

Fine Mapping the *MAP2K5* Region Identified rs7175517 As a Causal Variant Associated With BMI in China And The UK Populations

Ce Lu

Nanjing Medical University School of Public Health

Hai-Jun Wang

Peking University School of Public Health

Jie-Yun Song

Peking University School of Public Health

Shuo Wang

Peking University School of Public Health

Xue-Ying Li

Peking University School of Public Health

Tao Huang

Peking University School of Public Health

Hui Wang (✉ huiwang@bjmu.edu.cn)

Peking University <https://orcid.org/0000-0002-7639-5702>

Research

Keywords: BMI, eQTL, Fine mapping, GWAS, MAP2K5

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-978443/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Genome-wide association studies (GWASs) have consistently identified *MAP2K5* as the obesity susceptible gene. To deepen our understanding of the potential causal genetic variants of this region, a fine-mapping study of *MAP2K5* was conducted.

Results: First, a target single nucleotide polymorphisms (SNPs) sequencing of *MAP2K5* was carried out in 2030 Chinese children and followed by a fine mapping with the UK biobank GWAS data in 264,838 adults to identify the leading SNPs within the *MAP2K5* gene region. SNP rs7175517 (G>A) and rs4776970(T>A) were identified as the leading SNPs associated with BMI in both Chinese and the UK population. Second, colocalization of GWAS and expression quantitative trait loci (eQTL) analyses and bioinformatic functional annotations indicated that rs7175517 is the leading variants in *MAP2K5* gene region functionally. Additionally, dual luciferase assay indicated that the G allele of rs7175517 reduced the mRNA expression of *MAP2K5* in HEK293T cells. The possible mechanism was that the G allele of rs7175517 interacted with more RNA repressors from nuclei extracts, which was evidenced by electrophoretic mobility shift assays (EMSA). Furthermore, the pathway enrichment analyses of the products from DNA pull-down and protein mass spectrometry demonstrated that the G allele of rs7175517 might interact with RNA catabolic or splicing transcription factors, which consequentially increased the adiposity deposition.

Conclusions: SNP rs7175517 of *MAP2K5* gene was the putative causal variant associated with BMI. More precisely designed *in vitro* or animal experiments are warranted to further delineate the function of *MAP2K5* in the adipogenesis.

Introduction

Obesity is a serious health epidemic globally. A recent study of 195 countries estimated that 2.2 billion people were overweight or obesity in 2015[1]. The rapid rise of obesity is also a major public health problem in developing countries, including China. According to the most recent national survey, more than half Chinese adults are either overweight or obesity[2]. The estimated attribution percentage of overweight and obesity associated non-communicable diseases (NCDs) deaths increased from 5.7% in 1990 to 11.1% in 2019 in China[3]. The latest Chinese national prevalence of overweight and obesity in children and adolescents, based on the Chinese BMI screening criteria, were 6.8% and 3.6% in children under 6 years-old and 11.1% and 7.9% in children aged 6-17 years, respectively[4].

Although the increase of obesity prevalence was considered to be caused by changes in external environment such as hypercaloric diet and sedentary lifestyle, genetic factors and gene-environment interaction still play a critical role in obesity [4]. The heritability of body mass index (BMI) can reach up from 40–70%. Several genome-wide associated studies (GWASs) have revealed certain amount of susceptible genetic polymorphisms, such as fat-mass and obesity-associated gene (*FTO*) rs1421085, *SH2B* adapter protein 1(*SH2B1*) rs4788099 and mitogen-activated protein kinase kinase 5 (*MAP2K5*, the

encoded protein named MEK5) rs2241423[5, 6]. Since then, a number of researches have pinpointed that *MAP2K5* rs2241423 associated with both childhood and adulthood obesity in different Asian populations[7, 8]. In 2015, a MetaboChip meta-analysis for BMI identified 56 more novel loci and confirmed the association between of MAP2 protein complex and obesity[9]. In 2017, Abadi *et al.* selected 37 BMI-associated SNPs to observe the BMI percentiles distribution with 75,230 European ancestry participants and revealed that rs997295 of *MAPK5* had a positive association with BMI[10]. Recently, Pan *et al.* applied the promoter capture Hi-C in human adipocytes to decipher the transcription-regulation mechanism which was contributed to the adipogenesis[11]. The result revealed that *MAP2K5* rs4776984 was a cis-eQTL-eGene in the regulation of BMI. One *in vitro* experimental study demonstrated that *MAP2K5* encoded protein MEK5 protein which was the only known activator of *ERK5* and involved in the adipogenesis in the MAPK signaling pathway[12]. However, till now, no further study has been carried out to thoroughly delineate the association of variants within *MAP2K5* gene region with obesity.

Additionally, like other complex diseases, most of the *MAP2K5* variants reported by GWAS studies lie within non-coding regions, probably due to linkage disequilibrium (LD) which makes the causal variants inference and consequentially functional evaluation complicated[13]. Fine mapping is a complementary method for GWAS which can further elucidate the risk region/gene by investigating as many as possible variants either by imputation or sequencing for more detailed analysis and be combined with functional annotation to illustrate the biological mechanism of variants[14]. Therefore, the present study focused on the *MAP2K5* gene region and utilized Chinese children and UK Biobank (UKB) population to explore the causal variants among the *MAP2K5* gene region associated with body mass index (BMI). Furthermore, *in silico* functional annotation, bioinformatic colocalization analyses and *in vitro* experiments were performed to reveal possible molecular mechanisms.

Results

Associations between candidate variants and BMI in Chinese Children

First, to explore which genetic variants potentially have causal relationship with overweight/ obesity, several genetic variants of *MAP2K5* were validated in Chinese children. Based on the sample size and estimated statistical power ($\beta \leq 0.25$), six variants (MAF>0.10 in Chinese population) were selected for the validation in Chinese children. The basic characteristics of six variants was listed in the **Table S1**. Four variants rs16951006, rs3784711, rs7175517 and rs4776970 were significantly associated with BMI even after Bonferroni correction, as shown in Table 1 (all $P < 0.0083$, Bonferroni corrected for 6 SNPs). Additionally, the associations among the 6 variants with overweight/obesity were conducted as well (**Table S2**). All the aforementioned variants remained statistical significance even after Bonferroni correction (all $P < 0.0083$), except for rs16951006. SNP rs7175517 had the lowest P value among six SNPs in BMI association analyses and similar results were detected with central obesity phenotypes in Chinese children (**Table S3**) and the UK population (**Table S4**).

Table 1
Association of *MAP2K5* genetic variants and BMI in Chinese children

Gene	SNPs	BMI (N=2030)		
		β^*	SE	<i>P</i>
<i>MAP2K5</i>	rs11636408	0.25	0.13	0.051
<i>MAP2K5</i>	rs16951006	0.46	0.15	0.002
<i>MAP2K5</i>	rs8037318	-0.06	0.15	0.676
<i>MAP2K5</i>	rs3784711	0.37	0.14	0.007
<i>MAP2K5</i>	rs7175517	0.48	0.13	1.43×10⁻⁴
<i>MAP2K5</i>	rs4776970	0.48	0.15	0.001
SE: standard error				
*Effect sizes and <i>P</i> -values were estimated under additive genetic model adjusted for age, sex and study group.				

Associations between *MAP2K5* genetic variants and BMI in the UK population

Second, forward stepwise linear regressions were conducted between all selected genetic variants of *MAP2K5* and BMI. In total, 297 SNPs were significantly associated with BMI (all $P < 5 \times 10^{-8}$). To be noted, rs4776970 (T>A, $\beta = 0.024$, 2.33×10^{-23}) and rs7175517 (G>A, $\beta = 0.026$, 1.01×10^{-22}) were significantly associated with BMI, individually (Figure 1A). To be noticed, the effect size of rs7175517 was much higher in Chinese children ($\beta = 0.48$), which is almost 18 times as large as the value in UK population ($\beta = 0.026$). After conditioning on either rs4776970 or rs7175517, no variant in the *MAP2K5* region reached the predefined significance threshold [$P < 5 \times 10^{-8}$, Figure 1B (conditioning on rs7175517) and **Figure S1B** (conditioning on rs4776970)].

Colocalization of the expression quantitative trait locus (eQTL) and GWAS associations for genetics variances of *MAP2K5*

To gain insight into the leading genetic variants in *MAP2K5* related with BMI, colocalization of Genotype-Tissue Expression (GTEx) and GWAS data were conducted in subcutaneous adipose and whole blood eQTL, individually (Figure 2). SNPs in strong LD with rs7175517 have more favorable associations with BMI both in subcutaneous adipose and whole blood tissues, individually (Figure 2A & 2C). The variants located within 1 Mb of rs7175517 increased the expression of *MAP2K5* and BMI in subcutaneous adipose and whole blood tissues, individually (Figure 2B & 2D). No such prominent result was detected for rs4776970.

In-silico functional annotation and experimental verifications

The histone modification results indicated that rs7175517 marked with peaks of H3K4me3, H3K4me1 and H3K27ac in the ENCODE analyses from UCSC website. We assumed that SNP rs7175517 was a strong functional variant falling within promoter or enhancer region in comparison with rs4776970. (Figure S2). Therefore, only rs7175517 was chosen for the functional experiments. We first conducted dual luciferase reporter assays to determine how rs7175517 affect the mRNA expression of *MAP2K5*. Results showed that the construct containing rs7175517 [A] allele exhibited higher enhancer activity than that containing rs7175517 [G] allele (Figure 3A). Consequentially, electrophoretic mobility shift assays (EMSA) results indicated that rs7175517 [G] allele preferentially banded to more nuclear extracts compared with rs7175517 [A] allele in HEK293T cells (Figure 3B and Figure 3C). To pinpoint what kinds of nuclear proteins might bind to the G allele of rs7175517, DNA pull-down and protein mass spectrometry experiments were carried out. The protein mass spectrometry results were further analyzed with the Go, KEGG and RACE pathways with pathways enrichment method. Results implied that G allele of rs7175517 may recruited more RNA spliceosomes (Figure 4).

Discussion

Obesity is a highly complex trait caused by the reciprocal genetic and environmental factors. Understanding which variant statistically significantly contribute to adipogenesis at molecular level has been evidenced to be difficult. Deciphering the biological mechanisms of those significant signals in population is the vital step to clarify the molecular mechanism of obesity and can suggest better prevention strategy. As the activated kinase of ERK signaling pathway, *MAP2K5* was not only involved in the pathogenesis of adipose tissue but also played a critical role in protecting cells from stress-induced apoptosis, neuronal survival, cardiac development and angiogenesis[15]. Fine-mapping of the *MAP2K5* gene region could provide us the panorama picture on the associations between full genetic variants located in *MAP2K5* and BMI. In the present study, fine-mapping of *MAP2K5* was conducted in 2,030 Chinese children and further validated in 264,838 UK population. Furthermore, in-silico functional annotation and *in vitro* experiments were consequentially carried out. All results revealed that rs7175517 of *MAP2K5* was functionally correlated with obesity in both Chinese and UK population. The possible molecular mechanism was the G allele of rs7175517 bound with more spliceosomes, sequentially inhibited the expression of MEK5 and then triggered more adipogenesis.

The SNP rs2241423 has very strong LD with rs7175517 ($r^2=0.99$) and preferentially selected in most obesity related GWAS or candidate genetic variants studies. However, no signal for rs2241423 was detected in in-silico functional analyses in present study. Therefore, no further experiments were conducted for it. It is worthy to mention that previous research indicated that rs2241423 had stronger effect size on obesity in Chinese children than in Caucasian population[16]. Similar trends were observed in present study as well, SNP rs7175517 had stronger effect size on BMI in Chinese children than in the UK population, which further evidenced that *MAP2K5* had trans-ethnic difference in adipogenesis.

Taking results from our dual luciferase assay and EMAS assay together, the G allele of rs7175517 reduced the expression of *MAP2K5* and bound to more nuclear proteins. Intriguingly, our findings were

akin to the research of Pan *et.al.*, who identified that C allele of rs4776984 (high LD with rs7175517, $r^2=0.97$) increased nuclear protein binding in an allele-specific way via EMSA experiments as well[11]. Consistently, both EMSAs results indicated that nuclear proteins might interact with this region and suppress the expression of *MAP2K5*, which further triggered the adipogenesis. The recent research by Joslin *et.al.* revealed that rs4776984 (high LD with rs7175517, $r^2=0.97$) was associated with obesity via enhancer modulating variants analyses[17]. All that information indicated that this region is highly related with BMI due to high LD among those SNPs.

Our supershift assay did not find any significant transcription factor bound to rs7175517 region. Similar with the research conducted by Pan *et. al.*, positive transcription factors in the regulation of adipogenesis, such as CCAAT/enhancer binding protein beta (*CEBPB*) and peroxisome proliferator-activated receptor gamma (*PPARG*) were predicted to interact with rs4776984 region in adipocytes[18]. However, none of them was evidenced by the supershift assay[11]. This phenomenon might occur since there is a complex of transcription factors bind to rs7175517 rather than a solo transcription factor. In present study, whole blood tissue also indicated expression of *MAP2K5* related with BMI. Since *MEK5* was not only expressed in the adipocytes, it expressed ubiquitously in all type cells. Therefore, it may play an important role in the adipogenesis in other type of cells. For example, *MEK5* involved in the adipogenesis was evidenced in the mouse embryonic fibroblasts as well via *MEK5-ERK5* signaling pathway[12]. Additionally, Cristea's research revealed that *MEK5-ERK5* participated in the lipid metabolism in small-cell lung cancer through cholesterol synthesis pathway via regulating sterol regulatory element binding protein (*SREBP*)[19]. All these indicated that *MEK5-ERK5* may involve in the metabolic and adipogenesis pathway around whole body.

In addition, our results preferentially indicated that rs7175517 might bind to certain transcriptional factors which could regulate RNA splicing or expression with pathway enrichment analyses. The GO, KEGG and REAC pathway enrichment analyses all pinpoint that rs7175517 was interacted with RNA spliceosomes. As the G allele of rs7175517 reduced the expression of *MEK5* in the dual luciferase assay and bound with more nuclear proteins in EMSA, we speculated that certain RNA regulating transcription factors binding to the allele-specific region of rs7175517 and modulating the expression of *MEK5*. Consequentially, the *MEK5-ERK5* signaling pathway was down regulated and triggered adiposity accumulation. In fact, *MPA2K5* encoded protein has two isoforms, one is named *MEK5 α* (50-kDa) and another is named *MEK5 β* (40-kDa). *MEK5 α* is mainly expressed in the liver and brain, and is particulate, while *MEK5 β* is ubiquitously expressed and primarily cytosolic[20]. *MEK5 β* lacks an extended N terminus which was presented in *MEK5 α* . The N terminus of *MEK5 α* is the docking site for *ERK5* and *MEK5 α* is a stronger activator for the *ERK5* in comparison with *MEK5 β* [21]. However, the specific function of *MEK5 β* is not clear yet. Therefore, we presumed that certain RNA splicing regulator adjusted the expression of *MEK5 α* and *MEK5 β* to coordinate the activation of *ERK* signaling pathway in adipogenesis.

In summary, we first implemented fine mapping method to gain comprehensive view on the associations between *MAP2K5* gene region and BMI. The putative SNPs were further evidenced by in vitro experiments. Finally, we identified rs7175517 is the leading variant associated with obesity. However, the

present study only included Chinese children with limited sample size, so further validation studies with larger sample size and various places from China are warranted.

Overall, the present study deepened our understanding of the adipogenesis of the *MAP2K5* through whole genetic region and provided possible target for future obesity intervention or therapy. However, to gain insight of different isoforms' function in adipogenesis, more precisely designed molecular experiments should be conducted. In particular, *in vitro* experiments about interaction between the genetic variants (high LD variants) should be carried out to address which SNPs interact with what kind of transcriptional factors and how the MEK5-ERK5 signaling pathway was modulated in different tissues of whole body.

Conclusions

We fine mapped the *MAP2K5* region and identified SNP rs7175517 of *MAP2K5* gene was the putative causal variant associated with BMI. The results deepened our understanding of the adipogenesis of the *MAP2K5* through whole genetic region and provided possible target for future obesity intervention or therapy.

Materials And Methods

Study population

A two-stage case-control study was conducted. In the discovery stage participants were recruited from two independent case-control studies in the urban areas of Beijing, China. First, the study of Adolescent Lipids, Insulin Resistance and Candidate Genes (ALIR) included 151 normal-weight, 400 overweight and 386 obese children aged 7-18 years old. Second, the baseline of Comprehensive Prevention Project for Overweight and Obese Adolescents (CPOOA) collected 456 normal-weight, 318 overweight and 319 obese children aged 14-17[22, 23]. Individuals with any cardiovascular or metabolic related diseases were excluded as well. BMI is calculated by dividing weight (Kg) by square of height (m²). According to the BMI percentile criteria, children with an age- and sex-specific BMI \geq 95th percentile were classified into obese group, and those with a BMI between 85th to 95th percentile were classified into overweight group, whereas those with a BMI between 15th to 85th percentile were normal-weight. The children with BMI \geq 97th percentile were defined as severely obese [24]. Both studies were approved by the Ethics Committee Board of Peking University Health Science Center.

In the replication stage, information was extracted from UKB database. The UKB database is a cohort of approximately half a million individuals aged 40 to 69 years across the United Kingdom. The UKB data are available on application to the UKB (www.ukbiobank.ac.uk/). This research was conducted using the UKB data under Application Number 44430. A BMI between 18.5-25 is classified as normal, 25-30 as overweight, 30-35 as obesity and more than 35 as severely obesity[25]. The study only included Caucasian people and proposed individuals with no blood relationship. Applying the 1st, 2nd and 3rd Principal Component (PC) filtering and selecting those with information of age, sex, BMI and Townsend

deprivation index individuals. Finally, 264,838 non-related individuals of self-reported British descent from the UK Biobank were included in the present study. The summary-level of GWAS data used in present study are publicly available. Therefore, no specific ethical approval is needed.

Genotyping quality control and imputation

In the discovery stage, 9 variants located in *MAP2K5* genes, with 2 variants from published literatures[5, 26] and 7 Tag SNPs base on CHB database from 1000 Genomes Project were selected for genotyping. Genotyping was performed with the matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS, Agena, San Diego, CA, USA). The call rates were above 99.4% and 3 variants were excluded for strong linkage disequilibrium ($r^2 > 0.80$) in the present study. The remained 6 variants were shown for the basic information of genotyping (**Table S1**).

In the replication stage, the genotype data of SNPs located in the *MAP2K5* region were extracted from UKB GWAS database. The SNPs selection was conducted under the following criteria: (i) imputation quality score (INFO) ≥ 0.9 ; (ii) genotyping call rate $\geq 95\%$; (iii) minor allele frequency (MAF) in controls ≥ 0.01 ; or (iv) Hardy-Weinberg equilibrium (HWE) $\geq 1 \times 10^{-6}$. Finally, 2,994 SNPs were included into the subsequent analyses.

eQTL analyses and in-silico functional annotations

eQTL analyses was performed with Genotype-Tissue Expression (GTEx; <http://www.gtexportal.org/home/>) Project expression quantitative trait locus (eQTL) database (V8 release) [27]. Variant-gene paired eQTL analysis results were conducted in the subcutaneous adipose and whole blood tissues, individually.

To explore the potential molecular functions of the colocalization genes and corresponding variants, we performed in-silico functional annotations with several prediction aspects, including histone modification sites (H3K4me1, H3K4me3 and H3K27ac) from the Encyclopedia of DNA Elements (ENCODE). All outcomes were visualized in UCSC browser[28]. Transcription factor binding site analyses were performed with the JASPAR 2020[29] and enhancer element locator algorithm[30]. Finally, pathway enrichment analyses were adopted to explore which signaling pathway did the transcription factor involve via GO, KEGG and REAC pathways[31].

Dual luciferase reporter assays (DLRA)

Luciferase constructs were generated to encompass corresponding variants. The rs7175517 (chr15:68077130-68078130 GRCh37) is located in the intro region of *MAP2K5* and was cloned into the pGL3-Promoter vector (Promega, Madison, WI, USA). The constructed plasmids were sequenced to confirm the accuracy. HEK293T cells were plated into 24-well plates into each well (7.5×10^4) and co-transfected the plasmids (100ng/well) of interest next day with pRL-SV40 Renilla Luciferase Control Vector (10ng/well, Promega, Madison, WI, USA) using Lipofectamine 2000 reagent (Invitrogen, Carlsbad,

CA, USA). After 48 hours of culturing, the cells were lysed, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA). Relative luminescent signals were calculated by normalizing luciferase signals with Renilla signals. In total, 3 independent transfection experiments with triplicates for each condition were conducted.

Electrophoretic mobility shift assay (EMSA)

Nuclear extracts were prepared from HEK293T using the NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Scientific, Waltham, MA, USA). DNA oligonucleotides for each variant were synthesized with 5'-biotin labeling and HPLC purified by Genscript (Nanjing, China; probe sequences are listed in **Table S5**). Double-stranded DNA probes were prepared by combining sense and antisense oligonucleotides, heat annealing, and slow cooling. Probes and HEK293T cell nuclear extracts were then incubated by using the LightShift EMSA Optimization & Control Kit (Thermo Scientific) at 4°C for 20 mins. For competition assays, unlabeled competitors at 25-fold, 50-fold or 100fold excess oligonucleotides were added to the reaction mixture 10 mins before the addition of labeled probes. After incubation, binding reactions were separated on a 6% polyacrylamide gel and transferred blots were developed using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific) and signals were visualized with the ChemiDoc XRS+ scanner (BIO-RAD, Louisville, KY, USA).

DNA Pull-down and Protein mass spectrometry

The biotin labeled probe and magnetic beads were placed on the 4 °C freezer turn over incubation for 6 ~ 8 hours, the nuclear extracts were incubated with the magnetic beads DNA probe complexes placed on the 4 °C freezer turn over incubation overnight, after washing to remove nonspecific binding proteins. Finally, the eluate was subjected to elution to obtain the product of interest, which was then subjected to protein mass spectrometry to identify the protein. Protein mass spectrometry was conducted in the central lab of Nanjing Medical University.

Statistical analyses

Hardy-Weinberg equilibrium of all the genotypes were analyzed with χ^2 test. Logistic regression and linear regression analyses were conducted to analyze the effect of genetic variants on overweight and obesity (categorical variables) or BMI, individually. The adjusted covariables were age, sex and study population. For the UKB data analysis, age, sex, income, educational attainment (income and educational attainment was replaced by Townsend deprivation index for the forward step wise regression with 2,297 SNPs) and the top three principal components were adjusted for logistic and linear regressions. All genetic regression analyses were performed under an additive model and conducted with Plink software (v.1.9). The linkage disequilibrium (LD) between variants was tested by calculating r^2 with Haploview 4.2. In terms of genetic associations analyses, statistical significance were considered when P values $<5 \times 10^{-8}$. For the in vitro experiments, an unpaired student t test was used to compare the mean value of different conditional triplicates and two-sided P values <0.05 were considered significant unless otherwise specified. Then pathway enrichment analyses were performed for the proteins identified by mass

spectrometry using 'clusterProfiler' package in R software (version 3.5.1)
(<http://bioconductor.org/packages/clusterProfiler/>, version 3.6.1)

Abbreviations

BMI: body mass index;

eQTL: expression quantitative trait locus;

GWAS: genome-wide association study;

MAP2K5: mitogen-activated protein kinase kinase 5;

MEK: MAPK/extracellular signal-regulated kinase (ERK) kinase;

UKB: United Kingdom biobank;

SNP: single nucleotide polymorphism.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 23/09/2020.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by discipline construction funding of public health and preventive medicine from Peking University Health Science Center (BMU2020XY010 and BMU2021YJ046). The Chinese children SNP analysis was sponsored by research grants from National Natural Science Foundation of China (8157120468).

Authors' Contributions

Hai-Jun Wang, Tao Huang and Hui Wang conceived and designed the concept, Ce Lu and Tao Huang analysed the UKB data. Shuo Wang and Jie-Yun Song collected blood samples and analysed the data of Chinese children; Ce Lu and Xue-Ying Li completed all the *in vitro* experiments under the guide of Hui Wang. Cu Lu and Hai-Jun Wang did the functional annotation. Ce Lu and Hui Wang wrote the draft of the manuscript. All authors read, improved, and approved the manuscript.

The funder/sponsor did not participate in the work.

Acknowledgements

We thank individuals who participated in our study and those who participated in the UKB and GTEx studies. We also thank Prof. Zhibin Hu and Prof. Guangfu Jin provided us the experimental area and equipments to support us complete all the experiments.

References

1. Collaborators GBD, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med*. 2017;377(1):13–27.
2. The State Council Information Office of the People's Republic of China. Press briefing for the Report on Chinese Residents' Chronic Diseases and Nutrition 2020. Beijing; 2020.
3. Institute for Health Metrics and Evaluation *Global Health Data Exchange*. 2021 [cited 2021 Jan 10]; Available from: <http://ghdx.healthdata.org/gbd-resultstool>.
4. Pan XF, Wang L, Pan A. Epidemiology and determinants of obesity in China. *Lancet Diabetes Endocrinol*. 2021;9(6):373–92.
5. Speliotes EK, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42(11):937–48.
6. Thorleifsson G, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*. 2009;41(1):18–24.
7. Dorajoo R, et al. Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int J Obes (Lond)*. 2012;36(1):159–63.
8. Lee JS, Cheong HS, Shin HD. BMI prediction within a Korean population. *PeerJ*. 2017;5:e3510.
9. Locke AE, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197–206.
10. Abadi A, et al. Penetrance of Polygenic Obesity Susceptibility Loci across the Body Mass Index Distribution. *Am J Hum Genet*. 2017;101(6):925–38.
11. Pan DZ, et al. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat Commun*. 2018;9(1):1512.

12. Zhu H, et al. Role of extracellular signal-regulated kinase 5 in adipocyte signaling. *J Biol Chem.* 2014;289(9):6311–22.
13. Hutchinson A, Asimit J, Wallace C. Fine-mapping genetic associations. *Hum Mol Genet.* 2020;29(R1):R81–8.
14. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19(8):491–504.
15. Paudel R, Fusi L, Schmidt M. *The MEK5/ERK5 Pathway in Health and Disease.* *Int J Mol Sci,* 2021. 22(14).
16. Wang HJ, et al. Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case-control study. *BMC Med Genet.* 2016;17:7.
17. Joslin AC, et al. A functional genomics pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci. *Nat Commun.* 2021;12(1):5253.
18. Mosesti D, Regassa A, Kim WK. *Molecular Regulation of Adipogenesis and Potential Anti-Adipogenic Bioactive Molecules.* *Int J Mol Sci,* 2016. 17(1).
19. Cristea S, et al. The MEK5-ERK5 Kinase Axis Controls Lipid Metabolism in Small-Cell Lung Cancer. *Cancer Res.* 2020;80(6):1293–303.
20. English JM, et al. Isolation of MEK5 and differential expression of alternatively spliced forms. *J Biol Chem.* 1995;270(48):28897–902.
21. Seyfried J, et al. A novel mitogen-activated protein kinase docking site in the N terminus of MEK5 α organizes the components of the extracellular signal-regulated kinase 5 signaling pathway. *Mol Cell Biol.* 2005;25(22):9820–8.
22. Wang D, et al. Association of the MC4R V103I polymorphism with obesity: a Chinese case-control study and meta-analysis in 55,195 individuals. *Obesity (Silver Spring).* 2010;18(3):573–9.
23. Wang HJ, et al. Association of the common genetic variant upstream of INSIG2 gene with obesity related phenotypes in Chinese children and adolescents. *Biomed Environ Sci.* 2008;21(6):528–36.
24. Ji CY. and C. Working Group on Obesity in, *Report on childhood obesity in China (1)–body mass index reference for screening overweight and obesity in Chinese school-age children.* *Biomed Environ Sci.* 2005;18(6):390–400.
25. team WGI. The SuRF Report 2. Surveillance of chronic disease Risk Factors: Country-level data and comparable estimates. Geneva: World Health Organization; 2005.
26. Wen W, et al. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet.* 2012;44(3):307–11.
27. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
28. Haeussler M, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D853–8.
29. Fornes O, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2020;48(D1):D87–92.

30. Hallikas O, et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*. 2006;124(1):47–59.
31. Reimand J, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482–517.

Figures

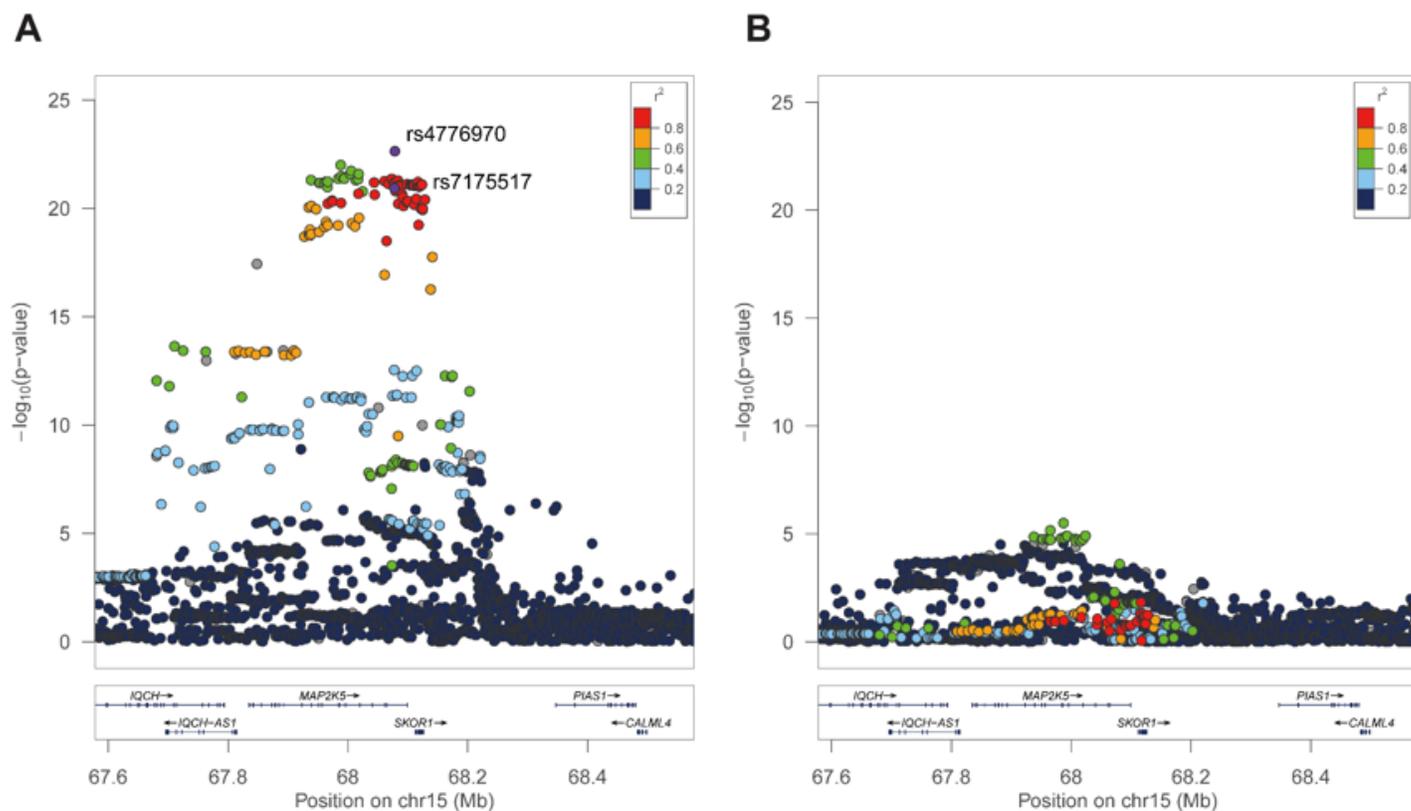


Figure 1

Regional association plots of MAP2K5 loci independently associated with BMI in UKB. The Y-axis represents P value in $-\log_{10}$ scale and the X-axis indicated the genetic variants localization. The r^2 was calculated based on rs7175517. (A) The strongest association was SNP rs4776970 and rs7175517 via forward stepwise regression analysis. (B) After condition for SNP rs7175517 and no other significant variant were pinpointed. The extent of linkage disequilibrium for all SNPs with rs7175517 is indicated by purple colors.

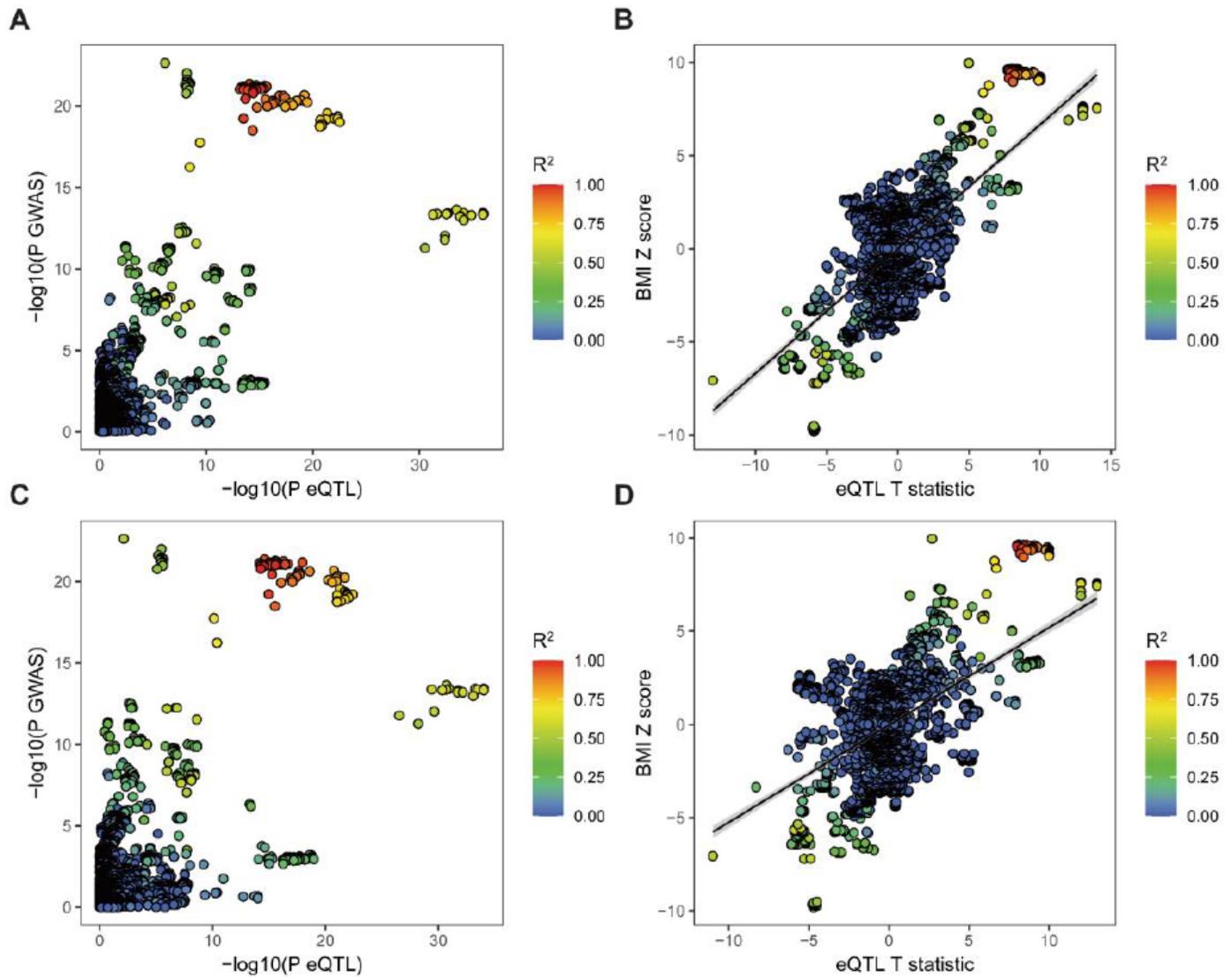


Figure 2

Colocalization of the eQTL and GWAS associations for MAP2K5 in UKB Scatterplot shows the overlap of GWAS and eQTL associations for genetic variants of MAP2K5 (Figure A and C). The Y-axis represents GWAS P value in $-\log_{10}$ scale for BMI. The X-axis represents eQTL P value in $-\log_{10}$ scale in subcutaneous adipose (A) and whole blood (C). Additionally, scatterplot shows SNP associations between BMI and MAP2K5 expression in subcutaneous adipose (B) and whole blood (D). The y-axis represents variants associated with BMI risk (Z score). The x-axis shows eQTLs associated with MAP2K5 gene expression (t statistic). The extent of linkage disequilibrium for all SNPs with rs7175517 is indicated by red colors.

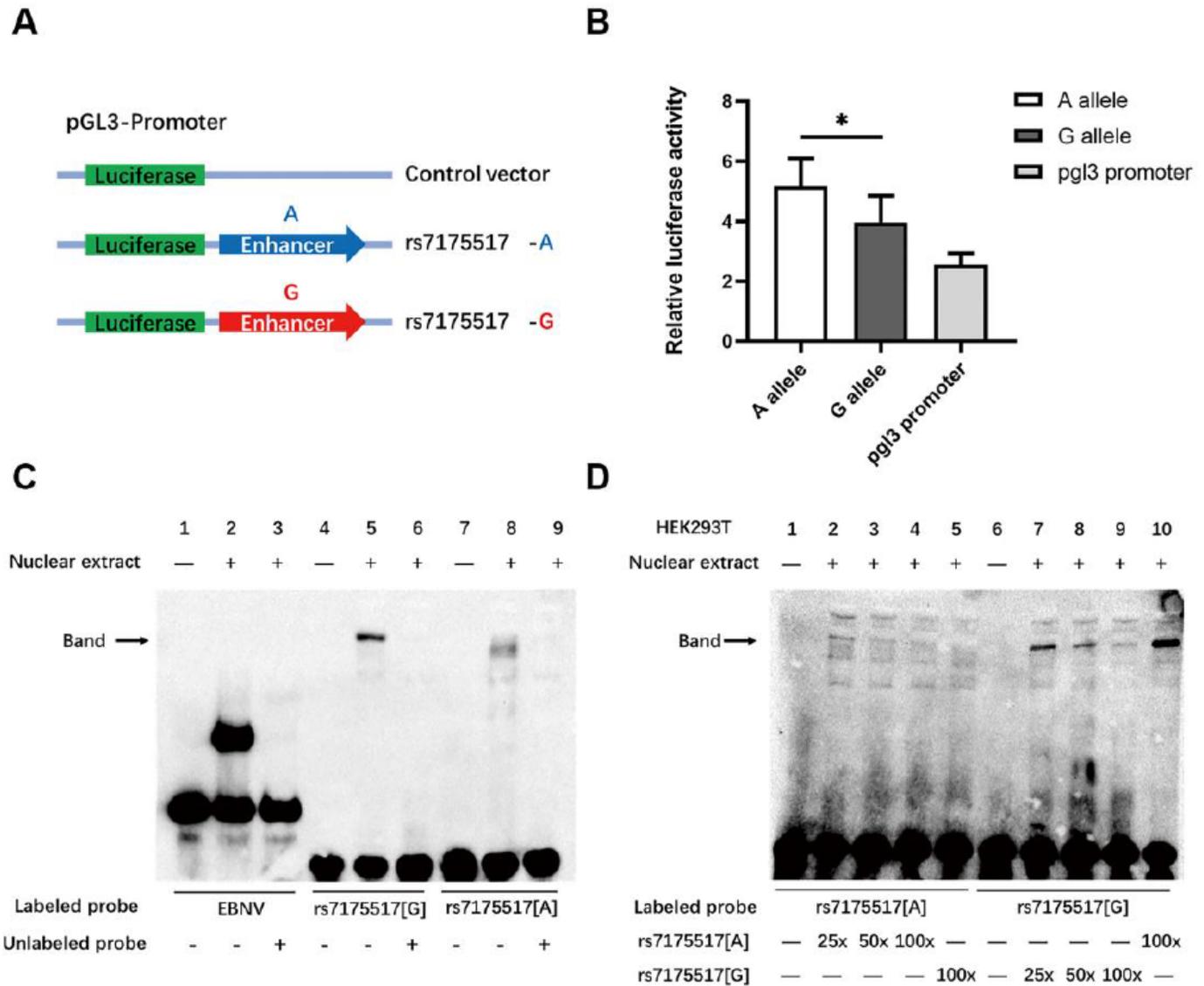
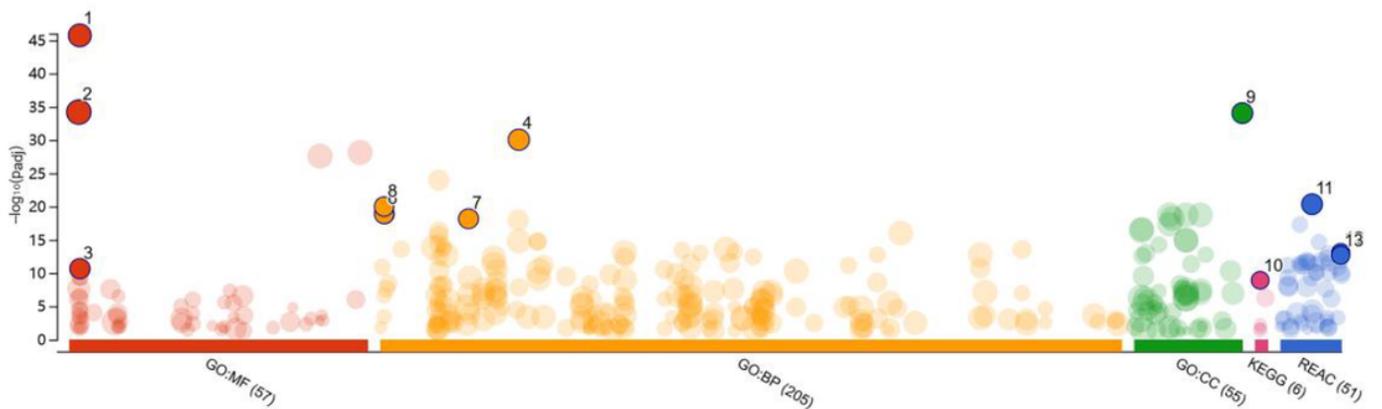


Figure 3

In vitro experiments of rs7175517 Luciferase reporter assay for region encompassing rs7175517(A). Each allele-specific construct contained the upstream 500bp and downstream 500bp of the putative mutated site. All constructs were introduced into the pGL3-promoter luciferase reporter vectors. All constructs as indicated (100ng per well into 24 well plates) were transiently transfected into human embryo kidney293T (HEK293T) cells with triplicates for 48h(B). EMSA with biotin-labeled oligonucleotides contained the rs7175517 G allele or rs7175517 A allele and nuclear was extracted from HEK293T cells (C). Lanes 1,4 and 7 showed the mobility of the corresponding labeled oligonucleotides without nuclear extracts; Lanes 2, 5 and 8 showed the mobility of the corresponding labeled oligonucleotides with nuclear extracts in the absence of competitor; lanes 3, 6 and 9 showed the mobility of the labeled oligonucleotides with nuclear extracts in the presence of unlabeled competitor(C). For example, more unlabeled rs7175517 oligonucleotides were added into the premixture of biotin labeled rs7175517 to compete for the interaction with nuclear extracts. The arrow indicates a DNA-protein

complex for rs7175517. EMSA competition assays(D). Lanes 1-5 indicates the competition assay for rs7175517 A allele and all lanes added with 80 fmol biotin-labeled oligonucleotides containing rs7175517 A allele. Line 1 refers to those without nuclear extracts; line 2, 3 and 4 refer to those with nuclear extracts and are in the presence of 25x, 50x and 100x unlabeled oligonucleotides containing rs7175517 A allele; line 5 refers to those with nuclear extracts and is in the presence of 100x unlabeled oligonucleotides containing rs7175517 G allele. With the increasing amount of unlabeled oligonucleotides, the band should slowly get shallow, particularly for adding stronger interaction allele. Similar results were presented in the line 6-10. For line 10, since rs7175517 G allele had stronger interaction with nuclear extracts, even 100x rs7175517 A allele was added, no fade was observed in the band. All the experiments were repeated three times, and only one experimental result was represented for each kind of experiment. For the dual luciferase assay, student t test was adopted to test the mean differences between different conditions and P value <0.05 was consider as statistically significant.



ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0003723	RNA binding	1.800×10^{-46}
2	GO:MF	GO:0003676	nucleic acid binding	6.197×10^{-35}
3	GO:MF	GO:0003729	mRNA binding	2.156×10^{-11}
4	GO:BP	GO:0016071	mRNA metabolic process	8.788×10^{-31}
5	GO:BP	GO:0000398	mRNA splicing, via spliceosome	1.291×10^{-19}
6	GO:BP	GO:0000377	RNA splicing, via transesterification reactions with ...	1.291×10^{-19}
7	GO:BP	GO:0008380	RNA splicing	6.728×10^{-19}
8	GO:BP	GO:0000375	RNA splicing, via transesterification reactions	9.950×10^{-21}
9	GO:CC	GO:1990904	ribonucleoprotein complex	8.794×10^{-35}
10	KEGG	KEGG:03040	Spliceosome	1.193×10^{-9}
11	REAC	REAC:R-HSA-89...	Metabolism of RNA	4.397×10^{-21}
12	REAC	REAC:R-HSA-72...	mRNA Splicing - Major Pathway	8.316×10^{-14}
13	REAC	REAC:R-HSA-72...	mRNA Splicing	1.984×10^{-13}

Figure 4

Pathway enrichment analyses with GO, KEGG and REAC. Core protein was highly expressed in the DNA pull-down assay in three databases (GO, KEGG and REAC). The GO pathways enrichment analyses including molecular function (MF), biological process (BP) and cellular component (CC) were indicated by red, orange and green. KEGG pathway enrichment analysis was indicated by pink color and the REAC pathway enrichment analysis was indicated by blue. All results indicated that co-expressed proteins were enriched in RNA splicing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfiles025.docx](#)