Coding Variants Affecting the Expression of Obesity-related Genes for Pediatric Adiposity

Ke Mao[1], Meixian Zhang[2], Jinshuai Cao[1], Xiaoyuan Zhao[2], Liwang Gao[2], Liwan Fu[3], Hong Cheng[2], Chun Yan[1], Xiaopeng Xu[1], Xiaofeng Shi[1], Zhuoyuan Jiang[1], Bingqing Wang[1], Yong-Biao Zhang[1], and Jie Mi[3]

[1]Beihang University
[2]Department of Epidemiology, Capital Institute of Paediatrics
[3]Beijing Children's Hospital, Capital Medical University, National Center for Children's Health

June 25, 2020

Abstract

Objective Heredity has a remarkable effect on obesity in an obesogenic environment. Despite numerous genetic variants contributing to obesity-related traits, none of them was identified from Chinese children. We aim to identify novel variants and genes associated with childhood obesity in China. Methods We obtain promising single nucleotide variants from 76 obese and 74 normal weight children by whole-exome sequencing, and interrogate their associations with obesity traits in an additional 6,334 children cohort. We then depict the effects of genome-wide significant ($P < 5E\text{-}8$) variants on expression of implicated genes in blood and adipose tissues by transcriptome sequencing. Results We identify two coding variants associated with obesity at genome-wide significance: rs1059491 ($P = 2.57E\text{-}28$) in *SULT1A2* and rs189326455 ($P = 8.98E\text{-}12$) in *MAP3K21*. In addition, rs1058491 is also significantly associated with several obesity traits. Transcriptome sequencing demonstrates that rs1059491 is an eQTL site associated with the expression levels of several obesity-related genes, such as *SULT1A2*, *ATXN2L*, and *TUFM*. Conclusions Our work identifies two coding variants affecting the expression of obesity-related genes based on a large Chinese pediatric adiposity cohort and provide new insights for the pathophysiology of Chinese childhood obesity.

Introduction

Childhood obesity is one of the most serious public health challenges worldwide. With a high estimated heritability(Elks et al., 2012), it increases the risk of many physical and mental conditions and contributes to the global burden of chronic diseases including type 2 diabetes, cardiovascular diseases, and several types of cancer(Levi et al., 2017; Twig et al., 2016; Twig et al., 2018). Over the past four decades, the number of obese children and adolescents (aged 5 to 19 years) worldwide has risen more than tenfold, from 11 million in 1975 to 124 million in 2016(Collaboration, 2017). Childhood obesity prevalence in China rocketed up since the 90s due to its rapid economic growth and industrialization. According to the Global Burden of Disease study, China has the highest number of obese children (15.3 million), and the second highest number of obese adults (57.3 million) in the world after US in 2015(Collaborators et al., 2017). About 70% of obese adolescents become obese adults, and they are more likely to develop diabetes and cardiovascular diseases at a younger age(Simmonds, Llewellyn, Owen, & Woolacott, 2016; Zhao & Grant, 2011).

Obesity results from a complex interplay of various genetic and environmental factors. Over the past decade, genome-wide association studies (GWASs) have successfully identified numerous common variants associated

with obesity-related traits, such as body mass index (BMI), waist-to-hip ratio, and body fat index(Akiyama et al., 2017; Locke et al., 2015; Shungin et al., 2015). Despite the progress, much of the obesity heritability remains yet to be explored. So far the identified single-nucleotide variants (SNVs) associated with obesity tend to be common, non-coding variants with small effect sizes, and the function and pathophysiology of these genetic variants are largely uncertain. Low-frequency or rare variants with large effects that may account for unexplained heritability in common obesity have not been systematically investigated in Chinese population. Exome sequencing has recently identified several large effects coding variants associated with obesity. Two SNVs, rs7238987 (*CYB5A*) and chr12:120990399 (RNF10:p.R151H), were identified as obesity-associated variants in American Pima Indians(Huang et al., 2014). SNVs of rs62623713 (*SYPL2*) and rs2076349 (*LAMB3*) were identified in morbidly obese European adults(Jiao et al., 2015; Jiao et al., 2016). A family-based association study in Hispanic children identified rs141510219 (*PEX1*) which was associated with several obesity traits(Sabo et al., 2017). However, most studies were carried out in adults, for whom the genetic influence of obesity-related genes might be compromised by diet, age, pregnancy, lifestyle and other factors. Studies on children were, on the other hand, exempt from the influence of those environmental factors to the utmost extent.

Here, we used whole-exome sequencing (WES) to detect novel obesity-susceptible gene loci in 150 Chinese children and further validated them in 6,334 children. We aim to identify significantly obesity associated coding variants. Association between variants and obesity traits such as BMI, fat mass percentage (FMP), fat mass index (FMI), and fat free mass index (FFMI) were also tested. Transcriptome sequencing were also performed in blood and adipose tissues, and the possible mechanisms of discovered variants in obesity pathology were investigated.

Material and Methods

Subjects

All subjects in whole-exome sequencing, subsequent validation, and whole blood transcriptome sequencing were obtained from two large school-based cohorts, the Beijing Child and Adolescent Metabolic Syndrome Study (BCAMS), and the China Child and Adolescent Cardiovascular Health Study (CCACH)(Liu et al., 2017; Shan et al., 2010). Adipose tissues were obtained from Plastic Surgery Hospital, Chinese Academy of medical Sciences. Individuals with hormone treatment, secondary obesity due to endocrinopathy and serious intercurrent illness were excluded. All cases and controls were unrelated Han Chinese in Beijing, China. Replication and validation for promising variants were performed on 6,334 (2,480 obese and 3,854 normal weight) children. Written informed consent were obtained from the parents or guardians of all subjects. The study was approved by the ethics committee and institutional review board of the Capital Institute of Paediatrics (IEC-C-008-A08-V.05.1).

Phenotype

Normal weight and common obesity were defined by age- and sex-specific BMI cutoff points recommended by the International Obesity Task Force (IOTF)(Cole, Flegal, Nicholls, & Jackson, 2007). Adiposity was defined as body fat mass percentage greater than or equal to 20% for boys and 25% for girls aged [?] 14 years and 30% for girls aged $>$ 14 years(Chen, Lu, & Department of Disease Control Ministry of Health, 2004). Weight-to-Height ratio (WHtR) was calculated to define abdominal obesity using a boundary value of 0.5(Browning, Hsieh, & Ashwell, 2010). Dyslipidemia was defined by serum total cholesterol (TC) [?] 5.20 mmol/L, or triacylglycerol (TG) [?] 1.70 mmol/L, or high-density lipoprotein cholesterol (HDL-C) [?] 1.04 mmol/L, or low-density lipoprotein cholesterol (LDL-C) [?] 3.37 mmol/L.

Whole-exome sequencing

Genomic DNA was isolated from peripheral blood leukocytes using the QIAamp DNA blood kit (Qiagen, Germany). Each DNA library was prepared from 5μg of qualified genomic DNA which was sheared to 180-200 base pairs. The SureSelect Human All Exon V4+UTR Kit (Agilent Technologies, USA) was used to capture 71 Mbps of exons and untranslated regions (UTRs) according to the manufacturer's protocol. Paired-end sequencing (2×100 bp) was carried out with the HiSeq 2500 Sequencing System (Illumina Inc., USA) at Berry Genomics, Co., Ltd.

Quality control and variant calling

Trimmed reads were aligned to the human reference genome (hg19) using the Burrows-Wheeler Alignment tool (v0.7.17)(H. Li & Durbin, 2009), with mapping rates above 99.4%, and > 90% of the target regions were completely covered with at least 10 × depth. Samples were required to reach at least 20 × coverage over 70% of the exome target. Duplicates were removed by Picard (v2.0.1).

Sequence variation, including SNVs and insertion/deletions, were detected using Genome Analysis Toolkit (v4.1.3.0)(McKenna et al., 2010). We excluded variants with a call rate of < 95% or Hardy-Weinberg equilibrium (HWE) P < 1.0E-6. Samples with a call rate lower than 95% or a heterozygosity rate more than three standard deviations away from the mean were removed. Samples were clear of gender-mismatches or relativeness.

Variants for validation were achieved through several filtering and comparison steps (**Figure 1**). InterVar (v2.0.1) was used for annotation(Q. Li & Wang, 2017). The highest impact effect was taken for variants that have different annotations due to multiple transcripts. SIFT, PolyPhen-2, MutationTaster, and FATHMM were used to predict the possible impact of coding variants on protein function(Yang & Wang, 2015).

Association analysis

Association test was performed with PLINK (v1.90)(Chang et al., 2015). Subjects were characterized according to their BMI and related anthropometrics. Association analysis was performed using linear or logistic regression assuming an additive model with sex and age as covariates. To weaken the effect of population stratification, we also included geographic information as a covariate because the maximal known stratification for Chinese is northern and southern ancestry. To prove the associated SNVs were independent, we performed a conditional analysis on each one by adjusting other significant associated SNVs or previous reported SNVs. Correction for multiple tests was performed using the PLINK adjust function, with genomic control-corrected Pvalues being calculated based on the genotypes of all variants in the final analysis. The threshold for genome-wide significance was set at aP value $< 5.0 \times 10^{-8}$.

Genotyping and validation

Candidate SNVs showing nominal association ($P < 0.05$) with obesity in the first stage were genotyped in an additional large case-control cohort consisting of 2,480 obese children and 3,854 controls. Genotyping was performed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (SEQUENOM) at Beijing Compass Biotechnology Co., Ltd. Multiplexed assays were designed using MassARRAY Assay Design v3.1. Allele-specific base extension was performed according to recommendations from Agena Bioscience. Sanger sequencing was performed to validate the candidate variants identified by WES and replication analysis.

Transcriptomic sequencing

Whole blood were obtained from 58 cases and 43 controls. Whole blood RNA were extracted using PAXgene® Blood RNA Kit according to its handbook protocols. Frozen subcutaneous adipose tissue were

obtained from 11 children during plastic surgeries. RNA from adipose tissues were extracted using mirVana miRNA Isolation Kit. Total cDNA library was constructed using NEB Next(r) Ultra RNA Library Prep Kit and paired-end sequencing was performed on Illumina novaseq6000 platform. Reads alignment, transcripts including novel splice variants assembly, abundance computation of these transcripts, and differentially expressed genes and transcripts compare were analyzed with a comprehensive software package of HISAT(Kim, Langmead, & Salzberg, 2015), StringTie(Pertea et al., 2015), and DESeq2(Love, Huber, & Anders, 2014). We focus on the expression pattern of genes, such as*SULT1A2* (NC_000016.10), *MAP3K21* (NC_000001.11), etc. which potentially affected by rs1059491 and rs189326455.

Functional analysis

We used MEME suite to predict the effect of SNPs on the binding affinity of transcription related factors. To evaluate the structural and conformational change of rs1059491 (SULT1A2: p.N235T), we used PSIPRED for secondary structure prediction. The differential expression data of eQTL rs1059491 in adipose tissues and blood were obtained from both transcriptome sequencing and GTEx website (https://gtexportal.org/). Visualization was achieved by R (www.r-project.org/).

Results

Clinical characteristics of subjects

Here, we collected the largest Chinese children obesity cohort, as far as we know, composed of 6,484 subjects for association study. We performed WES on 150 subjects (76 cases vs 74 controls) to discover the obesity-associated candidate variants and further validated them in a 6,334 subjects (2,480 cases vs 3,854 controls). The characteristics of all collected subjects were shown in **Table1** . For 150 subjects in discovery stage, sex, age, and birth weight were roughly matched between cases and controls ($P > 0.05$, Chi-square test), but obesity-related traits as weight, BMI, FMP, WHtR, serum lipid level (triacylglycerol and high-density lipoprotein cholesterol), and blood pressure were significantly discrepant ($P < 0.001$, Chi-square test). Similar characteristic patterns were observed between cases and controls in validation cohort except parameters as follows: age, sex, fat-free mass index, total cholesterol, and LDL-C (each has $P < 0.001$, Chi-square test).

Identification of obesity-associated variants

To discover the obesity associated variants, especially for those causal ones, we performed whole-exome sequencing on 76 obesity children and 74 controls at discovery stage. We totally obtained 6,725,533 variants from the 150 subjects with an average depth of 23x. After quality control (**Figure 1**), we finally got 921,287 variants for association test. For all variants showed nominal ($P < 0.05$) association to obesity, we just kept those lose-of-function and deleterious ones as candidates for further investigation. In the end, we identified 26 obesity associated candidates for validation stage. We genotype all candidates in a cohort of 6,334 individuals including 2,480 cases and 3,854 controls. After genotyping these candidates and quality control on call rate, we totally got 23 SNVs for association study by adjusting age, sex, and population stratification (**Figure 1** and**Supplementary Table 1**).

We performed logistic regression on the 23 variants and found 2 of them significantly associated with common obesity at genome-wide level: rs1059491 (*SULT1A2* :c.704T>G, P = 7.71E-24,OR = 2.296, 95% CI = 1.953-2.699) and rs189326455 (*MAP3K21* :c.162G>C, P = 6.16E-11, OR = 0.2187, 95% CI = 0.1387-0.3449) (**Table 2**). Here, we used BMI to define common obesity according to the WHO and IOTF standards. When using FMP as phenotype of the 'fatty obesity', only rs1059491 associated with obesity at genome-wide significant level (P = 1.17E-11, OR = 1.935, 95% CI = 1.599-2.342). Meanwhile, rs1059491 was also significantly (P = 5.01E-10,OR = 1.66, 95% CI = 1.415-1.948) associated with dyslipidaemia. Furthermore, we performed association test on genotype of rs1059491 from transcriptome of blood and adipose tissues, and found it significantly associated with obesity (P = 0.01408, OR = 3.533, 95% CI =

1.29-9.675). When combing WES, validation, and transcriptome data, we found no significant heterogeneity existed and rs1059491 (P = 2.57E-28, OR = 2.405, 95% CI = 2.058-2.811) and rs189326455 (P = 8.98E-12, OR = 0.2122, 95% CI = 0.1359-0.3313) became more significant in association with obesity in Chinese children.

To further examine whether the identified variants affect obesity related quantitative traits as BMI, FMI, FMP, FFMI, and LDL, we performed multiple linear regression analyses (**Table 3**). After adjusting by sex, age, and population stratification, rs1059491 demonstrated genome-wide significant associations with BMI (P = 1.41E-18, β = 1.953, 95% CI = 1.519 - 2.386), FMP (P = 7.22E-13, β = 3.282, 95% CI =4.175 - 7.197), FMI (P = 1.87E-11, β = 1.157, 95% CI = 0.82 - 1.493), FFMI (P = 8.14E-09, β = -0.8541, 95% CI = -1.144 - -0.5643), and LDL level (P = 5.42E-09, β = 0.1583, 95% CI = 0.1052 - 0.2114), while rs189326455 just showed genome-wide significant association with BMI (P = 3.36E-10,β = -2.655, 95% CI = -3.481 - -1.828, **Table 3**).

Functional predictions of genome-wide significant variants

From the association study, we identified two novel obesity associated SNVs in Chinese children. The rs1059491 (SULT1A2:p.N235T) is a missense variant on the exon 2 of *SULT1A2* (**Figure 2A**). The rs189326455 (MAP3K21:p.E54D) is a missense variant on the exon 1 of*MAP3K21* (**Figure 2B**). Both SNVs are highly conserved across mammals (**Figure 2A,B**) and rs1059491 is predicted as a deleterious mutation according to SIFT, Polyphen-2 and MutationTaster (**Supplementary Table 2**). However, neither SNV was predicted to affect on the secondary structure of SULT1A2 or MAP3K21 (**Supplementary Figure 1**). Therefore, neither of them might affect phenotype by changing protein conformation.

Considering both SNVs reside on promoter or promoter flanking regions (**Figure 2A,B**), we interrogated their roles in regulating gene expression. The rs1059491 has been reported to enhance the binding affinities to transcription factors like PPAR_2 and RXRA which were involved in pathway of regulation of lipid metabolism by peroxisome proliferator-activated receptor alpha (**Supplementary Table 3**). MEME suite also predicted that rs1059491 has affected the binding affinity of transcription factors like RXRA homodimer, PPARG, PPARG::RXRA, VDR, and NR1H2::RXRA (**Figure 2C**). The rs189326455 resided in the binding sites of transcription factors including POL2, CHD2_disc3, and ESR2(Consortium, 2011; Kheradpour & Kellis, 2014), which would drastically reduce the binding affinity of CHD2 and ESR2 (**Supplementary Table 3**). Taken together, both rs1059491 and rs189326455 may affect the expression of certain genes by changing the binding affinity to transcription factors.

The role of rs1059491 in gene expression

According to GTEx Portal, rs1059491 is an eQTL that related to the differential expression of multiple genes and transcripts (**Suplementary Table 4**). The genotype of rs1059491 is correlated with expression of 17 genes including *SULT1A2* , *SULT1A1* ,*EIF3C, LAT,* and *SH2B1* in subcutaneous adipose tissues data or blood data deposited in GTEx. From our transcriptome sequencing data, we confirmed that the genotypes of rs1059491 correlated with the expression of *SGF29* , *SPNS1* , *SULT1A2* , and*TUFM* in blood, and *EIF3C* in adipose tissues (**Figure 3A**).

We further examined the expression of those eQTL target genes between obese and normal weighted children in blood and adipose tissues. We found that *NPIPB9* were significantly differentially expressed between cases and controls in blood, whereas *ATXN2L* ,*SULT1A2* , and *TUFM* were differentially expressed in adipose tissues (**Figure 3B**). *SULT1A2* and *TUFM* were significantly up-regulated in obese individuals, while *ATXN2L* and*NPIPB9* were down-regulated. In conclusion, rs1059491 regulated the expression of several genes in blood and adipose tissue, which may have accounted for its association with obesity.

Discussion

Considering severe early onset obesity might enrich more low-frequency deleterious variants, we performed whole-exome sequencing and association study on obesity in Chinese children. We found two missense mutations (rs1059491 and rs189326455) are significantly associated with common obesity. The rs1059491 is not only associated with BMI, but also in association with FMI, FMP, FFMI, and LDL. It might result in obesity by affecting the expression level of target genes other than alternating protein structure because it is predictively unharmful to protein structure but acts as an eQTL that can dramatically change the expression of several obesity-related genes evidenced by transcriptome sequencing.

Previous GWAS studies have identified multiple SNVs associated with increased BMI(Cotsapas et al., 2009; Locke et al., 2015; Speliotes et al., 2010; Thorleifsson et al., 2009). We noted that our most significant SNV, rs1059491, was located in a previously reported obesity locus with lead SNV of rs7359397 which was identified in European(Speliotes et al., 2010). However, a fine mapping study on the locus failed to find that rs1059491 was in association with obesity in European children and adolescents(Volckmar et al., 2015). Explanations for such contradiction are as follows three reasons. Firstly, although rs1059491 is highly linked ($r^2 > 0.75$) to rs7359397 in European, no linkage disequilibrium ($r^2 = 0.00029$) was observed between them in Chinese. Secondly, the allele frequency of rs1059491 varies drastically among different populations, which has been exemplified in 1,000 Genomes Project (1KG) (**Supplementary Figure 2**). Minor allele frequency was 0.3434 in European while was around 0.1 in Asian populations. In our findings, minor allele frequency was 0.0421 in the control group, which was similar to that of 0.0571 in southern Han Chinese in 1KG. Thirdly, our conditional analysis found that rs1059491 (conditional P=7.13E-21) was independently associated with obesity in Chinese children when controlling rs7359397. Taken together, rs1059491 was a population specific obesity-associated SNV and significantly associated with children obesity in Chinese.

The rs1059491 was a conserved missense mutation and was predicted as deleterious (PolyPhen score = 1), while our findings show that it might result in obesity through altering gene expression other than affecting protein structure. Chip-seq have revealed that rs1059491 resided in transcription factors binding sites of PPARγ2 and RXRA in ENCODE datasets(Kheradpour & Kellis, 2014). Irregular behavior of PPARγ2 and RXRA can result in abnormal lipid metabolism which lead to obesity(Akyurek et al., 2013; Lima et al., 2013; Stossi et al., 2019; Takahashi, Morita, Yokoyama, Suzuki, & Yamamoto, 2012). Therefore, rs1059491 might affect the binding affinity to obesity-related transcription factors and subsequently alter obesity-related gene's expression. Our transcriptome data and GETx data have shown that rs1059491 is an eQTL affecting the expression of many genes, such as*SULT1A2* , *ATXN2L* , and *TUFM* . It's worthy to note that*SULT1A2* and *ATXN2L* were strongly implicated in obesity(Voisin et al., 2015; Volckmar et al., 2015).*SULT1A2*is up-regulated in adipose tissues of individuals carrying G allele of rs1059491, and high expression level of *SULT1A2* is a characteristic not only in lipid uptake tissues as duodenum and intestine, but also in fat tissues (GTEx database). On the other hand, rs1059491 regulates *ATXN2L* on translational level, and subsequently affects its interaction partner *ATXN2* , the down-regulation of which would likely lead to increased body size and fat levels in *Caenohabditis elegans* (Bar et al., 2016; Kaehler et al., 2012), and *ATXN2* knock-out mice show adult-onset obesity(Kiehl et al., 2006; Lastres-Becker et al., 2008; Meierhofer, Halbach, Sen, Gispert, & Auburger, 2016). In brief, the most likely mechanism of rs1059491 association with obesity is expression regulation of obesity-related genes on the transcription level and post-transcription level (**Figure 4**).

We also found that rs189326455 was another variant significantly associated with obesity. It located in the SH3 domain of MAP3K21, which was responsible for controlling or regulating protein-protein interactions in the signal transduction pathways, such as cytoplasmic signaling(Koch, Anderson, Moran, Ellis, & Pawson, 1991; Schlessinger, 1994). MAP3K21 is a member of the mixed lineage kinases family that acts as a specific modulator to inhibit lipopolysaccharide (LPS)-induced activation of c-Jun N-terminal kinase (JNK), or extracellular signal-regulated kinase (ERK)(Craige, Reif, & Kant, 2016), or as a negative regulator of Toll-like receptor 4 (TLR4) signaling(Seit-Nebi, Cheng, Xu, & Han, 2012). It's well known that JNK, ERK, and TLR4 are all implicated with obesity through elevating gene expression level(Ahmad et al., 2012; Hirosumi et al., 2002), promoting insulin resistance(Hirosumi et al., 2002; Ozaki et al., 2016; Pal et al., 2012; Solinas

& Becattini, 2017), or stimulating adipogenic differentiation(Gu et al., 2015). It's worthy to note that rs189326455 is also suggestively ($P = 5.63E-07$) associated with blood triacylglycerol levels which is highly associated with insulin resistance(Ormazabal et al., 2018). Thus, we speculate that rs189326455 might alter the affinity of SH3 domain and subsequently affects interaction between MAP3K21 and mis-activating down-stream signaling pathway proteins, such as JNK, ERK, and TLR4 (**Figure 4**), and results in insulin resistance and obesity(Seit-Nebi et al., 2012).

In this study, we investigated causal variants in severe early onset obesity by taking advantage of that child obesity is less affected by cumulative environmental factors and the heritability is higher compared with adulthood obesity. We identified two potential variants related to obesity with whole exome sequencing. We found that rs1059491 was an eQTL which might result in obesity through altering the expression of obesity-related genes, and rs189326455 might affect the regulation of MAP3k21 to subsequent obesity-related genes. Although the speculation should be proved by more functional experiments, our study is among the first to use exome sequencing technology to identify functional variants of both common obesity and abdominal obesity in Chinese children. Chinese people are more likely to develop abdominal obesity, which has been associated with a higher risk of developing type 2 diabetes and cardiovascular disease(Thomas et al., 2004). Therefore, identification of the SNVs associated with abdominal obesity is of great impact on prevention and control of adiposity-based chronic diseases.

Acknowledge. This work was supported by grants from the Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education(KZ202010025039) and National Natural Science Foundation of China (81973110, 81502872, and 31671312)

Author Contributions. K.M., M.Z., and J.C. wrote the manuscript, researched, analysed and interpreted the data, contributed to the discussion, and reviewed/edited the manuscript. J.M., and Y.Z. designed the study, and made critical revisions of the manuscript. C.M. revised the figures. M.Z., X.Z., L.G., and L. F. were involved in the sample collection, selection and preparation. H.C. collected the phenotype data. W.L., and Z.Z. designed the bioinformatics and experimental sections, coordinated the data collection, maintained project procedure and carried out bioinformatics analysis. M.Z. and L.G. performed the genotyping analysis. X.X and X.S. made critical revisions of the manuscript. All the authors contributed to the final paper.

Duality of Interest. The authors have nothing to declare.

Data Availability Statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmad, R., Al-Mass, A., Atizado, V., Al-Hubail, A., Al-Ghimlas, F., Al-Arouj, M., . . . Behbehani, K. (2012). Elevated expression of the toll like receptors 2 and 4 in obese individuals: its significance for obesity-induced inflammation. *J Inflamm (Lond), 9* (1), 48. doi:10.1186/1476-9255-9-48

Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., . . . Kamatani, Y. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population.*NATURE GENETICS, 49* (10), 1458-1467. doi:10.1038/ng.3951

Akyurek, N., Aycan, Z., Cetinkaya, S., Akyurek, O., Yilmaz Agladioglu, S., & Ertan, U. (2013). Peroxisome proliferator activated receptor (PPAR)-gamma concentrations in childhood obesity. *SCANDINAVIAN JOURNAL OF CLINICAL & LABORATORY INVESTIGATION, 73* (4), 355-360. doi:10.3109/00365513.2013.786121

Bar, D. Z., Charar, C., Dorfman, J., Yadid, T., Tafforeau, L., Lafontaine, D. L., & Gruenbaum, Y. (2016). Cell size and fat content of dietary-restricted Caenorhabditis elegans are regulated by ATX-2, an mTOR repressor. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 113* (32), E4620-4629. doi:10.1073/pnas.1512156113

Browning, L. M., Hsieh, S. D., & Ashwell, M. (2010). A systematic review of waist-to-height ratio as a screening tool for the prediction of cardiovascular disease and diabetes: 0.5 could be a suitable global boundary value. *NUTRITION RESEARCH REVIEWS, 23* (2), 247-269. doi:10.1017/S0954422410000144

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience, 4* , 7. doi:10.1186/s13742-015-0047-8

Chen, C., Lu, F. C., & Department of Disease Control Ministry of Health, P. R. C. (2004). The guidelines for prevention and control of overweight and obesity in Chinese adults. *BIOMEDICAL AND ENVIRONMENTAL SCIENCES, 17 Suppl* , 1-36.

Cole, T. J., Flegal, K. M., Nicholls, D., & Jackson, A. A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ, 335* (7612), 194. doi:10.1136/bmj.39238.399444.55

Collaboration, N. C. D. R. F. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *LANCET, 390* (10113), 2627-2642. doi:10.1016/S0140-6736(17)32129-3

Collaborators, G. B. D. O., Afshin, A., Forouzanfar, M. H., Reitsma, M. B., Sur, P., Estep, K., . . . Murray, C. J. L. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *NEW ENGLAND JOURNAL OF MEDICINE, 377* (1), 13-27. doi:10.1056/NEJMoa1614362

Consortium, E. P. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLOS BIOLOGY, 9* (4), e1001046. doi:10.1371/journal.pbio.1001046

Cotsapas, C., Speliotes, E. K., Hatoum, I. J., Greenawalt, D. M., Dobrin, R., Lum, P. Y., . . . Consortium, G. (2009). Common body mass index-associated variants confer risk of extreme obesity. *HUMAN MOLECULAR GENETICS, 18* (18), 3502-3507. doi:10.1093/hmg/ddp292

Craige, S. M., Reif, M. M., & Kant, S. (2016). Mixed - Lineage Protein kinases (MLKs) in inflammation, metabolism, and other disease states.*Biochim Biophys Acta, 1862* (9), 1581-1586. doi:10.1016/j.bbadis.2016.05.022

Elks, C. E., den Hoed, M., Zhao, J. H., Sharp, S. J., Wareham, N. J., Loos, R. J., & Ong, K. K. (2012). Variability in the heritability of body mass index: a systematic review and meta-regression.*Frontiers in Endocrinology, 3* , 29. doi:10.3389/fendo.2012.00029

Gu, H., Huang, Z., Yin, X., Zhang, J., Gong, L., Chen, J., . . . Cui, L. (2015). Role of c-Jun N-terminal kinase in the osteogenic and adipogenic differentiation of human adipose-derived mesenchymal stem cells.*EXPERIMENTAL CELL RESEARCH, 339* (1), 112-121. doi:10.1016/j.yexcr.2015.08.005

Hirosumi, J., Tuncman, G., Chang, L., Gorgun, C. Z., Uysal, K. T., Maeda, K., . . . Hotamisligil, G. S. (2002). A central role for JNK in obesity and insulin resistance. *NATURE, 420* (6913), 333-336. doi:10.1038/nature01137

Huang, K., Nair, A. K., Muller, Y. L., Piaggi, P., Bian, L., Del Rosario, M., . . . Baier, L. J. (2014). Whole exome sequencing identifies variation in CYB5A and RNF10 associated with adiposity and type 2 diabetes. *Obesity (Silver Spring), 22* (4), 984-988. doi:10.1002/oby.20647

Jiao, H., Arner, P., Gerdhem, P., Strawbridge, R. J., Naslund, E., Thorell, A., . . . Dahlman, I. (2015). Exome sequencing followed by genotyping suggests SYPL2 as a susceptibility gene for morbid obesity.*EUROPEAN JOURNAL OF HUMAN GENETICS, 23* (9), 1216-1222. doi:10.1038/ejhg.2014.255

Jiao, H., Kulyte, A., Naslund, E., Thorell, A., Gerdhem, P., Kere, J., . . . Dahlman, I. (2016). Whole-Exome Sequencing Suggests LAMB3 as a Susceptibility Gene for Morbid Obesity. *DIABETES, 65* (10), 2980-2989. doi:10.2337/db16-0522

Kaehler, C., Isensee, J., Nonhoff, U., Terrey, M., Hucho, T., Lehrach, H., & Krobitsch, S. (2012). Ataxin-2-like is a regulator of stress granules and processing bodies. *PLoS One, 7* (11), e50134. doi:10.1371/journal.pone.0050134

Kheradpour, P., & Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments.*Nucleic Acids Res, 42* (5), 2976-2987. doi:10.1093/nar/gkt1249

Kiehl, T. R., Nechiporuk, A., Figueroa, K. P., Keating, M. T., Huynh, D. P., & Pulst, S. M. (2006). Generation and characterization of Sca2 (ataxin-2) knockout mice. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, 339* (1), 17-24. doi:10.1016/j.bbrc.2005.10.186

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods, 12* (4), 357-360. doi:10.1038/nmeth.3317

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., & Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *SCIENCE, 252* (5006), 668-674. doi:10.1126/science.1708916

Lastres-Becker, I., Brodesser, S., Lutjohann, D., Azizov, M., Buchmann, J., Hintermann, E., . . . Auburger, G. (2008). Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice. *HUMAN MOLECULAR GENETICS, 17* (10), 1465-1481. doi:10.1093/hmg/ddn035

Levi, Z., Kark, J. D., Katz, L. H., Twig, G., Derazne, E., Tzur, D., . . . Afek, A. (2017). Adolescent body mass index and risk of colon and rectal cancer in a cohort of 1.79 million Israeli men and women: A population-based study. *CANCER, 123* (20), 4022-4030. doi:10.1002/cncr.30819

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *BIOINFORMATICS, 25* (14), 1754-1760. doi:btp324 [pii]

10.1093/bioinformatics/btp324

Li, Q., & Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *AMERICAN JOURNAL OF HUMAN GENETICS, 100* (2), 267-280. doi:10.1016/j.ajhg.2017.01.004

Lima, L. O., Bruxel, E. M., Hutz, M. H., Van der Sand, C. R., Van der Sand, L. C., Ferreira, M. E., . . . Almeida, S. (2013). Influence of PPARA, RXRA, NR1I2 and NR1I3 gene polymorphisms on the lipid-lowering efficacy and safety of statin therapy. *Arquivos Brasileiros de Endocrinologia e Metabologia, 57* (7), 513-519. doi:10.1590/s0004-27302013000700003

Liu, J., Wang, L., Sun, J., Liu, G., Yan, W., Xi, B., . . . Mi, J. (2017). Bone mineral density reference standards for Chinese children aged 3-18: cross-sectional results of the 2013-2015 China Child and Adolescent Cardiovascular Health (CCACH) Study. *BMJ Open, 7* (5), e014542. doi:10.1136/bmjopen-2016-014542

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., . . . Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *NATURE, 518* (7538), 197-206. doi:10.1038/nature14177

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15* (12), 550. doi:10.1186/s13059-014-0550-8

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *GENOME RESEARCH, 20* (9), 1297-1303. doi:gr.107524.110 [pii]

10.1101/gr.107524.110

Meierhofer, D., Halbach, M., Sen, N. E., Gispert, S., & Auburger, G. (2016). Ataxin-2 (Atxn2)-Knock-Out Mice Show Branched Chain Amino Acids and Fatty Acids Pathway Alterations. *MOLECULAR &*

CELLULAR PROTEOMICS, 15 (5), 1728-1739. doi:10.1074/mcp.M115.056770

Ormazabal, V., Nair, S., Elfeky, O., Aguayo, C., Salomon, C., & Zuniga, F. A. (2018). Association between insulin resistance and the development of cardiovascular disease. *Cardiovascular Diabetology, 17* (1), 122. doi:10.1186/s12933-018-0762-4

Ozaki, K. I., Awazu, M., Tamiya, M., Iwasaki, Y., Harada, A., Kugisaki, S., . . . Kohno, M. (2016). Targeting the ERK signaling pathway as a potential treatment for insulin resistance and type 2 diabetes. *Am J Physiol Endocrinol Metab, 310* (8), E643-E651. doi:10.1152/ajpendo.00445.2015

Pal, D., Dasgupta, S., Kundu, R., Maitra, S., Das, G., Mukhopadhyay, S., . . . Bhattacharya, S. (2012). Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *NATURE MEDICINE, 18* (8), 1279-1285. doi:10.1038/nm.2851

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol, 33* (3), 290-295. doi:10.1038/nbt.3122

Sabo, A., Mishra, P., Dugan-Perez, S., Voruganti, V. S., Kent, J. W., Jr., Kalra, D., . . . Butte, N. F. (2017). Exome sequencing reveals novel genetic loci influencing obesity-related traits in Hispanic children. *Obesity (Silver Spring), 25* (7), 1270-1276. doi:10.1002/oby.21869

Schlessinger, J. (1994). SH2/SH3 signaling proteins. *CURRENT OPINION IN GENETICS & DEVELOPMENT, 4* (1), 25-30. doi:10.1016/0959-437x(94)90087-6

Seit-Nebi, A., Cheng, W., Xu, H., & Han, J. (2012). MLK4 has negative effect on TLR4 signaling. *Cellular & Molecular Immunology, 9* (1), 27-33. doi:10.1038/cmi.2011.15

Shan, X. Y., Xi, B., Cheng, H., Hou, D. Q., Wang, Y., & Mi, J. (2010). Prevalence and behavioral risk factors of overweight and obesity among children aged 2-18 in Beijing, China. *Int J Pediatr Obes, 5* (5), 383-389. doi:10.3109/17477160903572001

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Magi, R., . . . Mohlke, K. L. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *NATURE, 518* (7538), 187-196. doi:10.1038/nature14132

Simmonds, M., Llewellyn, A., Owen, C. G., & Woolacott, N. (2016). Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obesity Reviews, 17* (2), 95-107. doi:10.1111/obr.12334

Solinas, G., & Becattini, B. (2017). JNK at the crossroad of obesity, insulin resistance, and cell stress response. *Mol Metab, 6* (2), 174-184. doi:10.1016/j.molmet.2016.12.001

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., . . . Loos, R. J. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet, 42* (11), 937-948. doi:10.1038/ng.686

Stossi, F., Dandekar, R. D., Johnson, H., Lavere, P., Foulds, C. E., Mancini, M. G., & Mancini, M. A. (2019). Tributyltin chloride (TBT) induces RXRA down-regulation and lipid accumulation in human liver cells. *PLoS One, 14* (11), e0224405. doi:10.1371/journal.pone.0224405

Takahashi, A., Morita, M., Yokoyama, K., Suzuki, T., & Yamamoto, T. (2012). Tob2 inhibits peroxisome proliferator-activated receptor gamma2 expression by sequestering Smads and C/EBPalpha during adipocyte differentiation. *MOLECULAR AND CELLULAR BIOLOGY, 32* (24), 5067-5077. doi:10.1128/MCB.00610-12

Thomas, G. N., Ho, S. Y., Lam, K. S., Janus, E. D., Hedley, A. J., Lam, T. H., & Hong Kong Cardiovascular Risk Factor Prevalence Study Steering, C. (2004). Impact of obesity and body fat distribution on cardiovascular risk factors in Hong Kong Chinese. *Obes Res, 12* (11), 1805-1813. doi:10.1038/oby.2004.224

Thorleifsson, G., Walters, G. B., Gudbjartsson, D. F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., . . . Stefansson, K. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *NATURE GENETICS, 41* (1), 18-24. doi:10.1038/ng.274

Twig, G., Tirosh, A., Leiba, A., Levine, H., Ben-Ami Shor, D., Derazne, E., . . . Kark, J. D. (2016). BMI at Age 17 Years and Diabetes Mortality in Midlife: A Nationwide Cohort of 2.3 Million Adolescents.*DIABETES CARE, 39* (11), 1996-2003. doi:10.2337/dc16-1203

Twig, G., Vivante, A., Bader, T., Derazne, E., Tsur, A. M., Levi, M., . . . Kark, J. D. (2018). Body Mass Index and Kidney Disease-Related Mortality in Midlife: A Nationwide Cohort of 2.3 Million Adolescents.*Obesity (Silver Spring), 26* (4), 776-781. doi:10.1002/oby.22144

Voisin, S., Almen, M. S., Zheleznyakova, G. Y., Lundberg, L., Zarei, S., Castillo, S., . . . Schioth, H. B. (2015). Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Medicine, 7* , 103. doi:10.1186/s13073-015-0225-4

Volckmar, A. L., Song, J. Y., Jarick, I., Putter, C., Gobel, M., Horn, L., . . . Hinney, A. (2015). Fine Mapping of a GWAS-Derived Obesity Candidate Region on Chromosome 16p11.2. *PLoS One, 10* (5), e0125660. doi:10.1371/journal.pone.0125660

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols, 10* (10), 1556-1566. doi:10.1038/nprot.2015.105

Zhao, J., & Grant, S. F. (2011). Genetics of childhood obesity. *J Obes, 2011* , 845148. doi:10.1155/2011/845148

Legends

Figure 1 . **The process diagram for investigating genetic basis of obesity in Chinese children** . At discovery stage, whole-exome sequencing identified promising candidates. At validation and transcriptome stage, two coding variants that are significant associated with obesity might act through altering gene expression.

Figure 2 . **The genomic context of the genome-wide significant SNVs associated with childhood obesity in Chinese** . Both rs1059491 (A) and rs189326455 (B) reside in conserved promoter region of*SULT1A2* (NC_000016.10) and *MAP3K21* (NC_000001.11), respectively. (C) The T allele (ancestral) and G allele (derived) show different binding ability to transcription factors. Red lines indicate the locus of rs1059491. Orange blocks indicate the possible binding site for transcription factors.

Figure 3 . **The expression change of genes affected by rs1059491** . (A) Gene expression differences in blood and adipose tissues among different genotypes (TT, TG, and GG) of rs1059491. Violin plots show the expression level of obesity-related genes from GTEx database and box plots show that of genes from our transcriptome data. (B) Transcriptome data show that rs1059491 results in significant differences in obesity-related genes between cases and controls in both blood and adipose tissues.

Figure 4 . **A scheme illustrates the potential implication of rs1059491 and rs189326455 in obesity** .

Supplementary Figure 1. The prediction of secondary structure changes of SULT1A2 and MAP3K21 caused by mutant allele of rs1059491 and rs189326455, respectively .

Supplementary Figure 2 . **The allele frequency of rs1059491 in different populations across the world** . ACB, African Caribbean in Barbados. ASW, Americans of African Ancestry in SW USA. CHB, Han Chinese in Beijing, China. CHS, Southern Han Chinese. CLM, Colombians from Medellin, Columbia. FIN, Finish in Finland. GBR, British in England and Scotland. ITU, Indian Telugu from the UK. JPT, Japanese in Tokyo, Japan. KHV, Kinh in HO Chi Minh City, Vietnam. PEL, Peruvians from Lima, Peru. PUR, Puerto Ricans from Puerto Rico.

Tables

Table 1. Clinical characteristics of subjects in this study

	WES	**WES**	**WES**	**Validation**	**Validation**	**Validatio**
	Obese (n = 76)	Control (n = 74)	P	Obese (n = 2,480)	Control (n = 3,854)	P
Male/female	61/15	55/19	0.501	1,754/649	1,705/2,073	<0.001
Age (years)	12.0 ± 2.5	12.1 ± 2.9	0.700	12.4 ± 3.0	12.8 ± 3.6	<0.001
Birth weight (g)	3216.8 ± 1109.1	3167.4 ± 984.2	0.786	3235.4 ± 1046.5	3159.8 ± 950.5	0.076
Height (cm)	158.2 ± 14.6	152.2 ± 16.7	0.009	158.2 ± 15.4	152.4 ± 17.2	<0.001
Weight (kg)	73.7 ± 20.2	42.9 ± 12.9	<0.001	74.0 ± 21.7	43.5 ± 14.0	<0.001
BMI (kg/m^2)	28.8 ± 3.6	18.0 ± 2.3	<0.001	28.8 ± 4.1	18.1 ± 2.7	<0.001
FMP (%)	34.7 ± 5.9	17.8 ± 5.7	<0.001	35.1 ± 7.2	18.4 ± 6.2	<0.001
FMI (kg/m^2)	10.1 ± 2.6	3.2 ± 1.4	<0.001	10.1 ± 2.9	3.4 ± 1.6	<0.001
FFMI (kg/m^2)	15.5 ± 4.0	14.5 ± 1.5	0.027	14.2 ± 4.9	14.4 ± 1.5	<0.001
WC (cm)	84.8 ± 11.4	64.4 ± 7.3	<0.001	87.5 ± 12.4	64.0 ± 8.8	<0.001
WHtR	0.56 ± 0.05	0.42 ± 0.03	<0.001	0.57 ± 0.05	0.42 ± 0.04	<0.001
TG (mmol/L)	1.26 (0.87-1.76)	0.77 (0.57-1.01)	<0.001	1.15 (0.83-1.62)	0.75 (0.56-0.99)	<0.001
TC (mmol/L)	4.13 ± 0.54	4.11 ± 1.01	0.890	4.13 ± 0.78	4.07 ± 0.79	0.001
HDL-C (mmol/L)	1.20 ± 0.27	1.48 ± 0.34	<0.001	1.21 ± 0.25	1.49 ± 0.31	<0.001
LDL-C (mmol/L)	2.59 ± 0.48	2.40 ± 0.87	0.068	2.51 ± 0.65	2.34 ± 0.69	<0.001
FPG (mmol/L)	5.79 ± 1.04	5.45 ± 1.15	0.025	5.55 ± 0.55	5.24 ± 0.68	<0.001
SBP (mmHg)	120 ± 15	105 ± 12	<0.001	119 ± 13	105 ± 12	<0.001
DBP (mmHg)	71 ± 9	64 ± 9	<0.001	73 ± 9	65 ± 9	<0.001

P value indicates the difference for anthropometric parameters between cases and controls. Except for birth weight, differences in traits between obese and controls were compared by analysis of covariance (ANCOVA) with sex and age adjustments.

BMI: body mass index; FFMI, fat-free mass index; FMI, fat mass index; FMP, fat mass percentage; WC, waist circumference; WHtR, waist-to-height ratio; TG, triacylglycerol; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; FPG, fasting plasma glucose; SBP, systolic blood pressure; DBP, diastolic blood pressure.

Table 2. SNVs associated with quality phenotypes in WES, RNA seq and validation (n=6,334) with P < 5E-08

SNV	**Gene**	**A1/A2**	**MAF**	**MAF**	**Obesity type**	OR	**95% *CI***	P	**Data s**
			Cases	**Controls**					
rs1059491	*SULT1A2*	G/T	0.0426	0.0909	BMI	4.052	1.692-9.703	0.001686	WES
					BMI	3.533	1.29-9.675	0.01408	TS(bloo
					BMI	3.816	1.98-7.355	6.31E-05	WES+T
					BMI	2.296	1.953-2.699	7.71E-24	Sequenc
					BMI	2.405	2.058-2.811	2.57E-28	Combin
					Adiposity	1.935	1.599-2.342	1.17E-11	Combin
					Dyslipidemia	1.66	1.415-1.948	5.01E-10	Combin
rs189326455	*MAP3K21*	C/G	0.00522	0.0241	BMI	0.1034	0.0124-0.8613	0.0359	WES
					BMI	0.2187	0.1387-0.3449	6.16E-11	Sequenc
					BMI	0.2122	0.1359-0.3313	8.98E-12	Combin

SNV, single-nucleotide variant; Chr, chromosome; A1, minor allele; A2, major allele; MAF, minor allele frequency of A1 in all samples; *OR* (estimated for A1), odds ratio; CI, confidence interval; WES, whole-exome sequence; TS, Transcriptome sequencing; *SULT1A2* , NC_000016.10; *MAP3K21* , NC_000001.11. Adiposity was defined by fat mass percentage. Dyslipidemia was defined by TC, TG, LDL or HDL level abnormality.

Table 3. Estimated change unit in quantitative traits per effect allele of significant variants (*P* < 5E-8) in validation stage after adjustment

SNV	Gene	Chr	A1/A2	Trait	β	CI	P
rs189326455	MAP3K21	1	C/G	BMI	-2.655	-3.481, -1.828	3.36E-10
rs1059491	SULT1A2	16	G/T	BMI	1.953	1.519, 2.386	1.41E-18
				FMP	3.282	4.175, 7.197	7.22E-13
				FMI	1.157	0.82, 1.493	1.87E-11
				FFMI	-0.8541	-1.144, -0.5643	8.14E-09
				LDL	0.1583	0.1052, 0.2114	5.42E-09

SNV, single nucleotide variant; Chr, chromosome; A1, minor allele; A2, major allele; β , partial regression coefficient (estimated for A1); BMI, body mass index (kg/m^2); FMP, fat mass percentage; FMI, fat mass index (kg/m^2); FFMI, fat-free mass index (kg/m^2); LDL, low density lipoproteins (mmol/L), *SULT1A2* , NC_000016.10; *MAP3K21* , NC_000001.11.

Supplementary Table 1. Obesity-related SNVs in WES validated with sequenom

SNV	Gene	Chr	Pos(hg19)	Ref	Alt	SNV type	F_case	F_ctrl	OR	95% CI
rs12410676	*VAV3*	1	108185309	G	A	nonsynonymous	0.106	0.1335	0.7716	0.6847-0
rs5030752	*EPRS*	1	220156704	T	C	nonsynonymous	0.001181	0.001802	0.6882	0.313-1.
rs189326455	*MAP3K21*	1	233463936	G	C	nonsynonymous	0.005172	0.02359	0.2187	0.1387-0
chr3_20161096	*KAT2B*	3	20161096	A	G	nonsynonymous	0.1914	0.1869	1.017	0.9252-1
rs143310118	*CCDC13*	3	42784452	G	A	nonsynonymous	0.01404	0.01319	1.046	0.7421-1
rs35597368	*PDGFRA*	4	55139771	T	C	nonsynonymous	0.1591	0.1472	1.12	1.002-1.
rs148000791	*AHI1*	6	135644371	T	C	nonsynonymous	0.06151	0.05155	1.212	1.017-1.
rs1801582	*PRKN*	6	161807855	C	G	nonsynonymous	0.0582	0.05938	0.9925	0.836-1.
rs2286428	*ZP3*	7	76054372	G	A	nonsynonymous	0.08817	0.08328	1.072	0.9285-1
rs104894047	*SHH*	7	155596114	C	T	nonsynonymous	0.04407	0.03843	1.185	0.9686-1
rs2305510	*ASAP1*	8	131138344	A	C	nonsynonymous	0.1555	0.1581	0.4797	0.2426-
rs33995374	*LOXL4*	10	100020880	C	T	nonsynonymous	0.08734	0.09892	0.86	0.7486-0
rs61759818	*OVCH2*	11	7721974	C	T	nonsynonymous	0.06396	0.06156	0.9936	0.8439-1
rs17099008	*MMP20*	11	102482504	T	G	nonsynonymous	0.1044	0.1028	1.04	0.9032-1
rs25680	*CD27*	12	6554628	G	A	nonsynonymous	0.04304	0.04167	1.084	0.8891-1
rs17076657	*RXFP2*	13	32371361	A	G	nonsynonymous	0.2017	0.2137	0.9163	0.8252-1
rs3742591	*TOGARAM1*	14	45433155	C	G	nonsynonymous	0.1179	0.128	0.9022	0.7984-1
rs75560163	*PLA2G4F*	15	42434336	G	C	nonsynonymous	0.1228	0.1161	1.054	0.9307-1
rs1059491	*SULT1A2*	16	28603655	T	G	nonsynonymous	0.0859	0.04209	2.296	1.953-2.
rs7359397	*SH2B1*	16	28885659	C	T	nonaynonymoua	0.1213	0.1343	0.8898	0.7911-
rs11079339	*EPX*	17	56270442	A	G	nonsynonymous	0.09887	0.1002	1.005	0.8809-1
rs56288451	*EMILIN2*	18	2909700	C	T	nonsynonymous	0.1113	0.1318	0.8128	0.7188-0
rs12457323	*LAMA3*	18	21424986	T	C	nonsynonymous	0.0782	0.06738	1.159	0.9925-1

SNV, single nucleotide variant; Chr, chromosome; Pos, physical position; Ref, reference base at the variant

site; Alt, alternate base in the sample at the variant site; F_case, effect allele frequency in case group; F_ctrl, effect allele frequency in control group; OR, odds ratio; *SULT1A2* , NC_000016.10; *MAP3K21* , NC_000001.11.

Supplemental Table 2. Functional prediction of genome-wide significant variants

SNP	Gene	SIFT	SIFT_predict	PolyPhen_2	PolyPhen_2_predict	MutationTaster	MutationTaster_p
rs189326455	*MAP3K21*	0.33	tolerated	0.13	benign	0.994094	disease_causing
rs1059491	*SULT1A2*	0.01	deleterious	0.999	probably_damaging	0.957296	disease_causing

SULT1A2 , NC_000016.10; *MAP3K21* , NC_000001.11

Supplementary Table 3. The regulatory motifs altered by rs1059491 and rs189326455

SNV	**Transcription factor**	**Strand**	**Binding affinity ref**	**Binding affinity Alt**	**The sequences of transcript**
rs1059491					Ref: GGCGGACGGTGGTGTA
					Alt: GGCGGACGGTGGTGTA
	PPAR_2	+	3	15	YNRGGTCATDGRGGTSRNK
	RXRA_known5	-	0.9	8.1	WNKRGGKSACNKTGACCYA
rs189326455					Ref: GCGCTCTATGACTACG
					Alt: GCGCTCTATGACTACG
	CHD2_disc3	+	10.4	-1.5	SSSSSSVGVNS
	Esr2	-	11.9	0.2	NRGGDYAVNVTGACCYWB

Supplementary Table 4. rs1059491 is significant associated with gene expression level changes in adipose tissue and blood in GTEx database

Gencode Id	Gene Symbol	*P*-Value	NES	Tissue
ENSG00000197165.10	*SULT1A2*	6.70E-44	0.59	Adipose - Subcutaneous
ENSG00000259982.1	*CDC37P1*	7.90E-34	0.65	Adipose - Subcutaneous
ENSG00000251417.2	*RP11-1348G14.4*	2.00E-23	-0.53	Adipose - Subcutaneous
ENSG00000178952.10	*TUFM*	1.20E-13	0.18	Adipose - Subcutaneous
ENSG00000178188.14	*SH2B1*	6.80E-12	-0.15	Adipose - Subcutaneous
ENSG00000184110.14	*EIF3C*	2.60E-10	0.29	Adipose - Subcutaneous
ENSG00000196502.11	*SULT1A1*	1.10E-07	-0.24	Adipose - Subcutaneous
ENSG00000196993.8	*NPIPB9*	0.00012	-0.23	Adipose - Subcutaneous
ENSG00000260517.2	*RP11-426C22.5*	0.00014	0.16	Adipose - Subcutaneous
ENSG00000233232.6	*NPIPB7*	0.00017	-0.23	Adipose - Subcutaneous
ENSG00000213658.11	*LAT*	0.00019	0.1	Adipose - Subcutaneous
ENSG00000197165.10	*SULT1A2*	6.20E-90	0.86	Whole Blood
ENSG00000178952.10	*TUFM*	1.20E-64	0.44	Whole Blood
ENSG00000196502.11	*SULT1A1*	1.30E-21	-0.32	Whole Blood
ENSG00000169682.17	*SPNS1*	1.40E-20	0.18	Whole Blood
ENSG00000233232.6	*NPIPB7*	4.90E-18	-0.37	Whole Blood
ENSG00000176476.8	*SGF29*	7.00E-14	-0.19	Whole Blood
ENSG00000188603.18	*CLN3*	0.000002	-0.1	Whole Blood
ENSG00000197272.2	*IL27*	0.000026	0.11	Whole Blood
ENSG00000168488.18	*ATXN2L*	0.00006	-0.061	Whole Blood

Gencode Id	Gene Symbol	*P*-Value	NES	Tissue
ENSG00000178188.14	*SH2B1*	0.000075	-0.067	Whole Blood
ENSG00000198156.10	*NPIPB6*	0.0001	0.18	Whole Blood

NES, Normalized effect size; *SULT1A2* , NC_000016.10;*MAP3K21* , NC_000001.11.