounts for $>20\%$ of body mass index (BMI) phenotypic variation, most of the genetic variability in BMI still remains undefined.[6,7]

In this study, we performed an exome-wide association study to examine obesity-susceptible copy number variation (CNV) between obese and normal weight children using whole-exome sequencing (WES) data. By this means, we identified a common CNV locus on PCDHB7-PCDHB8 at 5q31.3. The significance of this locus was replicated in WES data from independent samples. Also, association between CNV locus and increased BMI were highly underpinned by the correlation test in additional samples. Moreover, we investigated the effectiveness of this locus in predicting obesity using 2869 adults from community-based cohorts.

Our results indicated that genomic copy number differences that overlapped protocadherin genes were significantly associated with obesity in children and showed a possible association in adults.

MATERIALS AND METHODS

Discovery and replication

Subjects and phenotype. Study subjects, aged from 12 to 15 years old, were recruited from the Korean Children and Adolescents Obesity Cohort study. Informed parental consents of enrolled children were obtained. For categorization of subjects, we used definition of the Centers for Disease Control Prevention criteria for extreme obesity (a BMI-for-age of $\geqslant 1.2$ times the 95th percentile value or a BMI of $\geqslant 35\ \mathrm{kg\,m^{-2}}$).[8] Moreover, for obesity, we used the Western Pacific Regional Office of WHO criteria proposed an alternative cutoff points of a BMI $\geqslant 25\ \mathrm{kg\,m^{-2}}$ for Asian population[9] (Supplementary Table S1).

Whole-exome sequencing. WES data from 102 extreme obesity cases and 86 controls were used in the discovery stage. Moreover, those from 203 cases and 291 controls were used in the replication stage (Supplementary Table S2). An average sequencing depth of $\sim 67.2\times$ were sequenced on the Illumina HiSeq2500. We preprocessed WES data according to same procedures in our earlier study.[10]

[1]Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Republic of Korea and [2]Division of Metabolic Diseases, Center for Biomedical Sciences, National Institute of Health, Chungcheongbuk-do, Republic of Korea. Correspondence: Dr B-J Kim, Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, 187 Osongsaengmyeong 2ro osong-eup, heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28159, Korea.
E-mail: kbj6181@cdc.go.kr
Received 11 July 2016; revised 29 November 2016; accepted 18 December 2016; accepted article preview online 19 January 2017; advance online publication, 14 March 2017

Table 1. Summary of association results

CNV	Discovery			Replication		Correlation			Generalization		Fisher's combined P-value
	① Extreme obesity case-control			② Obesity case-control		③ Extreme + Obesity			④ Adult case-control		
	102 cases, 86 controls			203 cases, 291 controls		105 cases, 94 controls			273 cases, 2596 controls		
	P-value	Odds ratio	FDR	P-value	Odds ratio	Logistic regression P-value	Linear regression P-values		P-value	Odds ratio	(① + ②) (① + ② + ④)
							Case	Control			
chr5:140554655-140560171	6.40×10^{-4}	2.2189	0.0762	0.0174	1.4356	0.0381	2.0×10^{-16}	0.1055	0.054	1.156	3.76×10^{-5} 1.69×10^{-5}

Abbreviations: CNV, copy number variation; FDR, false discovery rate; QT, quantitative trait. Most CNV detection tools, including ExomeDepth, present CNV types, such as deletion (< 2 copies) and duplication (> 2 copies), rather than copy number, such as 0, 1, 2 and 3 copies. Thus, for association analysis, we encoded deletion, normal and duplication, to 1, 2 and 3, respectively.

CNV calling and statistical analysis. To perform CNV calling, we used the read depth-based exome CNV calling tool, ExomeDepth with the default parameters (Supplementary Figure S1(A), Supplementary Figure S2).[11] Using gender and age as covariates, a logistic regression analysis between obesity and CNV was undertaken with R package (http://www.r-project.org). To address multiple comparison issues, we calculated the Benjamini's false discovery rate. Also, the Fisher's combined probability test was used to combine results from discovery and replication using MADAM in the R package.

Validation, correlation and generalization

Validation. A total of 24 samples were randomly selected from the discovery stage. Two different platforms such as quantitative PCR (qPCR) and the Illumina HumanOmni2.5 BeadChip were used to evaluate estimated CNV genotypes (gain, loss and normal). A pre-designed probe (Assay ID: Hs02839484_cn) was used to qPCR experiment. For SNP array we defined the genomic position of each probe in the corresponding CNV region (Supplementary Figure S3).

Correlation test. Using additional 105 cases and 94 controls, we conducted a logistic regression analyses to examine significant differences between obesity and CNV (ΔΔCT). Moreover, liner regressions between BMI and ΔΔCT in each case and control group were calculated (Supplementary Table S2, Supplementary Figure S1(B)). Covariates such as age and sex were adjusted in regression analysis.

Generalization. To examine the effectiveness in adults, we used CNV genotypes from 4694 NimbleGen HD2 3 × 720K aCGH data used in our earlier study (Supplementary Figure S1(C)).[12] We performed a logistic regression between obesity and CNV adjusted for covariates using 273 cases (BMI ⩾ 30) and 2596 controls (BMI < 25).

RESULTS

Discovery and replication

Discovery stage. Only the CNV region overlapping PCDHB7-PCDHB8 (chr5:140554655–140560171, UCSC genome build hg19) showed great differences in frequency rates between cases and controls (CNV: $P=6.4\times10^{-4}$, odds ratio (OR) = 2.219, beta = 0.80/duplication: $P=3.8\times10^{-3}$, beta = 0.90/deletion: $P=7.3\times10^{-3}$, beta = − 1.30) (Table 1, Figure 1a, Supplementary Table S3 and Supplementary Table S4). Also, the *q*-value was 0.0762, indicating that only 7.6% of significance tests will result in false positives.

Replication stage. The significant CNV region successfully satisfied the statistical significance ($P=0.017$, OR = 1.4356). The meta-analysis on one-side *P*-values also satisfied a Bonferroni-adjusted *P*-value threshold of $P<0.05$ ($P=3.76\times10^{-5}$) (Table 1).

Validation, correlation and generalization

Validation. CNV genotypes of 24 samples were perfectly matched with those from the discovery stage (Figure 1b, c and e). We applied positive-predicted value as the measurement standard of accuracy. The mean positive-predicted value of the BeadChip and qPCR is 1.0 and 0.875, respectively (Supplementary Table S5).

Correlation test. Logistic regression analyses satisfied the statistical significance ($P=0.038$). Also, association between obesity and ΔΔCT in case was significant ($P<2\times10^{-16}$), whereas that of control was not significant ($P=0.1055$).

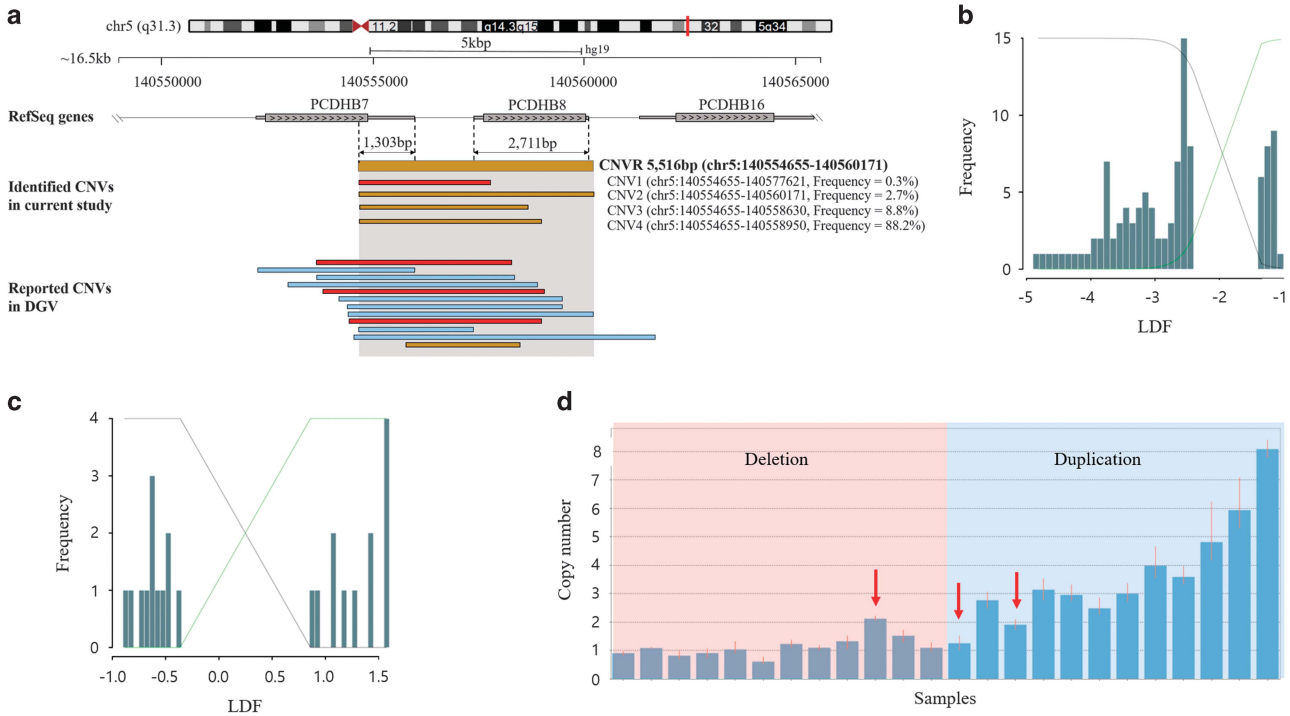

Figure 1. Results of the current study. (**a**) Goldenrod-filled rectangles indicate identified CNV regions. It consists of four sub-CNVs and spans one large exon from both PCDHB7 and PCDHB8. Goldenrod-filled rectangle represents a complex CNV genotype (a gain and a loss), whereas red- and sky blue-filled rectangle means a loss and a gain, respectively. RefSeq genes are shown as gray rectangles delineating single exons. The Database of Genomic Variants entries from other studies are shown at the bottom of the figure. (**b**) A histogram of each CNV genotype cluster of the discovery stage samples. Linear discriminant function (LDF) values from CNVtools were used for the clustering analysis. Three CNV genotypes were identified. (**c**) A histogram of each CNV genotype cluster of validation using the Illumina HumanOmni2.5 BeadChip. (**d**) Validation result using the Copy caller v2.0. Quantitative PCR results of 24 validation samples. Of 24 samples, only three samples (red arrows) have different CNV genotype from those of the discovery stage.

Generalization. The refined CNV region is ~4.3 kb (chr5: 140554875–140559203). And 39 consecutive CNV probes were targeted to within the boundaries of this region. From the logistic regression, it showed a suggestive trend ($P = 0.054$, OR = 1.156). Fisher's combined *P*-value of the three studies (discovery, replication and generalization) on one-side *P*-values was 1.69×10^{-5}.

DISCUSSION

Here, we have newly identified CNV region on PCDHB7-PCDHB8 gene that showed a notable significance in the obesity-control analysis of children. From the results, we suggest that the CNV difference on PCDHB7-PCDHB8 is more prevalent in severely obese children than in normal weight children. Particularly, duplication is positively associated with BMI, whereas deletion is inversely associated with BMI. Moreover, the effect of this region and its statistical significance are greater when the sampling population is highly obese and young. With regard to a well-known obesity-associated CNV (large and rare deletion) in 16p11.2, we found a deletion with ~525 kb in length (chr16:29675061-30215702).[13] However, this deletion was so rare that it was not found in the discovery stage. Moreover, there were only two 16p11.2 CNV carriers in the replication stage (one case and one control). *P*-values of discovery and replication were 0.48 and 0.98, respectively. Thus it is insufficient to examine the potential for clinical impact of 16p11.2 CNV using current data set.

Protocadherins are the largest subgroup in the cadherin superfamily and are predominantly expressed in the nervous system.[14] *PCDHB7* and *PCDHB8* genes encode potential calcium-dependent cell-adhesion proteins that may be involved in the establishment and maintenance of specific neuronal connections in brain (http://www.uniprot.org/uniprot). Also, Exome Aggregation Consortium demonstrated that brain-relavant genes including *PCDHB8* showed the greatest intolerance to gene dosage changes from CNVs.[15]

Recently, Mariman *et al.*[16] described a relationship between the variation in protocadherin genes and extreme obesity using WES data. They showed that rare variants of protocadherin genes were enriched at a higher frequency than the reference genes. For *PCDHB8*, structural changes, such as indels, were predicted to be damaging variants.[16] Moreover, several studies have suggested possible links between obesity and genetic variants at 5q31.3. A non-synonymous variant rs3749779 on *SLC25A2* associated with the obesity-related trait vitamin B12 deficiency was reported in Hispanic children.[17] Furthermore, Fox *et al.*[18] identified an intronic variant, rs31872, on *PCDHAC1* that was associated with visceral adipose tissue in men. Two SNPs, rs3749779 and rs31872, were located ~180 kb upstream and ~100 kb downstream from the CNV region, respectively (Supplementary Figure S4). In addition, Borghol *et al.* showed that a nearly 1 Mb region containing mainly protocadherins is more highly methylated in individuals with low socio-economic position during childhood, which is inversely associated with long-term weight gain.[19] Also, Balwierz *et al.*[20] suggested that high fat diet is assocated with *PCDHB7* gene regulation in obese mouse model. Given reported evidences, we hypothesized a possible function that gene dosage change caused by CNV may affect to change in neuronal connection called neuroplasticity, resulting in obesity.

We need more samples to increase the statistical power, but additional tests such as validation, replication and correlation test compensate for this limitation. Moreover, stringent selection criteria and multiple correction test were applied to minimize false positives. However, although the CNV genotype of

PCDHB7-PCDHB8 can vary, we simplified this genotype into three types, owing to the simple nature of the CNV calls. Thus, a CNV calling technique that can assign individuals to a precise CNV genotype is needed to more accurately assess genetic effects. In addition, functional investigations of the significant variant are needed, although we identified an exonic variant that may have deleterious effect.

To our knowledge we have identified a previously unreported common CNV region in the PCDHB7-PCHDB8 exons that is associated with the pathophysiology of obesity. Moreover, our data support previous evidences showing the importance of protocadherin gene families and the 5q31 region in obesity. Our findings may help improve the understanding of genetic variant contributions to obesity predisposition.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by intramural grants from the Korea National Institute of Health (2012-N73004-00, 2014-NI73001-00, 2012-E64001-00). Data were also provided by the Korean Genome Analysis Project (4845-301) and the Korean Genome and Epidemiology Study (4851-302) that were supported by the Korea Center for Disease Control and Prevention, Republic of Korea.

REFERENCES

1 Racette SB, Deusinger SS, Deusinger RH. Obesity: overview of prevalence, etiology, and treatment. *Phys Ther* 2003; **83**: 276–288.
2 Haslam DW, James WP. Obesity. *Lancet* 2005; **366**: 1197–1209.
3 Zhao J, Grant SF. Genetics of childhood obesity. *J Obes* 2011; **2011**: 845148.
4 Lim H, Xue H, Wang Y. Association between obesity and metabolic co-morbidities among children and adolescents in South Korea based on national data. *BMC Public Health* 2014; **14**: 279.
5 Fall T, Ingelsson E. Genome-wide association studies of obesity and metabolic syndrome. *Molec Cell Endocrinol* 2014; **382**: 740–757.
6 Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.
7 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
8 World Health Organization (WHO). *Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation on Obesity*. WHO: Geneva, 1998.
9 World Health Organization. Regional Office for the Western Pacific. *The Asia-Pacific Perspective: Redefining Obesity and its Treatment*. International Diabetes Institute, International Association for the Study of Obesity, International Obesity Task Force Publisher. Health Communications Australia Pty Limited: Sydney, New South Wales, Australia 2000.
10 Hwang MY, Moon S, Heo L, Kim YJ, Oh JH, Kim YK *et al.* Combinatorial approach to estimate copy number genotype using whole-exome sequencing data. *Genomics* 2015; **105**: 145–149.
11 Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012; **28**: 2747–2754.
12 Moon S, Jung KS, Kim YJ, Hwang MY, Han K, Lee JY *et al.* KGVDB: a population-based genomic map of CNVs tagged by SNPs in Koreans. *Bioinformatics* 2013; **29**: 1481–1483.
13 Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J *et al.* A new penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463**: 671–675.
14 Morishita H, Yagi T. Protocadherin family: diversity, structure, and function. *Curr Opin Cell Biol* 2007; **19**: 584–592.
15 Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE *et al.* Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet* 2016; **48**: 1107–1111.
16 Mariman EC, Bouwman FG, Aller EE, van Baak MA, Wang P. High frequency of rare variants with a moderate-to-high predicted biological effect in protocadherin genes of extremely obese. *Genes Nutr* 2014; **9**: 399.
17 Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS ONE* 2012; **7**: e51954.
18 Fox CS, Liu Y, White CC, Feitosa M, Smith AV, Heard-Costa N *et al.* Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet* 2012; **8**: e1002695.
19 Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* 2012; **41**: 62–74.
20 Balwierz A, Polus A, Razny U, Wator L, Dyduch G, Tomaszewska R *et al.* Angiogenesis in the New Zealand obese mouse model fed with high fat diet. *Lipids Health Dis* 2009; **8**: 13.

Supplementary Information accompanies this paper on International Journal of Obesity website (http://www.nature.com/ijo)