Observational Study

Medicine®

OPEN

# An investigation of obesity susceptibility genes in Northern Han Chinese by targeted resequencing

Yili Wu, PhD*, Weijing Wang, PhD, Wenjie Jiang, MD, Jie Yao, PhD, Dongfeng Zhang, MD

**Abstract**

Our earlier genome-wide linkage study of body mass index (BMI) showed strong signals from 7q36.3 and 8q21.13. This case–control study set to investigate 2 genomic regions which may harbor variants contributed to development of obesity.

We employed targeted resequencing technology to detect single nucleotide polymorphisms (SNPs) in 7q36.3 and 8q21.13 from 16 individuals with obesity. These were compared with 504 East Asians in the 1000 Genomes Project as a reference panel. Linkage disequilibrium (LD) block analysis was performed for the significant SNPs located near the same gene. Genes involved in statistically significant loci were then subject to gene set enrichment analysis (GSEA).

The 16 individuals aged between 30 and 60 years with BMI = $33.25 \pm 2.22\,kg/m^2$. A total of 12,131 genetic variants across all of samples were found. After correcting for multiple testing, 65 SNPs from 25 nearest genes (*INSIG1*, *FABP5*, *PTPRN2*, *VIPR2*, *WDR60*, *SHH*, *UBE3C*, *LMBR1*, *PAG1*, *IMPA1*, *CHMP4*, *SNX16*, *BLACE*, *EN2*, *CNPY1*, *LOC100506302*, *RBM33*, *LOC389602*, *LOC285889*, *LINC01006*, *NOM1*, *DNAJB6*, *LOC101927914*, *ESYT2*, *LINC00689*) were associated with obesity at significant level $q$-value $\leq 0.05$. LD block analysis showed there were 10 pairs of loci with $D' \geq 0.8$ and $r^2 \geq 0.8$. GSEA further identified 2 major related gene sets, involving lipid raft and lipid metabolic process, with FDR values $<0.12$ and $<0.4$, respectively.

Our data are the first documentation of genetic variants in 7q36.3 and 8q21.13 associated with obesity using target capture sequencing and Northern Han Chinese samples. Additional replication and functional studies are merited to validate our findings.

**Abbreviations:** AF = allele frequency, Alt. = alteration, *BLACE* = B-cell acute lymphoblastic leukemia expressed, BMI = body mass index, BWA = Burrows–Wheeler aligner, *CHMP4* = charged multivesicular body protein 4B, Chr. = chromosome, *CNPY1* = canopy FGF signaling regulator 1, DNA = deoxyribonucleic acid, *DNAJB6* = DnaJ heat shock protein family (Hsp40) Member B6, *EN2* = engrailed homeobox 2, *ESYT2* = extended synaptotagmin protein 2, *FABP5* = fatty acid binding protein 5, FDR = false discovery rate, GATK = Genome Analysis Toolkit, GO = genetic ontology, GSEA = gene set enrichment analysis, GWAS = genome-wide association studies, *IMPA1* = inositol monophosphatase 1, *INSIG1* = insulin induced gene 1, LD = linkage disequilibrium, *LINC00689* = long intergenic nonprotein coding RNA 689, *LINC01006* = long intergenic nonprotein coding RNA 1006, *LMBR1* = limb development membrane protein 1, *LOC100506302* = uncharacterized LOC100506302, *LOC101927914* = uncharacterized LOC101927914, *LOC285889* = uncharacterized LOC285889, *LOC389602* = uncharacterized LOC389602, NGSQC = Next Generation Sequencing Quality Control, *NOM1* = nucleolar protein with MIF4G domain 1, *PAG1* = phosphoprotein membrane anchor with glycosphingolipid microdomains 1, PCR = polymerase chain reaction, *PTPRN2* = protein tyrosine phosphatase, receptor type N2, *RBM33* = RNA binding motif protein 33, Ref. = reference, *SHH* = Sonic Hedgehog, SNP = single nucleotide polymorphisms, *SNX16* = sorting Nexin 16, *UBE3C* = ubiquitin protein ligase E3C, VIP = vasoactive intestinal polypeptide, *VIPR2* = vasoactive intestinal peptide receptor 2, *WDR60* = WD repeat domain 60.

**Keywords:** 7q36.3, 8q21.13, obesity, susceptibility genes, targeted resequencing

## 1. Introduction

There has been a worldwide epidemic in obesity linking increased morbidity and mortality as one of the major public health problems across countries. As a complex disorder, obesity is determined by both genetic and environmental factors, and the genetic influence accounts for 40% to 70% of the individual differences.[1] Recent genome-wide association studies (GWAS) of BMI have identified 97 genetic variants.[2] Nonetheless, these loci only explained 2.7% of the variance in BMI.[2] While epistatic and gene–environment interactions may contribute to the unexplained heritability of obesity, there is possibility that a significant fraction of the missing heritability is due to loci not yet identified or fully characterized.[3]

The advent of next-generation sequencing (NGS) with high-throughput screening provided both a broad spectrum and a precise vision for the genetic architecture of many diseases.[4] Although resequencing projects of whole human genome is still hampered by their high cost, the combination of target genomic region capture with NGS, as a low-cost technology with high efficiency and fidelity, has been used to investigate on several complex disorders and diseases.[4–7]

Editor: Marcello Iriti.

Funding: This study was supported by National Natural Science Foundation of China (grant number: 81302485), Natural Science Foundation of Shandong Province (grant number: BS2013SF002), China Postdoctoral Science Foundation (grant number: 2016M602102), and Qingdao Postdoctoral Science Foundation (grant number: 2016064).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Department of Epidemiology and Health Statistics, Public Health College, Qingdao University, Qingdao, Shandong Province, China.

* Correspondence: Yili Wu, Public Health College, Qingdao University, Qingdao, Shandong Province, China (e-mail: yiliwu79@163.com).



Medicine (2017) 96:7(e6117)

Received: 2 October 2016 / Received in final form: 18 January 2017 / Accepted: 20 January 2017

http://dx.doi.org/10.1097/MD.0000000000006117

Our previous genome-wide linkage analysis on BMI in 126 dizygotic twins identified a genome-wide significant linkage peak on chromosome 7 with a $\log_{10}$ the odds ratio (LOD) score of 4.06 and 3 suggestive linkage regions on additional regions with LOD score ≥2.2.[8] It is notable that our highest linkage region for BMI at 7q36 concurred with the result of a large multicenter linkage study of 4401 twin families from western countries.[9]

Greatly encouraged by these important findings, this study aims to determine the genetic variants on 7q36.3 and 8q21.13 (the top 2 ranked linkage regions in our previous study) involved in obesity through targeted resequencing technology.

## 2. Methods

### 2.1. Study samples

A total of 16 unrelated individuals with obesity were recruited from the Physical Examination Department of Qingdao Diabetes Hospital in October 2015. Information was collected through questionnaire, extraction of blood, together with anthropometric and laboratory measurements by well-trained clinicians face-to-face. BMI was derived by taking body weight (in kilogram) divided by height (in meter) squared. Subjects were included if the following criteria were met: aged 18 to 60 years; Han Chinese; ancestral home is in Shandong Province; BMI ≥ 30 kg/m$^2$; free of hypertension, diabetes, or cardiovascular disease. Those who were pregnant, breastfeeding or taking weight-reducing medication within 1 month were excluded. Written informed consent form was obtained from all participants and the study was approved by the Qingdao University Ethics Committee.

### 2.2. DNA extraction, target genomic region capture and sequencing

Genomic DNA was extracted from whole peripheral blood of the subjects using QIAamp DNA Blood Mini Kit (Qiagen, GmbH, Hilden, Germany). DNA quantification and integrity were determined by the Nanodrop spectrophotometer (Thermo Fisher Scientific, Inc., Wilmington, DE) and the 1% agarose electrophoresis, respectively. A custom capture library (Agilent Technologies, Inc., Santa Clara, CA) on genomic locations of interest: 155100001–159138663 on chromosome 7 and 81874071–82674071 on chromosome 8 (Table 1) was designed. The information about the 2 target capture genomic regions from "UCSC genome browser" was shown in Supplementary Files 1 and 2, http://links.lww.com/MD/B572.

Sixteen genomic DNA samples were captured on Agilent SureSelect custom library following the manufacturer's protocol (http://www.chem.agilent.com/library/usermanuals/Public/G7530–90000.pdf). Briefly, approximately 800 ng genomic DNA was sheared to 150 to 200 bp small fragments using sonicator (Covaris, Inc., Woburn, MA). The sheared deoxyribonucleic acid (DNA) was purified and treated with reagents supplied with the kit according to the protocol. Adapters from Agilent were ligated onto the polished ends and the libraries were amplified by polymerase chain reaction (PCR). The amplified libraries were hybridized with the custom probes. The DNA fragments bound with the probes were washed and eluted with the buffer provided in the kit. Then these libraries were sequenced on the Illumina sequencing platform (HiSeq X-10, Illumina, Inc., San Diego, CA) and 150 bp paired-end reads were generated.

**Table 1**

**Target genomic regions.**

| Chromosome | Reference sequence number | Start | End | Total length |
|---|---|---|---|---|
| 7 | rs712199 | 155,100,001 | 159,138,663 | 4,038,662 |
| 8 | rs4571694 | 81,874,071 | 82,674,071 | 800,000 |

### 2.3. Data analysis

We applied raw data filtering using Next Generation Sequencing Quality Control (NGSQC)—Toolkit v2.3.3 software. Raw reads which contained less than 70% high quality bases (Q20) or any N-base were removed. We then removed reads shorter than 70 bp afterwards to obtain clean reads. After quality control, raw data with 11.18 G bases were decreased to 10.41 G.

We used the Churchill software that integrated the following processes to call variants: Burrows–Wheeler aligner (BWA)—0.7.5a for mapping of paired end reads; Picard-tools—1.104 for marking duplicates that originate from PCR amplification (and that map at the same location); Genome Analysis Toolkit (GATK)—3.2 for realigning reads around indels, base recalibrate, and call variants with HaplotypeCaller method. Based on the recommended parameters on official website of GATK (https://software.broadinstitute.org/gatk/), the parameters stand_emit_conf was set to 30 to filter low quality variants.

We used ANNOVAR software to utilize up date-to-date information to functionally annotate genetic variants detected from Genome Reference Consortium GRCh37. And we identified variants documented in specific databases: the 1000 Genomes Project (http://annovar.openbioinformatics.org/en/latest/user-guide/filter/#1000-genomes-project-2015-aug-annotations) for allele frequency (AF) in populations; the LJB* databases for calculating SIFT scores, PolyPhen2 HDIV scores, PolyPhen2 HVAR scores, LRT scores, Mutation Taster scores, Mutation Assessor score, FATHMM scores, GERP++ scores, PhyloP scores, and SiPhy scores.

Based on the AF of each mutation in 16 samples and the AF in the 504 East Asians from the 1000 Genomes Project, we used R-3.0.0 software (http://web.mit.edu/people/jhaas/MacData/afs/sipb/project/r-project/arch/sun4x_59/lib/R/library/stats/html/fisher.test.html) for Fisher exact test. The resulting *P*-value and *q*-value were used as a basis for screening. The loci with *q*-value ≤ 0.05 were considered as statistically significant.

Linkage disequilibrium (LD) block analysis was performed for the multiple significant SNPs located near the same gene by using Haploview 4.2. A list of genes involved in statistically significant loci was then submitted to (http://software.broadinstitute.org/gsea/index.jsp) for gene set enrichment analysis (GSEA). False discovery rate (FDR) was calculated to obtain the significant gene sets.

## 3. Results

Basic information for 16 individuals with obesity is shown in Table 2. Through target genomic region capture sequencing, we obtained a total of 29,193,094 to 43,910,318 high-quality reads from these patients. The target region capture average ratio was 86.75%. All of the coverage ratios were over 92%, and the average depth of target regions was greater than 1500-fold. Therefore, sequencing coverage was fully adequate to detect gene variants within the majority of the targeted regions.

**Table 2**

**Sex-specific characteristics.**

| Characteristic | Male (n = 8) | Female (n = 8) | Total (n = 16) |
|---|---|---|---|
| Age, y | 36.36 ± 4.15 | 45.65 ± 12.81 | 41.01 ± 10.37 |
| Height, cm | 171.81 ± 2.67 | 157.06 ± 4.35 | 164.44 ± 8.38 |
| Weight, kg | 96.75 ± 5.33 | 82.98 ± 4.02 | 89.86 ± 8.45 |
| Systolic blood pressure, mm Hg | 123.13 ± 7.04 | 125.13 ± 5.67 | 124.12 ± 6.26 |
| Diastolic blood pressure, mm Hg | 78.75 ± 3.54 | 81.00 ± 5.71 | 79.88 ± 4.73 |
| Waist circumference, cm | 106.13 ± 6.99 | 100.13 ± 7.24 | 103.13 ± 7.54 |
| Hip circumference, cm | 110.63 ± 4.81 | 108.79 ± 4.09 | 109.71 ± 4.41 |
| Fasting blood glucose, mmol/L | 5.51 ± 0.55 | 5.24 ± 0.33 | 5.38 ± 0.46 |
| BMI, kg/m$^2$ | 32.78 ± 1.66 | 33.72 ± 2.70 | 33.25 ± 2.22 |

BMI = body mass index.

**Table 3**

**Sixty-five independent loci associated with obesity at $q \leq 0.05$.**

| Chr. | SNP ID | Ref./alt. alleles | Nearest gene | Mutation frequency in control (n = 504) | Mutation frequency in case (n = 16) | *q*-value |
|---|---|---|---|---|---|---|
| chr8 | rs10108428 | C/A | *PAG1* (intronic) | 0.86 | 0.50 | 0.038986 |
| chr8 | rs6988941 | C/T | *PAG1*, *FABP5* (intergenic) | 1.00 | 0.56 | 4.02E-18 |
| chr8 | rs7463736 | A/T | *PAG1*, *FABP5* (intergenic) | 1.00 | 0.63 | 7.57E-16 |
| chr8 | rs2955010 | C/T | *IMPA1* (intronic) | 0.87 | 0.41 | 2.23E-05 |
| chr8 | rs7830321 | G/A | *CHMP4C*, *SNX16* (intergenic) | 1.00 | 0.88 | 0.007744 |
| chr7 | rs145210930 | G/C | *INSIG1*, *BLACE* (intergenic) | 1.00 | 0.53 | 4.79E-21 |
| chr7 | rs367914749 | A/C | *INSIG1*, *BLACE* (intergenic) | 1.00 | 0.63 | 7.57E-16 |
| chr7 | rs149548396 | T/G | *INSIG1*, *BLACE* (intergenic) | 1.00 | 0.31 | 4.04E-34 |
| chr7 | rs4716993 | G/A | *BLACE*, *EN2* (intergenic) | 0.83 | 0.28 | 2.65E-07 |
| chr7 | rs6977647 | C/T | *EN2*, *CNPY1* (intergenic) | 0.88 | 0.38 | 7.17E-07 |
| chr7 | rs4716505 | A/G | *CNPY1*, *LOC100506302* (intergenic) | 0.73 | 0.28 | 0.002788 |
| chr7 | rs1636869 | G/A | *RBM33*, *SHH* (intergenic) | 0.92 | 0.56 | 0.002403 |
| chr7 | rs1636874 | T/C | *RBM33*, *SHH* (intergenic) | 0.89 | 0.38 | 1.94E-07 |
| chr7 | rs6459964 | T/C | *SHH*, *LOC389602* (intergenic) | 1.00 | 0.75 | 3.33E-09 |
| chr7 | rs898694 | T/C | *SHH*, *LOC389602* (intergenic) | 1.00 | 0.75 | 3.33E-09 |
| chr7 | rs11765221 | T/C | *SHH*, *LOC389602* (intergenic) | 1.00 | 0.88 | 0.007744 |
| chr7 | rs12698345 | C/T | *SHH*, *LOC389602* (intergenic) | 0.96 | 0.31 | 1.43E-17 |
| chr7 | rs62481754 | T/C | *SHH*, *LOC389602* (intergenic) | 0.96 | 0.63 | 0.00015 |
| chr7 | rs6956657 | T/C | *LOC389602*, *LOC285889* (intergenic) | 0.65 | 0.22 | 0.016584 |
| chr7 | rs6460015 | A/G | *LOC389602*, *LOC285889* (intergenic) | 1.00 | 0.44 | 3.41E-25 |
| chr7 | rs116435518 | T/G | *LOC285889*, *LINC01006* (intergenic) | 1.00 | 0.31 | 4.04E-34 |
| chr7 | rs6459716 | T/C | *LMBR1*, *NOM1* (intergenic) | 0.98 | 0.66 | 2.14E-06 |
| chr7 | rs1182423 | T/C | *UBE3C*, *DNAJB6* (intergenic) | 1.00 | 0.56 | 4.02E-18 |
| chr7 | rs1182422 | G/A | *UBE3C*, *DNAJB6* (intergenic) | 0.99 | 0.56 | 1.65E-11 |
| chr7 | rs7794423 | T/C | *DNAJB6* (intronic) | 0.74 | 0.31 | 0.007463 |
| chr7 | rs6974402 | C/T | *DNAJB6* (intronic) | 0.99 | 0.81 | 0.02242 |
| chr7 | rs7783247 | G/C | *LOC101927914* (ncRNA_intronic) | 0.98 | 0.72 | 0.002434 |
| chr7 | rs7802775 | A/G | *LOC101927914* (ncRNA_intronic) | 0.98 | 0.72 | 0.002434 |
| chr7 | rs2009450 | T/C | *LOC101927914* (ncRNA_intronic) | 0.97 | 0.47 | 3.39E-12 |
| chr7 | rs10281282 | T/C | *PTPRN2* (intronic) | 0.80 | 0.31 | 6.28E-05 |
| chr7 | rs10279215 | A/G | *PTPRN2* (intronic) | 1.00 | 0.75 | 3.33E-09 |
| chr7 | rs4475428 | G/T | *PTPRN2* (intronic) | 0.83 | 0.44 | 0.013143 |
| chr7 | rs13309736 | C/T | *PTPRN2* (intronic) | 0.88 | 0.22 | 4.38E-13 |
| chr7 | rs111359155 | G/A | *PTPRN2* (intronic) | 0.19 | 0.59 | 0.008378 |
| chr7 | rs71544556 | A/C | *PTPRN2* (intronic) | 0.01 | 0.28 | 3.14E-06 |
| chr7 | rs62480891 | C/T | *PTPRN2* (intronic) | 0.14 | 0.75 | 2.80E-10 |
| chr7 | rs4909181 | C/A | *PTPRN2* (intronic) | 0.00 | 0.19 | 0.000431 |
| chr7 | rs28645127 | A/C | *PTPRN2* (intronic) | 0.86 | 0.28 | 5.49E-09 |
| chr7 | rs11773238 | A/G | *PTPRN2* (intronic) | 0.77 | 0.28 | 0.000103 |
| chr7 | rs138997616 | C/A | *PTPRN2* (intronic) | 0.01 | 0.28 | 3.44E-07 |
| chr7 | rs4909069 | C/T | *PTPRN2* (intronic) | 0.88 | 0.44 | 5.57E-05 |
| chr7 | rs13247120 | G/C | *PTPRN2* (intronic) | 0.00 | 0.94 | 2.01E-50 |
| chr7 | rs11771431 | C/T | *PTPRN2* (intronic) | 1.00 | 0.81 | 0.001056 |
| chr7 | rs35665468 | A/G | *PTPRN2* (intronic) | 0.86 | 0.50 | 0.035758 |
| chr7 | rs34127805 | C/T | *PTPRN2* (intronic) | 0.99 | 0.81 | 0.02242 |
| chr7 | rs117626388 | A/G | *PTPRN2* (intronic) | 0.99 | 0.81 | 0.008045 |

| Chr. | SNP ID | Ref./alt. alleles | Nearest gene | Mutation frequency in control (n = 504) | Mutation frequency in case (n = 16) | *q*-value |
|---|---|---|---|---|---|---|
| chr7 | rs116990591 | G/A | *PTPRN2* (intronic) | 0.99 | 0.81 | 0.02242 |
| chr7 | rs28502325 | A/G | *PTPRN2* (intronic) | 1.00 | 0.69 | 4.67E-10 |
| chr7 | rs10949726 | A/G | *PTPRN2* (intronic) | 0.96 | 0.28 | 2.63E-19 |
| chr7 | rs11767650 | G/C | *PTPRN2* (intronic) | 0.96 | 0.28 | 2.63E-19 |
| chr7 | rs4909087 | T/C | *PTPRN2* (intronic) | 0.74 | 0.19 | 1.94E-06 |
| chr7 | rs9770794 | G/A | *NCAPG2* (intronic) | 0.94 | 0.47 | 1.59E-08 |
| chr7 | rs4909097 | T/C | *NCAPG2* (upstream) | 1.00 | 0.81 | 3.75E-05 |
| chr7 | rs2486025 | A/G | *ESYT2* (intronic) | 0.95 | 0.22 | 4.62E-21 |
| chr7 | rs75396698 | T/G | *WDR60* (intronic) | 0.94 | 0.25 | 2.67E-17 |
| chr7 | rs2527207 | A/G | *WDR60*, *LINC00689* (intergenic) | 0.94 | 0.50 | 1.75E-07 |
| chr7 | rs262142 | A/G | *WDR60*, *LINC00689* (intergenic) | 1.00 | 0.63 | 7.57E-16 |
| chr7 | rs2730252 | C/T | *VIPR2* (intronic) | 1.00 | 0.38 | 6.59E-29 |
| chr7 | rs2730251 | C/T | *VIPR2* (intronic) | 1.00 | 0.31 | 9.19E-33 |
| chr7 | rs382552 | C/G | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.88 | 0.007744 |
| chr7 | rs115829008 | A/G | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.44 | 1.82E-26 |
| chr7 | rs115317985 | C/T | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.38 | 3.17E-30 |
| chr7 | rs1640157 | T/G | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.44 | 1.82E-26 |
| chr7 | rs1640158 | A/G | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.44 | 1.82E-26 |
| chr7 | rs370300356 | C/G | *VIPR2*, *NONE* (intergenic) | 1.00 | 0.47 | 1.24E-24 |

SNP ID from dbSNP 138.

Alt. = alteration, *BLACE* = B-cell acute lymphoblastic leukemia expressed, *CHMP4* = charged multivesicular body protein 4B, Chr. = chromosome, *CNPY1* = canopy FGF signaling regulator 1, *DNAJB6* = DnaJ heat shock protein family (Hsp40) member B6, *EN2* = engrailed homeobox 2, *ESYT2* = extended synaptotagmin protein 2, *FABP5* = fatty acid binding protein 5, GO = genetic ontology, *IMPA1* = inositol monophosphatase 1, *INSIG1* = insulin induced gene 1, *LINC00689* = long intergenic nonprotein coding RNA 689, *LINC01006* = long intergenic nonprotein coding RNA 1006, *LMBR1* = limb development membrane protein 1, *LOC100506302* = uncharacterized LOC100506302, *LOC101927914* = uncharacterized LOC101927914, *LOC285889* = uncharacterized LOC285889, *LOC389602* = uncharacterized LOC389602, *NOM1* = nucleolar protein with MIF4G domain 1, *PAG1* = phosphoprotein membrane anchor with glycosphingolipid microdomains 1, *PTPRN2* = protein tyrosine phosphatase, receptor type N2, *RBM33* = RNA binding motif protein 33, Ref. = reference, *SHH* = Sonic Hedgehog, SNP = single nucleotide polymorphisms, *SNX16* = sorting Nexin 16, *UBE3C* = ubiquitin protein ligase E3C, VIP = vasoactive intestinal polypeptide, *VIPR2* = vasoactive intestinal peptide receptor 2, *WDR60* = WD repeat domain 60.

A total of 12,131 genetic variants across the 16 samples were found. After comparison with the 504 East Asians in the 1000 Genomes Project, 65 SNPs involved 25 genes were associated with obesity significantly at $q$-values $\leq 0.05$, as shown in Table 3 and Figs. 1 and 2.

As considering several SNPs were found located near the same gene, LD block analysis was conducted and identified 10 pairs of loci with $D' \geq 0.8$ and $r^2 \geq 0.8$ (Table 4). The LD pattern of mutations among 65 significant SNPs is shown in Supplementary File 3, http://links.lww.com/MD/B572.

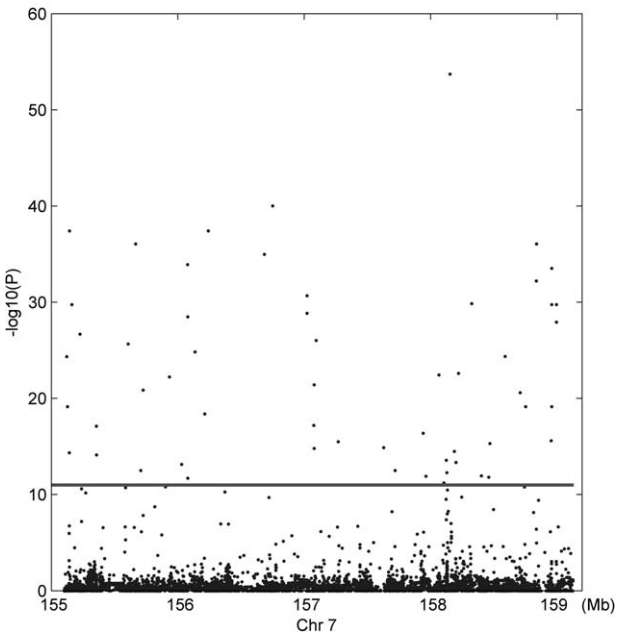


**Figure 1.** Manhattan plot showing the results of the association with obesity in 7q36.3.

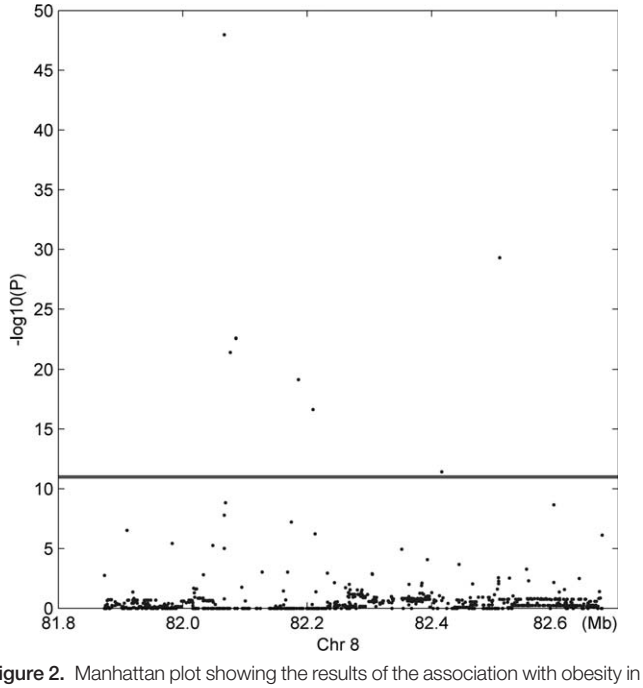


**Figure 2.** Manhattan plot showing the results of the association with obesity in 8q21.13.

**Table 4**

**Linkage disequilibrium block analysis with $D' \geq 0.8$ and $r^2 \geq 0.8$.**

| Loci 1 | Loci 2 | $D'$ (95% CI) | $r^2$ | Dist |
|---|---|---|---|---|
| rs6459964 | rs898694 | 1.00 (0.76, 1.00) | 1.00 | 18 |
| rs1182423 | rs1182422 | 1.00 (0.82, 1.00) | 1.00 | 7 |
| rs7783247 | rs7802775 | 1.00 (0.78, 1.00) | 1.00 | 1 |
| rs71544556 | rs138997616 | 1.00 (0.77, 1.00) | 1.00 | 10,363 |
| rs34127805 | rs117626388 | 1.00 (0.72, 1.00) | 1.00 | 13 |
| rs34127805 | rs116990591 | 1.00 (0.72, 1.00) | 1.00 | 14 |
| rs117626388 | rs116990591 | 1.00 (0.72, 1.00) | 1.00 | 1 |
| rs10949726 | rs11767650 | 1.00 (0.78, 1.00) | 1.00 | 3 |
| rs4909097 | rs2486025 | 1.00 (0.61, 1.00) | 0.82 | 91,735 |
| rs1640157 | rs1640158 | 1.00 (0.82, 1.00) | 1.00 | 2 |

$D'$ is the value of $D$ prime between the 2 loci.
$r^2$ is the correlation coefficient between the 2 loci.
Distance is the distance between the loci in base.
CI = confidence interval.

To identify potential enriched gene sets, a genetic ontology (GO) pathway analysis was performed. The overrepresented gene sets (GO gene sets) were as follows: lipid raft, lipid metabolic process, phosphoric monoester hydrolase activity, cellular protein metabolic process, cellular macromolecule metabolic process, plasma membrane part, protein metabolic process, phosphoric ester hydrolase activity, plasma membrane, signal transduction, cellular lipid metabolic process, membrane part, hydrolase activity acting on ester bonds, as shown in Table 5. As for FDR, although all gene sets were above 0.05, the gene sets of lipid raft and lipid metabolic process had FDR values <0.4.

## 4. Discussion

This is our first attempt to explore the obesity-related SNPs in target genomic regions as informed from our previous genome-wide linkage study. We were able to identify 65 SNPs in 25 genes in association with obesity. On pathway level, 2 major gene sets were suggested: LIPID_RAFT and LIPID_METABOLIC_PROCESS.

Among the 25 genes, *INSIG1*, *FABP5*, *PTPRN2*, *VIPR2* have been reported to be associated with obesity. *INSIG1* encodes an endoplasmic reticulum membrane protein which regulates cholesterol concentrations in cell. Studies[10,11] on *INSIG1* gene and obesity suggested the gene plays a critical role in feedback regulation of lipid metabolism and may be involved in obesity development. *FABP5* encodes the fatty acid binding protein found in epidermal cells and relevant pathways include glucose/energy metabolism. Canas et al[12] have found overweight prepubertal boys showed elevated FABP5. A possible mechanism is that FABP5 regulates diet-induced obesity via GIP (gastric inhibitory polypeptide) secretion from enteroendocrine K cells in response to fat ingestion.[13]*PTPRN2* plays a role in insulin secretion in response to glucose stimuli. Interestingly, most studies in Chinese have shown that the obesity-predisposing alleles were associated with insulin secretion, which is distinctive from the observations reported in Caucasians, among whom the obesity-related loci were primarily associated with insulin resistance.[14] GO annotations related to *VIPR2* include G-protein coupled receptor activity and vasoactive intestinal polypeptide (VIP) receptor activity. Note that none of the individual genes in the VIP pathway reached the genome-wide significance level in single-marker GWAS on obesity, however

**Table 5**

**Gene set enrichment analysis.**

| Database | Gene set | Genes in overlap | GSEA, *P*-value | FDR |
|---|---|---|---|---|
| GO cellular component | LIPID_RAFT | *SHH*, *PAG1* | 8.01E-5 | 1.17E-1 |
| GO biological process | LIPID_METABOLIC_PROCESS | *SHH*, *IMPA1*, *FABP5* | 4.24E-4 | 3.08E-1 |
| GO molecular function | PHOSPHORIC_MONOESTER_HYDROLASE_ACTIVITY | *IMPA1*, *PTPRN2* | 1.20E-3 | 4.03E-1 |
| GO biological process | CELLULAR_PROTEIN_METABOLIC_PROCESS | *SHH*, *PTPRN2*, *DNAJB6*, *UBE3C* | 1.49E-3 | 4.03E-1 |
| GO biological process | CELLULAR_MACROMOLECULE_METABOLIC_PROCESS | *SHH*, *PTPRN2*, *DNAJB6*, *UBE3C* | 1.56E-3 | 4.03E-1 |
| GO cellular component | PLASMA_MEMBRANE_PART | *SHH*, *PAG1*, *PTPRN2*, *VIPR2* | 1.7E-3 | 4.03E-1 |
| GO biological process | PROTEIN_METABOLIC_PROCESS | *SHH*, *PTPRN2*, *DNAJB6*, *UBE3C* | 2.13E-3 | 4.03E-1 |
| GO molecular function | PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY | *IMPA1*, *PTPRN2* | 2.22E-3 | 4.03E-1 |
| GO cellular component | PLASMA_MEMBRANE | *SHH*, *PAG1*, *PTPRN2*, *VIPR2* | 3.62E-3 | 5.85E-1 |
| GO biological process | SIGNAL_TRANSDUCTION | *SHH*, *PAG1*, *IMPA1*, *VIPR2* | 5.87E-3 | 7.45E-1 |
| GO biological process | CELLULAR_LIPID_METABOLIC_PROCESS | *SHH*, *IMPA1* | 6.01E-3 | 7.45E-1 |
| GO cellular component | MEMBRANE_PART | *SHH*, *PAG1*, *PTPRN2*, *VIPR2* | 6.34E-3 | 7.45E-1 |
| GO molecular function | HYDROLASE_ACTIVITY_ACTING_ON_ESTER_BONDS | *IMPA1*, *PTPRN2* | 6.66E-3 | 7.45E-1 |

*DNAJB6* = DnaJ heat shock protein family (Hsp40) member B6, *FABP5* = fatty acid binding protein 5, FDR = false discovery rate, GO = genetic ontology, GSEA = gene set enrichment analysis, *IMPA1* = inositol monophosphatase 1, *PAG1* = phosphoprotein membrane anchor with glycosphingolipid microdomains 1, *PTPRN2* = protein tyrosine phosphatase, receptor type N2, *SHH* = Sonic Hedgehog, *UBE3C* = ubiquitin protein ligase E3C, *VIPR2* = vasoactive intestinal peptide receptor 2.

one study via pathway-based analysis of GWA-data[15] suggested the VIP pathway was important for obesity.

Although there was no strong indication that *WDR60*, *SHH*, *UBE3C*, or *LMBR1* polymorphism was the main causal variant of obesity in the population, studies showed that variation in these genes may be part of the multifactorial etiology of this complex condition. *WDR60* encodes a member of the WD repeat protein family which plays a role in formation of cilia. The influences of cilia-related genes on adipogenesis via retrograde transport of SHH receptors and SHH signaling have been reported in a most recent review.[16] *LMBR1* and *UBE3C* were known to be related to Coenzyme Q10 Deficiency, Primary, 2 and Kabuki Syndrome 1, respectively. The clinical symptoms for both diseases include obesity.

The rest of the obesity-related genes we have identified, including *PAG1*, *IMPA1*, *CHMP4*, *SNX16*, *BLACE*, *EN2*, *CNPY1*, *LOC100506302*, *RBM33*, *LOC389602*, *LOC285889*, *LINC01006*, *NOM1*, *DNAJB6*, *LOC101927914*, *ESYT2*, *LINC00689* have not been reported. The biological function of these particular variants remains to be characterized.

Our data are the first documentation via target capture sequencing to identify obesity related rare variants in Northern Han Chinese. The highly interesting genomic regions were derived from our previous genome-wide linkage study in dizygotic twins from the same population, that could serve as important prior information. A major limitation of our study was relatively small sample size and it is desirable to replicate our findings in other studies.

In sum, based on our previous genome-wide linkage study, we identified genes and gene sets associated with adult obesity in 7q36.3 and 8q21.13 chromosome regions through targeted resequencing technology. We believe findings of this study contribute to further replication and functional studies.

## References

[1] Sandholt CH, Hansen T, Pedersen O. Beyond the fourth wave of genome-wide obesity association studies. Nutr Diabetes 2012;2:e37.
[2] Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197–206.
[3] Cohen JC, Kiss RS, Pertsemlidis A, et al. Multiple rare Alleles contribute to low plasma levels of HDL cholesterol. Science 2004;305:869–72.
[4] Li ZM, Lin Q, Huang WQ, et al. Target gene capture sequencing in Chinese population of sporadic Parkinson disease. Medicine 2015;94: e836.
[5] Paz G, Boguszewski MCS, Mastronardi CA, et al. Whole exome sequencing of extreme morbid obesity patients: translational implications for obesity and related disorders. Genes (Basel) 2014;5:709–25.
[6] Bonnefond A, Durand E, Sand O, et al. Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. PLoS ONE 2010;5:e13630.
[7] Wei XM, Ju XC, Yi X, et al. Identification of sequence variants in genetic disease-causing genes using targeted next-generation sequencing. PLoS ONE 2011;6:e29500.
[8] Zhang DF, Pang ZC, Li SX, et al. High-resolution genome-wide linkage mapping identifies susceptibility loci for BMI in the Chinese population. Obesity 2012;20:830–3.
[9] Kettunen J, Perola M, Martin NG, et al. Multicenter dizygotic twin cohort study confirms two linkage susceptibility loci for body mass index at 3q29 and 7q36 and identifies three further potential novel loci. Int J Obesity 2009;33:1235–42.
[10] Smith EM, Zhang Y, Baye TM, et al. INSIG1 influences obesity-related hypertriglyceridemia in humans. J Lipid Res 2010;51:701–8.
[11] Liu FH, Song JY, Shang XR, et al. The gene-gene interaction of INSIG-SCAP-SREBP pathway on the risk of obesity in Chinese children. Biomed Res Int 2014;2014:538564.
[12] Canas JA, Damaso L, Hossain J, et al. Fatty acid binding proteins 4 and 5 in overweight prepubertal boys: effect of nutritional counselling and supplementation with an encapsulated fruit and vegetable juice concentrate. J Nutr Sci 2015;4:e39.
[13] Shibue K, Yamane S, Harada N, et al. Fatty acid-binding protein 5 regulates diet-induced obesity via GIP secretion from enteroendocrine K cells in response to fat ingestion. Am J Physiol Endocrinol Metab 2015;308:E583–91.
[14] Liang J, Sun Y, Liu X, et al. Genetic predisposition to obesity is associated with insulin secretion in Chinese adults: the Cardiometabolic Risk in Chinese (CRC) study. J Diabetes Complications 2016;30:1229–33.
[15] Liu YJ, Guo YF, Zhang LS, et al. Biological pathway-based genome-wide association analysis identified the vasoactive intestinal peptide (VIP) pathway important for obesity. Obesity (Silver Spring) 2010;18: 2339–46.
[16] Mariman EC, Vink RG, Roumans NJ, et al. The cilium: a cellular antenna with an influence on obesity risk. Br J Nutr 2016;116:576–92.