

## ORIGINAL ARTICLE

# Identification of a genetic locus on chromosome 4q34-35 for type 2 diabetes with overweight

Mi-Hyun Park<sup>1,2</sup>, Soo Heon Kwak<sup>3</sup>, Kwang Joong Kim<sup>1</sup>, Min Jin Go<sup>1</sup>, Hye-Ja Lee<sup>1</sup>, Kyung-Seon Kim<sup>1</sup>, Joo-Yeon Hwang<sup>1</sup>, Kuchan Kimm<sup>1</sup>, Young-Min Cho<sup>3</sup>, Hong Kyu Lee<sup>3</sup>, Kyong Soo Park<sup>3,4</sup> and Jong-Young Lee<sup>1</sup>

The incidence of type 2 diabetes is rising rapidly because of an increase in the incidence of being overweight and obesity. Identification of genetic determinants for complex diseases, such as type 2 diabetes, may provide insight into disease pathogenesis. The aim of the study was to investigate the shared genetic factors that predispose individuals to being overweight and developing type 2 diabetes. We conducted genome-wide linkage analyses for type 2 diabetes in 386 affected individuals (269 sibpairs) from 171 Korean families and association analyses with single-nucleotide polymorphisms of candidate genes within linkage regions to identify genetic variants that predispose individuals to being overweight and developing type 2 diabetes. Through fine-mapping analysis of chromosome 4q34-35, we detected a locus potentially linked (nonparametric linkage 2.81, logarithm of odds 2.27,  $P = 6 \times 10^{-4}$ ) to type 2 diabetes in overweight or obese individuals (body mass index, BMI  $\geq 23 \text{ kg m}^{-2}$ ). Multiple regression analysis with type 2 diabetes-related phenotypes revealed a significant association (false discovery rate (FDR)  $P = 0.006$  for rs13144140; FDR  $P = 0.002$  for rs6830266) between *GPM6A* (rs13144140) and BMI and waist-hip ratio, and between *NEIL3* (rs6830266) and insulin level from 1314 normal individuals. Our systematic search of genome-wide linkage and association studies, demonstrate that a linkage peak for type 2 diabetes on chromosome 4q34-35 contains two type 2 diabetes-related genes, *GPM6A* and *NEIL3*.

*Experimental & Molecular Medicine* (2013) 45, e7; doi:10.1038/emm.2013.5; published online 8 February 2013

**Keywords:** affected sibpairs; association; linkage; obesity; overweight; type 2 diabetes

## INTRODUCTION

Type 2 diabetes mellitus is one of the most common metabolic disorders in the world. According to the International Diabetes Federation estimates, nearly 194 million people had type 2 diabetes in 2003, and this number is expected to increase to 333 million by 2025.<sup>1</sup> Although the pathogenesis of type 2 diabetes is not completely understood, it is well established that the disease is a consequence of complex interactions between both genetic and environmental factors. The quest to elucidate the genetic causes of type 2 diabetes was advanced with the recent advent of genome-wide association studies, from which nearly 25 new genetic loci robustly associated with

type 2 diabetes risk have been described.<sup>2–9</sup> However, it is likely that additional type 2 diabetes risk genes are yet to be discovered.

A complementary approach to genome-wide association studies for discovering susceptibility genes is genome-wide linkage analysis, which has relatively more power for identifying rare high-risk disease alleles.<sup>10</sup> This approach uses affected sibpairs (ASPs), nuclear families and multigenerational kindred to define chromosomal loci that contain candidate disease genes. More than 50 genome-wide linkage studies have demonstrated the presence of different type 2 diabetes susceptibility loci in different ethnic groups or

<sup>1</sup>Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Korea; <sup>2</sup>Division of Intractable Diseases, Center for Biomedical Sciences, National Institute of Health, Chungcheongbuk-do, Korea; <sup>3</sup>Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea and <sup>4</sup>WCU Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Medicine, Seoul National University, Seoul, Korea

Correspondence: Dr KS Park, WCU Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Medicine, Seoul National University, 28 Yongon-Dong Chongno-Gu, Seoul 110-744, Korea.

E-mail: kspark@snu.ac.kr

or Dr J-Y Lee, Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do 363-951, Korea.

E-mail: leeey63@nih.go.kr

Received 9 May 2012; accepted 10 September 2012

endophenotypes.<sup>11</sup> The usual paradigm for exploiting genome-wide linkage analysis is to focus on chromosomal regions having an established disease linkage and perform a fine-mapping and candidate gene study.

Among the various environmental risk factors for type 2 diabetes, obesity is considered to be the most important determinant.<sup>12</sup> Obesity is influenced by increased caloric intake and physical inactivity, but individual susceptibility also has a genetic basis. Recent studies have revealed genes, such as *FTO*, associated with both diabetes and obesity,<sup>13</sup> which may increase the risk of type 2 diabetes by modulating body mass index (BMI). Obesity and type 2 diabetes have complex interactions, and therefore, evaluation of the genetic risk factors according to BMI subgroups could be valuable. Furthermore, dividing the subjects into similar BMI subgroups would increase the chance of detecting a positive interaction.

In this study, we report the results of the first genome-wide linkage study for type 2 diabetes in the Korean population, followed by fine mapping using subgroup analysis according to BMI to increase the probability of identifying type 2 diabetes-linked genetic loci.

## METHODS

### Subjects

For the linkage study, nuclear Korean families with at least two siblings with type 2 diabetes were recruited from Seoul National University Hospital (SNUH). A total of 386 affected individuals (269 ASPs in 171 families) were considered for the study; parents and other normal siblings were not included. All subjects enrolled in this study were of Korean ethnicity. Diabetes was diagnosed based on the American Diabetes Association criteria.<sup>14</sup>

For the association study, 378 unrelated type 2 diabetes patients and 382 unrelated non-diabetic control subjects were recruited from SNUH. The diabetic subjects were randomly recruited from patients in the SNUH outpatient clinic. Non-diabetic control subjects were recruited from an unselected population undergoing a routine health checkup at SNUH. Subjects' height, weight, waist and hip circumferences were measured. Fasting plasma glucose and plasma insulin concentrations were measured. For the replication study, 949 unrelated type 2 diabetes subjects and 932 control subjects were selected from projects of the Korean Health and Genome Study (KHGS).<sup>15</sup> A total of 932 participants in the cohorts who had no history of diabetes, and no first-degree relatives with diabetes were recruited as normal control subjects. In addition, the normal control subjects were not taking medication for diabetes, hypertension or dyslipidemia.

The Institutional Review Boards of the Clinical Research Institute at the SNUH and the Korean National Institute of Health approved the study protocol, and written informed consent was obtained from each subject. Table 1 presents the clinical characteristics of the subjects in the linkage and association studies.

### Genotyping of microsatellite markers

Genotyping was performed using the fluorescently labeled human ABI Prism Linkage Mapping Set MD-10 (Applied Biosystems, Foster City, CA, USA), comprising 400 informative microsatellite markers with an average intermarker spacing of 9.7 cM. Each marker set included a fluorescently labeled forward primer and a tailing reverse

**Table 1 Subject characteristics**

Characteristic	Linkage study				Association study			
	SNUH		SNUH		KHGS		KHGS	
	T2DM (all)	T2DM (BMI $\geq 23$ )	T2DM (all)	T2DM (BMI $\geq 23$ )	T2DM (all)	T2DM (BMI $\geq 23$ )	T2DM (all)	T2DM (BMI $\geq 23$ )
Number of subjects	386 (269 Pairs, 171 families)	216 (132 Pairs, 101 families)	378	282	949	835	932	491
Sex (M/F)	170/216	93/123	181/197	136/146	479/470	414/421	447/485	194/297
Age (years)	56.51	56.77	61.44	60.82	57.17	56.04	64.13	63.91
BMI ( $\text{kg m}^{-2}$ )	24.60 $\pm$ 3.00	26.20 $\pm$ 2.34	24.87 $\pm$ 2.92	26.03 $\pm$ 2.30	26.03 $\pm$ 3.08	26.70 $\pm$ 2.57	23.34 $\pm$ 3.05	25.62 $\pm$ 2.08
WHR	0.90 $\pm$ 0.06	0.91 $\pm$ 0.05	0.91 $\pm$ 0.06	0.92 $\pm$ 0.05	0.92 $\pm$ 0.06	0.93 $\pm$ 0.06	0.91 $\pm$ 0.07	0.94 $\pm$ 0.06
Fasting insulin ( $\text{mg dl}^{-1}$ )	11.91 $\pm$ 15.32	11.66 $\pm$ 6.97	12.70 $\pm$ 10.57	13.80 $\pm$ 11.53	8.80 $\pm$ 5.89	9.11 $\pm$ 6.00	7.22 $\pm$ 6.91	8.05 $\pm$ 7.65

Abbreviations: BMI, body mass index; KHGS, Korean Health and Genome Study; NC, normal controls; SNUH, Seoul National University Hospital; T2DM, type 2 diabetes mellitus; WHR, waist-to-hip ratio.

primer. All markers were dinucleotide repeats of the type (CA)<sub>n</sub>, originally chosen from the Genethon/CEPH map.<sup>16</sup>

PCR (95 °C for 2 min, then 35 cycles of 95 °C for 20 s, 58 °C for 40 s and 72 °C for 30 s, followed by a final 72 °C for 45 min) was performed in a 384-well plate on a GeneAmp PCR system (9700 Block; Applied Biosystems) in 5 µl reactions containing: 10–20 ng genomic DNA, 2.5 mmol l<sup>-1</sup> MgCl<sub>2</sub>, 0.25 mmol l<sup>-1</sup> dNTPs, 1 pmol of primer and 0.2 U EF-Taq DNA polymerase in 1 × PCR buffer (SolGent, Daejeon, Korea). Automated 96-channel and 384-channel pipettes (Hydra I, II; Matrix Technologies, Hudson, NH, USA) were used for the pipetting steps. Electrophoresis and signal recording were performed on ABI 3730 automated sequencers (Applied Biosystems) using standard protocols. We used GeneScan 500 Liz (Applied Biosystems) as the internal size standard because it assists polymorphic fragment length calling and allows more accurate allele calling and unambiguous comparison of data across experimental conditions.

Genotypes were initially scored using GENEMAPPER 3.7 software, and were reviewed independently to confirm the accuracy of allele calling. All genotyped markers were checked for incompatibilities using the program MARKERINFO from the SAGE (statistical analysis for genetic epidemiology) package. Database-aided quality-control procedures included confirmation of standard individual genotypes (CEPH standard 1347-02; Coriell Institute for Medical Research, Camden, NJ, USA), plate identity, orientation and allele size.

### Linkage analyses and linkage programs

For ASP linkage analysis screening, nonparametric multipoint analyses of all autosomal chromosomes were performed with the programs GENEHUNTER v2.1 and Merlin v1.0.1 considering all the pairs as independent. For the X chromosome, the multipoint nonparametric linkage (NPL) score was calculated using GENEHUNTER Plus, which implemented both the NPL Z-score and the allele-sharing logarithm of odds (LOD).<sup>17,18</sup> Allele frequencies for the 400 markers were estimated based on the entire data set using Recode (Division of Statistical Genetics, Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA). Marker order and intermarker distances were taken from published Genethon/CEPH maps.<sup>16</sup> Parental genotypes were missing, estimated probabilities were calculated for all possible parental genotypes conditioned on the sibship genotype information.

We also used the SAGE package, which is based on methods proposed by Haseman and Elston.<sup>19</sup> Allele frequencies for the markers were estimated using the SAGE program FREQ. Pedigree relationships were tested using the RELTEST program. Alleles shared identical by descent (IBD) among independent sibpairs were calculated using the GENIBD program. For linkage testing, the mean proportion of alleles shared IBD among ASPs ( $\pi$ ) was estimated and tested against the null hypothesis of no linkage ( $\pi = 0.5$ ) (H<sub>0</sub>: mean IBD sharing (mean of the  $\pi_i$ ) =  $\frac{1}{2}(0 \times P_{(0)} + 0.5 \times P_{(1)} + 1 \times P_{(2)})$ ), H<sub>A</sub>: mean IBD sharing  $> \frac{1}{2}$ ), with significant excess of sharing taken as evidence for linkage by the SIBPAL program of the SAGE software package.

To explore possible heterogeneity in our data set and to overcome any associated reduction in power, subgroup analysis was performed according to BMI. We used a BMI of 23 kg m<sup>-2</sup> as a cutoff value for being overweight (including obesity), and both siblings who had a BMI  $\geq 23$  kg m<sup>-2</sup> were included for subgroup analysis.

### Association studies

Single-nucleotide polymorphisms (SNPs) were chosen from gene-centric SNPs in linkage region 4q34-35; 175 924 556 ~ 185 473 235,

and included SNPs located directly within coding, promoter, non-coding and intron regions.

Genotyping was performed using the Illumina's Golden Gate genotyping system<sup>20</sup> (Illumina, San Diego, CA, USA) and data quality was assessed using duplicate DNAs. All data considered for genotyping analysis had a genotype quality score  $\geq 0.25$ . SNPs that did not satisfy the following criteria were excluded: (i) a minimum call rate of 90%; (ii) no duplicate errors; (iii) Hardy-Weinberg equilibrium *P* value  $> 0.001$ . In total, 201 SNPs (SNUH: 166, KHGS: 66, overlap: 31) were used for association analysis. Associations between SNPs and type 2 diabetes-related phenotypes among normal subjects were determined by linear regression analysis while controlling for age, gender and BMI. Association was tested using co-dominant, dominant and recessive models. Statistical analyses were performed using SAS software (SAS institute, Cary, NC, USA).

## RESULTS

### Genome-wide linkage analysis in the whole study group

Whole-genome linkage analysis using ASPs with 400 micro-satellite markers was performed on 269 Korean ASPs (171 families) with type 2 diabetes. The observed z-scores of multipoint NPL scores from the genome-wide scan are shown in Supplementary Figure S1. The maximum score was on chromosome 4, with a peak multipoint NPL score of 1.52 (*P* = 0.06) at marker D4S415 in a region encompassing 185 cM.

### Subgroup analysis of individuals with BMI $\geq 23$ kg m<sup>-2</sup>

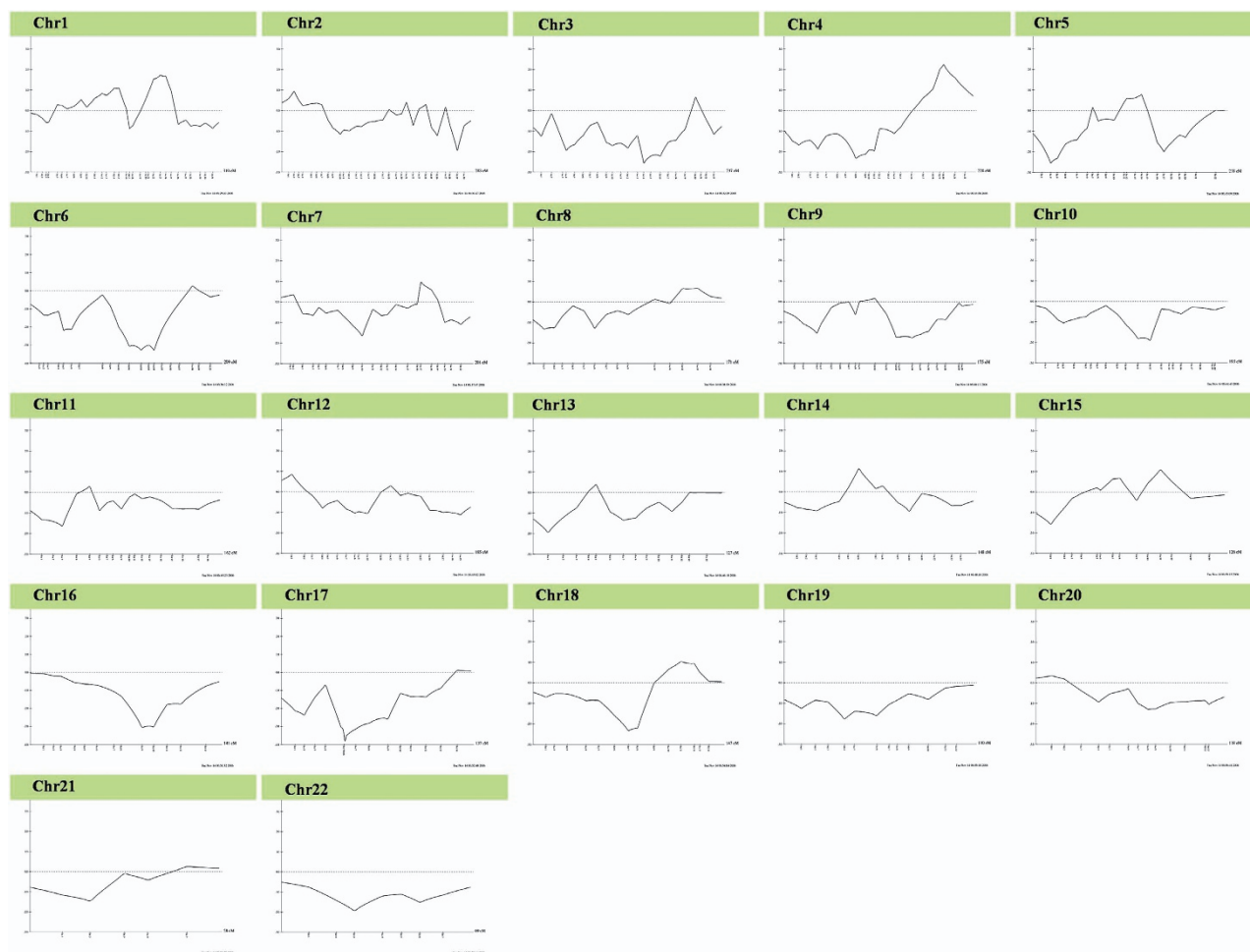
To assess a possible interaction between BMI and diabetes, we carried out subgroup analysis with a BMI cutoff of 23 kg m<sup>-2</sup>. The BMI  $< 23$  kg m<sup>-2</sup> group had a multipoint NPL score of 2.13 for 18q22 (Supplementary Figure S2). The BMI  $\geq 23$  kg m<sup>-2</sup> group had a multipoint NPL score of 1.73 for 1q31 and 2.24 for 4q34 (Figure 1).

The 4q34 region in the BMI  $\geq 23$  kg m<sup>-2</sup> group had the highest NPL score of all the regions examined. To confirm this linkage that had been analyzed with GENEHUNTER v2.1, we again analyzed chromosome 4 (Subset: BMI  $\geq 23$  kg m<sup>-2</sup>) with 132 ASPs (101 families) using the SAGE software package. Among the 22 markers on chromosome 4, D4S1539, D4S415 and D4S1535 passed the statistical threshold for linkage. Marker D4S415 showed an especially significant result with the mean proportion of alleles sharing an IBD score of 0.56 with *P* = 0.007 (Supplementary Table S1).

We also carried out a linkage analysis for BMI on chromosome 4 using the traditional Haseman-Elston procedure included in the SAGE package. Again, we found evidence for linkage at marker D4S415 on 4q34 with *P* = 0.009 (data not shown).

### Fine mapping of the chromosome 4q linkage region

For high-resolution mapping, 132 ASPs in the BMI  $\geq 23$  kg m<sup>-2</sup> group were genotyped by adding eight microsatellite markers (D4S3033, D4S2952, D4S2979, D4S1595, D4S2991, D4S1607, D4S3015 and D4S2920) that covered the candidate regions on chromosome 4q. Multipoint NPL analysis showed a significantly increased *Z* value, with an NPL score corresponding to marker D4S3015 of 2.81 (LOD



**Figure 1** Multipoint NPL score maps of all autosomal chromosomes in the BMI  $\geq 23 \text{ kg m}^{-2}$  group with 132 ASPs (101 families). Multipoint NPL scores were calculated using GENEHUNTER v2.1. In each graph, the left vertical axis indicates the NPL score, the horizontal axis indicates the length of each chromosome, and the locus numbers on the horizontal axis indicate the positions of the microsatellite markers.

2.27,  $P = 6 \times 10^{-4}$ ) at 188.8 cM (Figure 2). We also examined the  $P$ -values of the mean proportion test calculated by the SIBPAL program. Among the 30 markers of chromosome 4 including the additional eight markers listed above, D4S415, D4S1607, D4S3015, D4S2920 and D4S1535 passed the statistical threshold for linkage. The D4S3015 marker showed an especially significant result with the mean proportion of alleles sharing an IBD score of 0.58 with  $P = 0.001$  (Supplementary Table S2).

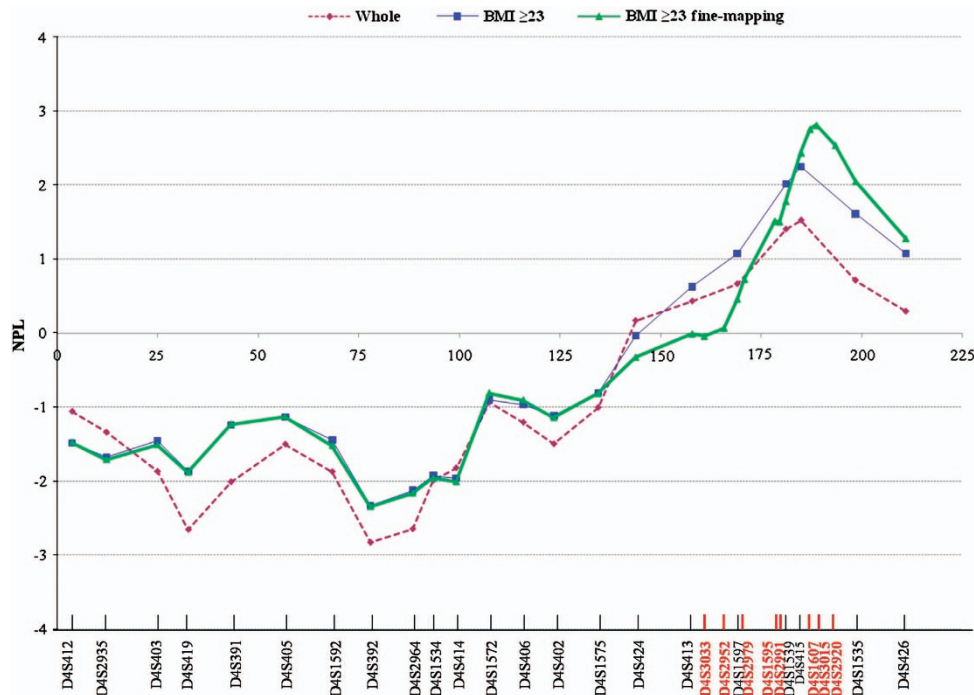
The 'one LOD drop' support interval was identified as a 9.5-Mb region between markers D4S1539 and D4S1535, with an NPL score  $> 1.81$ .

#### Association studies in the chromosome 4q34-35 linkage region

To identify candidate genes predisposing individuals to type 2 diabetes, we examined possible associations between type 2 diabetes and SNPs from 23 genes in the 9.5-Mb 4q34-35 linkage region. In the first association study, we analyzed the association of 166 SNPs with type 2 diabetes in 411 SNUH

subjects (282 cases, 129 controls) with BMI  $\geq 23 \text{ kg m}^{-2}$ . In the logistic analysis, several SNPs in the gene encoding glycoprotein M6a (*GPM6A*) showed marginal association with type 2 diabetes (Supplementary Table S3). We also analyzed the association of 66 SNPs with type 2 diabetes in 1326 KHGS subjects (835 cases, 491 controls) with BMI  $\geq 23 \text{ kg m}^{-2}$ . In the logistic analysis, several SNPs in the gene encoding Nei endonuclease VIII-like 3 (*NEIL3*) showed marginal association with type 2 diabetes (Supplementary Table S3). We next conducted a combined analysis with 1737 SNUH and KHGS individuals with BMI  $\geq 23 \text{ kg m}^{-2}$ . Among the 31 SNPs, two SNPs in *NEIL3*, rs11940019 and rs17676249, showed an odds ratio of 4.80 ( $P = 0.02$ ) using the recessive model. In the case of *GPM6A*, haplotype analysis (rs23332251(T)/rs7675676(G)) showed an association ( $P = 0.008$ ) with type 2 diabetes (Supplementary Table S3).

We next performed quantitative trait analysis by assessing the association of *GPM6A* and *NEIL3* with type 2 diabetes-related phenotypes in normal control subjects. Results for the multiple regression analysis of association between *GPM6A*



**Figure 2** Multipoint NPL score map of chromosome 4. The whole and BMI  $\geq 23 \text{ kg m}^{-2}$  groups were analyzed using 22 markers at 9.5 cM resolution, and BMI  $\geq 23 \text{ kg m}^{-2}$  fine mapping was analyzed with 30 markers and an additional eight markers in the 40.6-cM interval at a resolution of 157.9 to 198.5 cM. Multipoint NPL scores were calculated using GENEHUNTER v2.1. In the graph, the left vertical axis indicates the NPL score, and the horizontal axis indicates the markers of chromosome 4. The additional eight markers are shown in red.

and *NEIL3*, and BMI, waist-hip ratio (WHR), and fasting insulin level are shown in Table 2. The *GPM6A* SNPs were associated with BMI and WHR, and the *NEIL3* SNPs were associated with fasting insulin level in 382 SNUH and 932 KHGS subjects. In the combined 1314 normal subjects (SNUH + KHGS), the rs13152426 and rs13144140 SNPs in *GPM6A* were significantly associated with BMI ( $P = 0.0004$ , false discovery rate (FDR) 0.0062) and WHR ( $P = 0.0007$ , FDR 0.0109). The rs6850861, rs6823018, rs6830266 and rs2048077 SNPs in *NEIL3* were significantly associated ( $P = 0.0003$ , FDR 0.0023) with fasting insulin level. However, the fasting glucose level was not significantly associated with any *GPM6A* or *NEIL3* SNPs (data not shown).

## DISCUSSION

We report here the first genome-wide search for chromosome loci associated with type 2 diabetes susceptibility in Korean subjects. Our results reveal evidence of linkage at the 4q34-35 locus in subjects with BMI  $\geq 23 \text{ kg m}^{-2}$ . The study is comprehensive because we performed genome-wide linkage analysis, fine mapping and association analysis with quantitative traits. Another strength of our study is that the ethnicity of the Korean population is relatively homogeneous, resulting in a higher probability of identifying diabetes-linked loci.

There have been more than 50 type 2 diabetes linkage analysis studies conducted in various populations, but few loci

with strong evidence for linkage have been replicated.<sup>11</sup> The 4q34-35 region showed a replicated linkage signal that was reported in several populations. Significant evidence for linkage has been obtained for marker D4S1501 on 4q34 in Ashkenazi Jewish individuals,<sup>21</sup> and modest evidence for linkage between type 2 diabetes and chromosome 4q34-q35 was detected in Finnish families.<sup>22</sup> In addition, the 4q34 region contains susceptibility loci in French whites.<sup>23</sup> In the French study, the loci were detected when subjects were subdivided according to a BMI of  $27 \text{ kg m}^{-2}$ , which is similar to our approach.

We used a BMI of  $23 \text{ kg m}^{-2}$  as a cutoff value for being overweight. A WHO expert consultant considered whether a population-specific cutoff point for BMI was necessary, and concluded that a substantial proportion of the Asian population is at high risk for type 2 diabetes and that many Asians have BMIs lower than the existing WHO cutoff point for being overweight ( $\geq 25 \text{ kg m}^{-2}$ ) compared with Caucasians (in general) or European populations.<sup>24</sup> Another WHO report indicated that Asian adults with a BMI  $> 23.0$  should be considered overweight.<sup>25</sup>

We found evidence of linkage for type 2 diabetes from subgroups with BMI  $\geq 23 \text{ kg m}^{-2}$ . These results indicate an interaction between susceptibility loci and obesity. Moreover, subgrouping by BMI may have increased our chance of discovering risk loci. Conversely, subgrouping could lead to false-positive results. Because we confirmed our results in two

Table 2 Association results for candidate genes and type 2 diabetes-related phenotypes in normal subjects

	rs#	Gene	Phenotype	C/C <sup>a</sup>	C/R <sup>a</sup>	R/R <sup>a</sup>	P-value			N
							Co-dominant	Dominant	Recessive	
SNUH	rs13152426	GPM6A	BMI	507 (23.56 ± 3.06)	376 (23.05 ± 3)	49 (23.16 ± 3.2)	<b>0.008</b>	<b>0.007</b>	<b>0.24</b>	382
	rs13144140	GPM6A	BMI	507 (23.56 ± 3.06)	376 (23.05 ± 3)	49 (23.16 ± 3.2)	<b>0.008</b>	<b>0.007</b>	<b>0.24</b>	382
	rs6553892	GPM6A	WHR	737 (0.91 ± 0.06)	185 (0.91 ± 0.07)	9 (0.85 ± 0.07)	<b>0.005</b>	<b>0.006</b>	<b>0.37</b>	382
	rs13152426	GPM6A	WHR	506 (0.91 ± 0.07)	376 (0.91 ± 0.06)	49 (0.89 ± 0.06)	<b>0.03</b>	0.05	0.14	382
	rs13144140	GPM6A	WHR	506 (0.91 ± 0.07)	376 (0.91 ± 0.06)	49 (0.89 ± 0.06)	<b>0.03</b>	0.05	0.14	382
	rs1471797	GPM6A	WHR	379 (0.9 ± 0.07)	445 (0.91 ± 0.06)	106 (0.92 ± 0.07)	0.27	0.85	<b>0.03</b>	382
	rs6850861	NEIL3	log <sub>2</sub> Ins	507 (0.73 ± 0.35)	355 (0.78 ± 0.29)	70 (0.8 ± 0.32)	<b>0.005</b>	<b>0.008</b>	<b>0.08</b>	382
	rs6823018	NEIL3	log <sub>2</sub> Ins	548 (0.74 ± 0.34)	335 (0.78 ± 0.29)	49 (0.79 ± 0.36)	<b>0.0004</b>	<b>0.0007</b>	<b>0.04</b>	382
	rs6830266	NEIL3	log <sub>2</sub> Ins	523 (0.73 ± 0.34)	354 (0.79 ± 0.28)	55 (0.78 ± 0.36)	<b>0.001</b>	<b>0.002</b>	<b>0.04</b>	382
	rs2048077	NEIL3	log <sub>2</sub> Ins	506 (0.73 ± 0.35)	356 (0.78 ± 0.29)	70 (0.8 ± 0.32)	<b>0.005</b>	<b>0.008</b>	<b>0.08</b>	382
	rs11940019	NEIL3	log <sub>2</sub> Ins	729 (0.76 ± 0.33)	194 (0.75 ± 0.31)	9 (0.51 ± 0.36)	<b>0.04</b>	0.23	<b>0.03</b>	382
	rs17676249	NEIL3	log <sub>2</sub> Ins	731 (0.76 ± 0.33)	192 (0.75 ± 0.31)	9 (0.51 ± 0.36)	<b>0.005</b>	<b>0.01</b>	0.10	382
	rs13152426	GPM6A	BMI	196 (22.78 ± 3.20)	157 (21.93 ± 2.73)	27 (21.77 ± 2.68)	<b>0.03</b>	<b>0.02</b>	0.73	932
	rs13144140	GPM6A	BMI	196 (22.78 ± 3.20)	157 (21.93 ± 2.73)	27 (21.77 ± 2.68)	<b>0.03</b>	<b>0.02</b>	0.73	932
	rs6553892	GPM6A	WHR	260 (0.88 ± 0.06)	63 (0.91 ± 0.07)	2 (0.93 ± 0.01)	0.40	0.80	<b>0.006</b>	932
	rs13152426	GPM6A	WHR	168 (0.90 ± 0.06)	134 (0.88 ± 0.06)	23 (0.87 ± 0.06)	<b>0.01</b>	<b>0.04</b>	<b>0.03</b>	932
KHGS	rs13144140	GPM6A	WHR	168 (0.90 ± 0.06)	134 (0.88 ± 0.06)	23 (0.87 ± 0.06)	<b>0.01</b>	<b>0.04</b>	<b>0.03</b>	932
	rs1471797	GPM6A	WHR	145 (0.89 ± 0.06)	147 (0.89 ± 0.06)	33 (0.87 ± 0.05)	<b>0.004</b>	<b>0.01</b>	<b>0.03</b>	932
	rs6850861	NEIL3	log <sub>2</sub> Ins	190 (0.79 ± 0.22)	139 (0.83 ± 0.21)	29 (0.88 ± 0.18)	<b>0.006</b>	<b>0.01</b>	<b>0.05</b>	932
	rs6823018	NEIL3	log <sub>2</sub> Ins	213 (0.78 ± 0.22)	119 (0.84 ± 0.22)	26 (0.89 ± 0.15)	<b>0.01</b>	<b>0.02</b>	0.17	932
	rs6830266	NEIL3	log <sub>2</sub> Ins	205 (0.79 ± 0.22)	126 (0.83 ± 0.22)	27 (0.89 ± 0.15)	<b>0.008</b>	<b>0.008</b>	0.22	932
	rs2048077	NEIL3	log <sub>2</sub> Ins	190 (0.79 ± 0.22)	139 (0.83 ± 0.21)	29 (0.88 ± 0.18)	<b>0.006</b>	<b>0.01</b>	<b>0.05</b>	932
	rs11940019	NEIL3	log <sub>2</sub> Ins	124 (0.80 ± 0.22)	170 (0.80 ± 0.22)	64 (0.85 ± 0.21)	0.28	0.54	<b>0.02</b>	932
	rs17676249	NEIL3	log <sub>2</sub> Ins	284 (0.83 ± 0.21)	71 (0.75 ± 0.24)	3 (0.50 ± 0.34)	0.30	0.57	<b>0.02</b>	932
	rs13152426	GPM6A	BMI	703 (23.35 ± 3.12)	533 (22.72 ± 2.96)	76 (22.67 ± 3.08)	<b>0.0008 (0.0124)<sup>b</sup></b>	<b>0.0004 (0.0062)</b>	0.26	1314
	rs13144140	GPM6A	BMI	703 (23.35 ± 3.12)	533 (22.72 ± 2.96)	76 (22.67 ± 3.08)	<b>0.0008 (0.0124)</b>	<b>0.0004 (0.0062)</b>	0.26	1314
	rs6553892	GPM6A	WHR	997 (0.90 ± 0.06)	248 (0.91 ± 0.07)	11 (0.86 ± 0.07)	0.49	0.22	<b>0.03</b>	1314
	rs13152426	GPM6A	WHR	674 (0.91 ± 0.07)	510 (0.90 ± 0.06)	72 (0.88 ± 0.06)	<b>0.0007 (0.0109)</b>	<b>0.005</b>	<b>0.006</b>	1314
	rs13144140	GPM6A	WHR	674 (0.91 ± 0.07)	510 (0.90 ± 0.06)	72 (0.88 ± 0.06)	<b>0.0007 (0.0109)</b>	<b>0.005</b>	<b>0.006</b>	1314
	rs1471797	GPM6A	WHR	524 (0.90 ± 0.07)	592 (0.91 ± 0.06)	139 (0.91 ± 0.07)	<b>0.03</b>	<b>0.02</b>	0.32	1314
	rs6850861	NEIL3	log <sub>2</sub> Ins	697 (0.75 ± 0.32)	494 (0.79 ± 0.27)	99 (0.82 ± 0.29)	<b>0.0003 (0.0023)</b>	<b>0.001 (0.0077)</b>	<b>0.012 (0.0372)</b>	1314
	rs6823018	NEIL3	log <sub>2</sub> Ins	761 (0.75 ± 0.31)	454 (0.80 ± 0.27)	75 (0.83 ± 0.31)	<b>0.0003 (0.0023)</b>	<b>0.0007 (0.0077)</b>	<b>0.03</b>	1314
Total	rs6830266	NEIL3	log <sub>2</sub> Ins	728 (0.75 ± 0.32)	480 (0.80 ± 0.27)	82 (0.82 ± 0.31)	<b>0.0003 (0.0023)</b>	<b>0.0004 (0.0077)</b>	<b>0.04</b>	1314
	rs2048077	NEIL3	log <sub>2</sub> Ins	696 (0.75 ± 0.32)	495 (0.79 ± 0.27)	99 (0.82 ± 0.29)	<b>0.0003 (0.0023)</b>	<b>0.001 (0.0077)</b>	<b>0.012 (0.0372)</b>	1314
	rs11940019	NEIL3	log <sub>2</sub> Ins	1013 (0.78 ± 0.30)	265 (0.75 ± 0.29)	12 (0.51 ± 0.34)	<b>0.04</b>	0.11	<b>0.0056 (0.0248)</b>	1314
	rs17676249	NEIL3	log <sub>2</sub> Ins	1015 (0.78 ± 0.30)	263 (0.76 ± 0.29)	12 (0.51 ± 0.34)	0.05	0.16	<b>0.0056 (0.0248)</b>	1314

Abbreviations: CI, confidence interval; FDR, false discovery rate; GPM6A, glycoprotein M6a; KHGS, Korean Health and Genome Study; Ins, insulin; NEIL3, Nei endonuclease VIII-like 3; OR, odds ratio; SNUH, Seoul National University Hospital; WHR, waist-hip ratio.

<sup>a</sup>C/C, C/R and R/R represent homozygotes for the common allele, and heterozygotes and homozygotes for the rare allele, respectively. Genotype distributions, means, and s.d. of each value and P-values are shown.

Numbers in bold are statistically significant.

<sup>b</sup>0 represents the P-value analyzed in FDR.

replication sets and performed a quantitative trait analysis, however, the possibility of false positives seems low.

*GPM6A* (glycoprotein M6A) acts as a stress-response gene during hippocampal formation. The relationship between *GPM6A*, type 2 diabetes and being overweight is unknown. However, a small leucine-rich glycoprotein called decorin was reported to be associated with type 2 diabetes and obesity, possibly by upregulating the expression of decorin in adipose tissue in type 2 diabetes subjects.<sup>26</sup> To assess the putative function of the *GPM6A* variant, we analyzed the effect of the rs13144140 SNP (intron1, NM\_201591) using an *in silico* approach. The FASTSNP program allows users to efficiently identify and prioritize high-risk SNPs according to their phenotypic risks and putative functional effects.<sup>27</sup> The analysis of rs13144140 with FASTSNP revealed that it is predicted to be a functional change in the protein by causing a change in a transcription factor binding sites. Using TRANSFAC,<sup>28</sup> we found that rs13144140 correlated with binding of HNF-1, a transcription factor that controls multiple genes implicated in pancreatic  $\beta$ -cell function.<sup>29</sup> These observations were further supported by matching the HNF-1 DNA binding domain sequence (TGCAAATCAT TTTC) using the Transcriptional Regulatory Element Database.<sup>30</sup> This intronic variant (rs13144140) could be an internal enhancer element that regulates *GPM6A* expression in an adipose tissue-specific manner.

*NEIL3* belongs to a class of DNA glycosylases homologous to the bacterial Fpg/Nei family. These glycosylases initiate the first step in base excision repair by cleaving bases damaged by reactive oxygen species and introducing a DNA strand break via the associated lyase reaction.<sup>31</sup> Three human genes, designated *NEIL1*, *NEIL2* and *NEIL3*, encode proteins that contain sequence homologies to Fpg and Nei.<sup>32</sup> Deletion of *NEIL1* results in a metabolic syndrome. In the absence of exogenous oxidative stress, *neil1* knockout (*neil1*<sup>-/-</sup>) and heterozygous (*neil1*<sup>+/-</sup>) mice develop severe obesity, dyslipidemia and fatty liver disease, and also have a tendency to develop hyperinsulinemia.<sup>33</sup> Although the role of *NEIL3* in type 2 diabetes or obesity is not yet known, it may have a similar function as *NEIL1*.

In summary, we have identified *GPM6A* and *NEIL3* as being associated with overweight and type 2 diabetes using a systematic search through genome-wide linkage and linkage region-based association analyses in Koreans. The chromosomal locus 4q34-35 was linked to type 2 diabetes in subjects with BMI  $\geq 23$  kg m<sup>-2</sup>. The genes located in this region were associated with metabolic traits, such as insulin level, BMI and WHR. Further studies of these two genes in different populations are required.

## ACKNOWLEDGEMENTS

This work was supported by an intramural grant from the Korea National Institute of Health, Centers for Disease Control and Prevention, Republic of Korea (4800-4845-301-210) and a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (00-PJ3-PG6-GN07-0001).

- 1 Sicree R, Shaw JE, Zimmet PZ. The global burden of diabetes. In: Gan D (ed.). *Diabetes Atlas*. International Diabetes Federation: Brussels, 2003, pp 15–71.
- 2 Garup N, Andersen G, Krarup NT, Albrechtsen A, Schmitz O, Jorgensen T *et al*. Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* 2008; **57**: 2534–2540.
- 3 Lyssenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spegel P *et al*. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 2009; **41**: 82–88.
- 4 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL *et al*. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
- 5 Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D *et al*. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
- 6 Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB *et al*. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007; **39**: 770–775.
- 7 van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA *et al*. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes* 2008; **57**: 3122–3128.
- 8 Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T *et al*. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008; **40**: 638–645.
- 9 Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H *et al*. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
- 10 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
- 11 Barroso I. Genetics of Type 2 diabetes. *Diabet Med* 2005; **22**: 517–535.
- 12 Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. *JAMA* 1999; **282**: 1523–1529.
- 13 Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A *et al*. Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* 2008; **57**: 1419–1426.
- 14 Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 1997; **20**: 1183–1197.
- 15 Lim S, Jang HC, Lee HK, Kimm KC, Park C, Cho NH. A rural-urban comparison of the characteristics of the metabolic syndrome by gender in Korea: the Korean Health and Genome Study (KHGS). *J Endocrinol Invest* 2006; **29**: 313–319.
- 16 Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A *et al*. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; **380**: 152–154.
- 17 Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
- 18 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 19 Rao S, Wang QK. SAGE programs: model-free linkage analysis for complex cardiovascular phenotypes. *Methods Mol Med* 2006; **128**: 61–89.
- 20 Oliphant A, Barker DL, Stuelcpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002; **56**: 60–61.
- 21 Permutt MA, Wasson JC, Suarez BK, Lin J, Thomas J, Meyer J *et al*. A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 2001; **50**: 681–685.
- 22 Lindgren CM, Mahtani MM, Widen E, McCarthy MI, Daly MJ, Kirby A *et al*. Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am J Hum Genet* 2002; **70**: 509–516.
- 23 Vionnet N, Hani EH, Dupont S, Gallina S, Francke S, Dotte S *et al*. Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type

- 2-diabetes locus on chromosome 1q21-q24. *Am J Hum Genet* 2000; **67**: 1470–1480.
- 24 WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004; **363**: 157–163.
- 25 WHO/IASO/IOTF. *The Asia-Pacific Perspective: Redefining Obesity and its Treatment*. Health Communications Australia Pty Ltd., 2000.
- 26 Bolton K, Segal D, McMillan J, Jowett J, Heilbronn L, Abberton K *et al*. Decorin is a secreted protein associated with obesity and type 2 diabetes. *Int J Obes* 2008; **32**: 1113–1121.
- 27 Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ *et al*. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 2006; **34** (Web Server issue): W635–W641.
- 28 Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R *et al*. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003; **31**: 374–378.
- 29 Wang H, Antinozzi PA, Hagenfeldt KA, Maechler P, Wollheim CB. Molecular targets of a human HNF1 alpha mutation responsible for pancreatic beta-cell dysfunction. *EMBO J* 2000; **19**: 4257–4264.
- 30 Zhao F, Xuan Z, Liu L, Zhang MQ. TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res* 2005; **33** (Database issue): D103–D107.
- 31 Bandaru V, Sunkara S, Wallace SS, Bond JP. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair* 2002; **1**: 517–529.
- 32 Rosenquist TA, Zaika E, Fernandes AS, Zharkov DO, Miller H, Grollman AP. The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA Repair* 2003; **2**: 581–591.
- 33 Vartanian V, Lowell B, Minko IG, Wood TG, Ceci JD, George S *et al*. The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proc Natl Acad Sci USA* 2006; **103**: 1864–1869.



**This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>**

Supplementary Information accompanies the paper on the Experimental & Molecular Medicine website (<http://www.nature.com/emm>)