Whole Exome Sequencing Study in a Family with Type 2 Diabetes Mellitus

Xiaowei Zhou¹,⁎
Weichang Guo²,⁎
Hejia Yin¹
Jie Chen¹
Liju Ma³
Qiuping Yang⁴
Yan Zhao¹
Shaoyou Li⁵
Weijun Liu²
Huifang Li¹

¹Department of Diabetes, The First Affiliated Hospital of Kunming Medical University, Kunming, People’s Republic of China; ²Department of Physical Education, Kunming Medical University, Kunming, People’s Republic of China; ³Department of Clinical Laboratory, First Affiliated Hospital of Kunming Medical University, Kunming, People’s Republic of China; ⁴Department of Geriatrics, The First Affiliated Hospital of Kunming Medical University, Kunming, People’s Republic of China; ⁵Department of NHC Key Laboratory of Drug Addiction Medicine, The First Affiliated Hospital of Kunming Medical University, Kunming, People’s Republic of China

⁎These authors contributed equally to this work

Background: Type 2 diabetes mellitus (T2DM) is characterized by β cell decline in the pancreas and insulin resistance. This study aimed to investigate the possible pathogenetic gene mutation sites of T2DM patients using whole exome sequencing.

Materials and Methods: We recruited a Chinese family with 3-generation history of diabetes. The whole blood genomic DNA of seven members of the family was extracted and sent for whole exome sequencing. Biological information was analyzed with in silico prediction methods, including significance analysis of single nucleotide polymorphism (SNP)/Indel site, and analysis of specific SNP/Indel proteins and their potential mechanisms.

Results: Six out of seven members of the family were diagnosed with diabetes. All DNA samples (23 kb) met quality requirements of library construction. Clean reads of each sample demonstrated high Q20 and Q30 (>80%), indicating good sequencing quality of sequencing data. A total of 130,693 SNPs and 15,928 Indels were found in DNA samples. A total of 22 significant SNPs and Indel mutation sites located on 19 genes were obtained, including ZCCHC3, SYN2, RPL14, SRRD, AMID, CAMKK2, ZNF787, RNF157, NPIPB15, AIG3, KIAA0040, MAST2, ESRRB, Chor58, PNLIPRP1, DACH1, MACCI, CAPN9 and DMKN. An rs2305205 mutation of PNLIPRP1 gene and an rs778701848 mutation of CAMKK2 gene may be associated with the pathogenesis of T2DM in this family.

Conclusion: Exons of these diabetic patients demonstrated an rs2305205 mutation in PNLIPRP1 gene and an rs778701848 mutation in CAMKK2 gene. These two mutations might promote T2DM occurrence through reducing sensitivity of peripheral tissue to insulin and reducing insulin secretion.

Keywords: type 2 diabetes, gene mutation, whole exome sequencing, PNLIPRP1, CAMKK2

Introduction
Type 2 diabetes mellitus (T2DM), as a polygenic genetic disease, is mainly characterized by β cell decline in the pancreas and insulin resistance, usually resulting from a combination of genetic and environmental factors. The main clinical manifestations of T2DM are glycolipid metabolism associated disorders. In recent years, with aging of the population and improvement of living standards, the incidence of diabetes is rising year by year worldwide. According to the data of International Diabetes Federation in 2017, there are approximately 451 million diabetes mellitus patients in the whole world, with the potential to achieve 693 million by the year 2045.¹ Among these diabetes mellitus patients, T2DM accounts for more than 90%. A previous study² reported that the occurrence of T2DM is related to genetic susceptibility, with a certain degree of familial aggregation.² In Arab, Tunisia, France, Sweden, Greece, and China, more than...
50% of diabetic patients demonstrated a positive family history of diabetes mellitus. It is suggested that family history, genetic factors, and similar environmental factors are associated with an increased risk of T2DM. In a similar living environment, the prevalence of siblings is 4.2 times higher than that of spouses, indicating that genetic factors play a critical role in the development of T2DM in the home. Due to the common incidence of familial diabetes clinically, the screening of disease-associated susceptibility genes in members of a family with T2DM is important for preventing and treating T2DM.

Some previous studies focusing on genetic associations demonstrated that single nucleotide polymorphisms (SNPs) are obviously associated with T2DM susceptibility. There is a lot of information on genetic architecture of T2DM, such as high levels of polygenicity and small effect-sizes for most genetic risk variants. At least 75 independent genetic loci related to T2DM have been discovered until now. Some of the variants are associated with the protection against T2DM risk, while some variants are considered as risk markers and correlated with T2DM susceptibility. Therefore, it is critical to clarify the risk and pathogenesis of T2DM through discovering the genetic variants.

Whole exome sequencing offers a useful strategy for identifying genes involved in human disorders, such as diabetes. Whole exome sequencing, in a few studies, demonstrated that the occurrence of T2DM in family members was associated with low frequency and rare mutations of genes. In recent years, whole exome sequencing has been widely applied for identifying susceptibility gene variants or insulin mutations in diabetes patients, and other diseases caused by potential genetic variations in China. The incidence of T2DM is high clinically, with an obvious genetic predisposition. Presently, there have been no studies focusing on familial diabetes based on whole exome sequencing method in China until now. Therefore, whole exome sequencing for family members with T2DM and discovery of disease-related susceptibility genes are of great significance for preventing and treating diabetes.

T2DM, as a genetically heterogeneous disease, is commonly characterized by a few specific genes involved in different races or families of the same race. Presently, there is no research focusing on application of whole exome sequencing technology for determining diabetic families in Yunnan province of China. Therefore, the purpose of this study was to discover the possible pathogenic genes and mutation sites of the pathogenic genes in T2DM patients, using whole exome sequencing method in Han nationality of Kunming in China. This study will provide promising clues for preventing and treating familial diabetes.

### Materials and Methods

#### Study Design and Patients

We recruited Chinese patients with a three-generation family history of diabetes (6 members with diabetes and 1 member without diabetes) (Figure 1). The pro-band was a male patient with a ten-year history of diabetes in the second generation. The oral glucose tolerance test (OGTT) confirmed the diagnosis of diabetes. T2DM was considered in all patients except for the wife of pro-band according to the characteristics of medical history and laboratory examinations. There was no information suggesting consanguineous marriage in this family.

#### Type 2 Diabetes Mellitus Diagnosis

Diabetes mellitus was diagnosed according to the criteria of diabetes mellitus published by WHO in 1999 including: fasting plasma glucose (FPG) ≥7 mmol/L, plasma glucose 2 hours after meal (2hPG) and OGTT 2hPG>11.1 mmol/L. The diagnostic criteria for T2DM were based on several items: ① the symptoms of polyuria, polydipsia, and without typical weight loss. ② Oral hypoglycemic drugs were effective in controlling blood sugar and there was no spontaneous diabetic ketoacidosis history. ③ Islet cell antibody (ICA), insulin antibody (IAA) and glumatic acid decarboxylase (GAD) antibodies were negative. ④ Fasting serum C-peptide level was within the lower limit of normal reference range. ⑤ The clinical findings of insulin resistance (acanthosis nigricans, hypertension, dyslipidemia). Together with the age of patients (all more than 40 years old), all these supported the diagnosis of T2DM. Moreover, type 1 diabetes mellitus (T1DM) was excluded with presence of autoantibody positivity, severe deficiency and absence of any suggestive symptoms of the other reasons for diabetes.

This study has been approved by The first affiliated hospital of Kunming medical University (Approval No. 2020-L-18). This study was conducted in accordance with the Declaration of Helsinki. All the family members provided written formed consent and approved this study.
Whole Blood Genomic DNA Extraction
Total of 2 mL peripheral blood was collected from each of the seven family members. The DNA samples were extracted from the peripheral blood of these seven family members with whole blood genomic DNA rapid extraction kit (Biomed Corporation, China), as instructed by the manufacturer. Then, the DNA samples from the seven members of this family were sent to BGI-Shenzhen (Shenzhen, China) for whole exome sequencing.

Whole Exome Sequencing
Agilent SureSelect kit V6 was used for the exome capture. Sequencing data were generated using the pair-end 100bp sequencing strategy on the BGISEQ-500 platform.

In silico Prediction Methods for Analyzing SNPs

Biological Information Data Analysis with Significance Analysis of SNP/Indel Site
The probability of exon DNA mutation at any point in the family members can be calculated according to the following binomial distribution probability formula: \( P_v = p^m (1-p)^{n-m} \) Where, “p” was the minor allele mutation frequency (MAF) in the normal population (0<p<1), “n” was the total sample, “m” was the number of samples with the mutation observed, “Pv” was the probability of “m” mutations observed in “n” samples. In this study, according to the calculation, n=7, m=6. Mutation frequency of each site in the normal population was analyzed using the corresponding MAF in 1000 human genome database. Some single nucleotide polymorphisms (SNP) and Indel had no mutation frequency information in the 1000 human genome database. At this time, rs of SNP and Indel were used to query dbSNP database to obtain the maximum MAF value of this site. There was no MAF value corresponding to the new SNP and Indel sites in 1000 human genome database and dbSNP database. A relatively small MAF value was set for Manhattan map and subsequent analysis. In this study, the minimum MAF value was set as 1e-4.

Biological Information Data Analysis with Manhattan Map
The Manhattan map is a general method for displaying and analyzing the significance of each mutation in a genome. In order to obtain significant SNP and Indel site of clustered diabetic members in this family, Manhattan map was used to display information of SNP and Indel site. The specific process was described as follows: ① the binomial distribution probability formula was used to calculate the mutation probability \( P_v \) of each SNP and Indel site. ② Manhattan function in qqman, a software package used in R language for analyzing SNP, was applied to display information of diabetic exon SNP and Indel in this family. ③ The cutoff value of rs shown in Manhattan map was set as \( P_v=1e-9 \). ④ There were many significant SNP and Indel sites. In order to obtain significant and relatively clear biological significance of SNP and Indel sites, the Manhattan map was divided into two categories, including showing all SNP/Indel sites and only showing SNP/Indel sites with the altered protein functions.

Biological Information Data Analysis with Analysis of Specific SNP/Indel Proteins and Their Potential Mechanisms
To determine whether a particular SNP/Indel is directly related to occurrence of diabetes, the SNP/Indel is first mapped into structure of corresponding protein. The effect of SNP/Indel on protein function was predicted using PDB sum. The protein with SNP/Indel mutation was then localized in the KEGG pathway to analyze whether this pathway was related to the onset of diabetes.

Statistical Analysis
All of the statistical analyses for the bioinformatic data were carried out using R software (version: 2.15).
Biological information data were analyzed according to the SNP/Indel site and Manhattan map. The average number of original bases, clean bases and average GC content were counted and analyzed. The SNPs and Indel were also counted. The value of Q20 or Q30 more than 80% was assigned as good sequencing quality.

**Results**

**DNA Sample Quality Assessment**

Quality evaluation was conducted on 7 DNA samples. The concentration of DNA samples was detected with Qubit, and then the total amount of DNA was calculated. If the total amount of DNA was more than 1μg, the standard was reached. DNA integrity was then measured by 1% agarose gel electrophoresis. It can be seen from the electro-phoretogram that the main bands of DNA samples were all over 23kb (Figure 2). Although some samples were slightly degraded, they met the quality requirements of the library construction.

**Exon Sequencing Quality**

All exon sequencing was carried out on 7 DNA samples, and the average number of original bases was 16065.63Mb, of clean bases was 16024.34Mb. Clean reads of each sample demonstrated high Q20 and Q30 (>80%), indicating that the sequencing data had good sequencing quality. The average GC content was 50.68%. The average sequencing depth of the target region was about 137.37X (Table 1).

**Statistical Results of Mutation Detection**

**SNP Detection Results**

A total of 130,693 SNPs were discovered in all samples. Within the coding region, there were 10,367 synonymous mutations, 10,368 missense mutations, 115 stop gain, 38 stop loss, 30 start loss, and 147 splicing.

**Indel Detection Results**

A total of 15,928 Indel were discovered in all samples. Within the coding region, there were 278 frame shift, 91 non-frame shift insertion, 144 non-frame shift deletion, 0 stop loss, 3 start loss, and 58 splicing.

**Bioinformatic Data Analysis**

Whole exome sequencing results showed that there were 8383 exon DNA mutations in 6 diabetic family members, while no mutation was found in 1 non-diabetic control member. Among them, there were 7603 single nucleotide polymorphism mutation sites and 780 insertion-deletion mutation sites.

After calculating the significance of each SNP/Indel locus, Manhattan diagram was used to display mutation sites. By setting a specific cutoff value, the significant SNP/Indel sites can be discovered. In Figure 3A, there were 59 SNP sites with high significance ($P=1e-9$), among which there were 24 SNP sites with moderate and high biological functions. After removing the SNP site without coding function in the intergenic region, 10 SNP sites with significant and high biological functions were finally discovered (Figure 3B). Using the same method, 23 Indel mutation sites with high significance ($P=1e-9$) were discovered (Figure 4A), among which 12 Indel mutation sites with significant and moderately high biological functions were discovered (Figure 4B).

In summary, 22 significant SNP mutation sites and Indel mutation sites in this family were obtained (Table 2). These gene mutation sites were located on 19 genes, including zinc finger CCHC domain-containing protein 3 (ZCCHC3), synapsin 2 (SYN2), ribosomal protein L14 (RPL14), selenite [Se(IV)] reductase D (SRRD), adenosyl methionine decarboxylase 1 (AMDC1), calcium/calmodulin dependent protein kinase kinase 2 (CAMKK2), zinc finger protein 787 (ZNF787), ring finger protein 157 (RNF157), nuclear pore

![Figure 2](https://doi.org/10.2147/IJGM.S335090) Agarose gel electrophoresis of 7 DNA samples from family members. M1 is λ-Hind III digest DNA Ladder. M2 is D2000 DNA Marker, and 1–7 are DNA samples.
Table 1 Statistical Graph of Exon Sequencing Data from 7 Family Members

<table>
<thead>
<tr>
<th></th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
<th>Subject 5</th>
<th>Subject 6</th>
<th>Subject 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw reads</td>
<td>158,603,486</td>
<td>156,383,178</td>
<td>155,608,204</td>
<td>165,373,174</td>
<td>158,013,548</td>
<td>162,535,482</td>
<td>168,076,764</td>
</tr>
<tr>
<td>Raw bases</td>
<td>158,603.35</td>
<td>156,383.12</td>
<td>155,608.82</td>
<td>165,373.32</td>
<td>158,013.35</td>
<td>162,535.55</td>
<td>168,076.68</td>
</tr>
<tr>
<td>Clean reads</td>
<td>158,214,218</td>
<td>156,000,878</td>
<td>155,232,348</td>
<td>164,969,052</td>
<td>157,635,416</td>
<td>161,235,706</td>
<td>167,666,246</td>
</tr>
<tr>
<td>Clean bases</td>
<td>158,195.11</td>
<td>155,979.60</td>
<td>155,202.60</td>
<td>164,944.42</td>
<td>157,615.58</td>
<td>162,111.28</td>
<td>167,666.36</td>
</tr>
<tr>
<td>Clean read1 Q20 (%)</td>
<td>99.74</td>
<td>99.74</td>
<td>99.74</td>
<td>99.74</td>
<td>99.75</td>
<td>99.75</td>
<td>99.75</td>
</tr>
<tr>
<td>Clean read2 Q20 (%)</td>
<td>96.63</td>
<td>96.69</td>
<td>96.66</td>
<td>96.67</td>
<td>96.67</td>
<td>96.65</td>
<td>96.66</td>
</tr>
<tr>
<td>Clean read1 Q30 (%)</td>
<td>96.51</td>
<td>96.63</td>
<td>96.89</td>
<td>96.65</td>
<td>96.72</td>
<td>96.55</td>
<td>96.52</td>
</tr>
<tr>
<td>Clean read2 Q30 (%)</td>
<td>88.25</td>
<td>88.43</td>
<td>88.39</td>
<td>88.44</td>
<td>88.58</td>
<td>88.80</td>
<td>88.38</td>
</tr>
<tr>
<td>Sequencing Depth</td>
<td>128.99</td>
<td>126.18</td>
<td>118.85</td>
<td>147.66</td>
<td>135.58</td>
<td>150.89</td>
<td>153.44</td>
</tr>
</tbody>
</table>

Notes: Raw reads: total number of original sequences. Raw bases: number of original bases. Clean reads: number of sequences obtained after removal of low-quality reads. Clean bases: number of bases obtained after removal of low-mass bases. Clean read1 and Clean read2: Reads obtained by double-end sequencing. Software will give a quality value for each base, which represents the correct recognition rate of bases as 99.9%. Q20 (%): represents the percentage of bases with a mass value ≥20 (good sequencing quality). Q30 (%): the percentage of bases with a mass value ≥ 30 (good sequencing quality).

complex interacting protein family-B15 (NPIPB15), asparagine-linked glycosylation 3 (ALG3), KIAA0040, microtubule associated serine/threonine kinase 2 (MAST2), estrogen related receptor alpha (ESRRA), chromosome 8 open reading frame 58 (C8orf58), pancreatic lipase-related protein 1 (PNLIPRP1), dachshund homolog 1 (DACH1), metastasis associated colon cancer 1 (MACC1), calpain 9 (CAPN9), and dermokine (DMKN).

Genes Related to Diabetes
In order to analyze how the previously mentioned significant SNP and Indel mutation sites affect the occurrence of diabetes, we mapped the genes corresponding to the 22 mutation sites in KEGG pathway. Finally, the PNLIPRP1 gene and CAMKK2 gene were mapped to the KEGG diabetes-related metabolic pathway.

PNLIPRP1 Gene rs2305205 Site Mutation
The PNLIPRP1 gene is located on chromosome 10 and contains 13 exons. The rs2305205 mutation site in the 8th exon resulted in the replacement of 271 alanine (Ala) of PNLIPRP1 protein with valine (Val) (Figure 5). This SNP site was located in the N-terminal domain of PNLIPRP1 and was involved in triglyceride metabolism. After the mutation, the surface accessibility of the whole protein increased by 27.4Å², and the relative surface accessibility increased from 55% to 58.2%. It is suggested that mutations in the rs2305205 site might alter the function of the protein.

According to the KEGG pathway map for PNLIPRP1 involved triglyceride metabolism (Figure 6), 3.1.1.3 in pink is pancreatic lipase, including pancreatic triglyceride lipase (PLT), PLRP1 and PLRP2. Meanwhile, the 3.1.1.3 in pink mainly involved in metabolism of the fatty acid, monoaoyl-glycerol and 1.2-diacyl-sn-glycerol directly (Figure 6).

CAMKK2 Gene rs778701848 Site Mutation
Positioning in KEGG signaling pathway revealed that CAMKK2 was a key node linking endogenous and metabolically related adiponectin (ADIPOQ), leptin and adenylate-activated protein kinase (AMPK) (Figure 7). Adiponectin and leptin activated AMPK through CAMKK2 phosphorylation to regulate glucose metabolism. As Figure 7 shows, AMPK phosphorylation is involved in the gene expression of plenty of genes, including receptor gamma coactivator 1α (PGC-1α), alcohol regulatory element binding protein 1c (SREBP1c), CREB transcriptional coactivator 2 (TORC2), glucose transporter 4 (GLUT4), acetyl-CoA carboxylase (ACC), rapamycin target protein (mTOR), S6 kinase 1 (S6K1) and insulin receptor substrate (IRS). Therefore, mutations in the CAMKK2 gene may be involved in the development and progression of diabetes by inhibiting the downstream AMPK pathway.

Discussion
The PNLIPRP1 gene was cloned from cDNA in the year 1992 consisting of 467 amino acids and localized in the chromosomal region 10q24-q26 in humans. PTL could hydrolyze triglyceride into diglyceride, and subsequently convert into mono-glycerides and free fatty acid, which are then absorbed by intestinal epithelial cells.
Both PLRP1 and PTL are secreted by pancreatic acinar cells and demonstrated the same affinity with colipase.\textsuperscript{25,26} In order to study the biological function of PLRP1, a previous study\textsuperscript{27} showed that food intake could promote the secretion of PLRP1 in the pancreas, therefore, PLRP1 may play a certain role in food digestion. The in vitro studies\textsuperscript{26,28} showed that PLRP1 could regulate PTL activity by competing for colipase and modulating body fat, obesity, insulin resistance and blood glucose levels in vivo.

In this study, 22 significant SNP mutation sites and Indel mutation sites were discovered, which were then mapped in the KEGG pathway. We found that the PNLIPRP1 gene and CAMKK2 gene were mapped to the KEGG diabetes-related metabolic pathway, therefore, these two variants were discussed. According to the in silico prediction findings, six diabetic patients in this family demonstrated PNLIPRP1 gene rs2305205 site mutation, while normal control did not. The mutation of rs2305205 caused the amino acid encoded by PNLIPRP1 gene from alanine to valine at 271.

**Figure 3** Manhattan map of all SNP sites (A) and Manhattan map of significant SNP mutation sites, marks the SNP sites with high significance ($P<1e-9$) (B).

**Note:** X-axis was the position of the chromosome where the SNP was located. Y-axis was the $-\log_{10}(P)$ value corresponding to each SNP site. $-\log_{10}(P)$ reflects the degree of association between mutation site and disease occurrence. The greater the value of $-\log_{10}(P)$, the stronger the association. The blue horizontal line cutoff value was $P=1e-5$, and the red horizontal line cutoff value was $P=1e-8$. 

\url{https://doi.org/10.2147/IJGM.S335090}

\textsuperscript{DovePress} International Journal of General Medicine 2021:14

\textsuperscript{Zhou et al} Dovepress

\textsuperscript{Dovepress} International Journal of General Medicine downloaded from https://www.dovepress.com/ by 54.187.248.227 on 19-Nov-2021

For personal use only.
The mutation of rs2305205 reduced the function of PLRP1, weakened function of colipase competing with PTL, increased activity of PTL, promoted digestion and absorption of fatty acids. Meanwhile, the mutation of rs2305205 could also increase the content of body fat, and then cause obesity, insulin resistance, and increased blood glucose. Therefore, we concluded that PNLIPRP1 gene mutation may promote the development of T2DM through the previously mentioned pathways.

As the KEGG pathway map of PNLIPRP1 involved in triglyceride metabolism, 3.1.1.3 in pink was considered to be a pancreatic lipase, including pancreatic triglyceride lipase (PLT), PLRP1, and PLRP2. Because lipid metabolism disorder can cause triglyceride to deposit in the target tissue of insulin action, reduce insulin sensitivity, and increase the risk of type 2 diabetes, it is further speculated that the rs2305205 mutation of PNLIPRP1 was associated with the occurrence of diabetes in this family. Presently, there has been no research on the correlation between PNLIPRP1 gene mutation and occurrence of diabetes at home and abroad, and no research reports on the mutation of this gene in pedigree diabetes.

The calcium/calmodulin-dependent protein kinase kinase 2 (CAMKK2) gene is located on chromosome 12q24.2 and contains 16 exons. The rs778701848 site mutation on the 16th exon results in insertion of amino acid 538. The protein encoded by CAMKK2 gene is involved in energy metabolism and glucose homeostasis.
through modulating adiponectin, leptin, and insulin.\textsuperscript{31} The present KEGG pathway map indicated that adiponectin and leptin could activate AMPK through CAMKK2 phosphorylation to regulate glucose metabolism. Meanwhile, AMPK phosphorylation is involved in gene expression of plenty of genes in the process of metabolism. Therefore, loss of CAMKK2 function not only reduced function of adiponectin and leptin, but also reduced levels of

<table>
<thead>
<tr>
<th>SNP/Indel</th>
<th>CHR</th>
<th>BP</th>
<th>P value</th>
<th>Func.</th>
<th>Gene</th>
<th>MAF1000G</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1482434100</td>
<td>20</td>
<td>298,046</td>
<td>6.4E-29</td>
<td>Disruptive_inframe_insertion</td>
<td>ZCCHC3</td>
<td>NA</td>
</tr>
<tr>
<td>Novel SNP</td>
<td>3</td>
<td>12,004,751</td>
<td>1.56E-26</td>
<td>Conservative_inframe_insertion</td>
<td>SYN2</td>
<td>NA</td>
</tr>
<tr>
<td>Novel SNP</td>
<td>22</td>
<td>26,483,980</td>
<td>1E-24</td>
<td>Disruptive_inframe_deletion</td>
<td>SRD</td>
<td>NA</td>
</tr>
<tr>
<td>rs34637914</td>
<td>6</td>
<td>1.11E+08</td>
<td>9.99E-19</td>
<td>Splice_acceptor_variant+intron_variant</td>
<td>AMD1</td>
<td>NA</td>
</tr>
<tr>
<td>rs778701848</td>
<td>12</td>
<td>1.2E+08</td>
<td>9.99E-19</td>
<td>Disruptive_inframe_deletion</td>
<td>CAMKK2</td>
<td>NA</td>
</tr>
<tr>
<td>rs1195035360</td>
<td>19</td>
<td>56,088,071</td>
<td>9.99E-19</td>
<td>Disruptive_inframe_insertion</td>
<td>ZNF787</td>
<td>NA</td>
</tr>
<tr>
<td>rs145150447</td>
<td>17</td>
<td>76,173,781</td>
<td>1.66E-17</td>
<td>Missense_variant</td>
<td>RNF157</td>
<td>0.001597</td>
</tr>
<tr>
<td>rs77563894</td>
<td>16</td>
<td>74,391,834</td>
<td>6.39E-17</td>
<td>Missense_variant</td>
<td>NPIPB15</td>
<td>NA</td>
</tr>
<tr>
<td>rs79144888</td>
<td>3</td>
<td>1.84E+08</td>
<td>1.52E-15</td>
<td>Missense_variant</td>
<td>ALG3</td>
<td>0.003395</td>
</tr>
<tr>
<td>rs774012991</td>
<td>1</td>
<td>1.75E+08</td>
<td>4.64E-14</td>
<td>Conservative_inframe_insertion</td>
<td>KIAA0040</td>
<td>NA</td>
</tr>
<tr>
<td>rs3737738</td>
<td>1</td>
<td>46,035,056</td>
<td>2.58E-13</td>
<td>Missense_variant</td>
<td>MAST2</td>
<td>0.007987</td>
</tr>
<tr>
<td>rs759464632</td>
<td>11</td>
<td>64,315,823</td>
<td>2.6E-13</td>
<td>Structural_interaction_variant</td>
<td>ESRRRA</td>
<td>NA</td>
</tr>
<tr>
<td>rs145988500</td>
<td>8</td>
<td>22,600,944</td>
<td>3.97E-13</td>
<td>Missense_variant</td>
<td>C8orf58</td>
<td>0.008586</td>
</tr>
<tr>
<td>rs2305205</td>
<td>10</td>
<td>1.17E+08</td>
<td>5.66E-12</td>
<td>Missense_variant+splice_region_variant</td>
<td>PNLPPIP1</td>
<td>0.013379</td>
</tr>
<tr>
<td>rs1294271615</td>
<td>13</td>
<td>71,866,526</td>
<td>4.62E-11</td>
<td>Conservative_inframe_insertion</td>
<td>DACH1</td>
<td>NA</td>
</tr>
<tr>
<td>rs117142233</td>
<td>7</td>
<td>20,161,835</td>
<td>5.85E-11</td>
<td>Missense_variant</td>
<td>MACC1</td>
<td>0.019768</td>
</tr>
<tr>
<td>rs1267170</td>
<td>7</td>
<td>20,154,356</td>
<td>9.31E-11</td>
<td>Missense_variant</td>
<td>MACC1</td>
<td>0.021366</td>
</tr>
<tr>
<td>rs28359631</td>
<td>1</td>
<td>2.31E+08</td>
<td>3.57E-10</td>
<td>Splice_acceptor_variant+intron_variant</td>
<td>CAPN9</td>
<td>0.026757</td>
</tr>
<tr>
<td>rs74005499</td>
<td>19</td>
<td>35,511,478</td>
<td>3.77E-10</td>
<td>Frameshift_variant</td>
<td>DMKN</td>
<td>NA</td>
</tr>
<tr>
<td>rs12562749</td>
<td>1</td>
<td>2.31E+08</td>
<td>5.05E-10</td>
<td>Missense_variant</td>
<td>CAPN9</td>
<td>0.028355</td>
</tr>
<tr>
<td>rs776716602</td>
<td>19</td>
<td>35,511,479</td>
<td>5.78E-10</td>
<td>Frameshift_variant</td>
<td>DMKN</td>
<td>NA</td>
</tr>
</tbody>
</table>

Notes: SNP/Indel illustrated the rs number in the dbSNP database, CHR illustrated the chromosome of the SNP/Indel locus, BP illustrated the chromosome position of the SNP/Indel, P value illustrated the probability of SNP/Indel appearing in DNA of 6 diabetic patients simultaneously. This table retained the data with P<1e-9. Func. indicated the mutation type. Gene indicated the gene name corresponding to SNP/indel. MAF1000G represented the frequency of SNP/indel in thousands of human genomes.

Figure 5 Diagram of the effect of rs2305205 mutation on protein function. (A) Indicates the mode of interaction of Ala271 with other amino acid residues in the unmutated local structure. The hydrogen bond is indicated by a green dotted line, the bond length is identified by a number, and the hydrophobic effect is represented by a semicircle. (B) Mode of interaction of Val271 with other amino acid residues in the local structure after mutation. The effect on protein structure before and after the mutation was obtained from PDB sum, where the PDB ID number of PNLPPIP1 was 2 ppl.
Figure 6 The KEGG pathway map of PNLPFR1 involved in the process of triglyceride metabolism.

Note: The rectangle in the figure indicates the enzyme that catalyzes the reaction. Each enzyme is labeled with the EC (Enzyme Commission) number. The circles in the figure indicate the reaction substrate and product. 3.1.1.3 in pink was pancreatic lipase.
Figure 7 The KEGG pathway map of CAMKK2 participating in AMPK signaling pathway.

Note: The reactions involved in CAMKK2 are marked in light red.
adiponectin and leptin in the blood. Moreover, the finding of the present KEGG pathway map for CAMKK2 gene also illustrated plenty of targeting genes and associated signaling pathways, including PGC-1α, SREBP1c, TORC2, GLUT4, ACC, mTOR, S6K1 and IRS, which are consistent with some previous studies. Therefore, this study further proved that CAMKK2 gene was closely correlated with glucose homeostasis, gluconeogenesis, and hepatic glucose production.

In the present study, we found that six diabetic patients in a family demonstrated CAMKK2 gene rs778701848 site mutation and PNLIPRP1 gene rs2305205 site mutation, while normal family member (control) did not. Both of the mutations rs778701848 and rs2305205 could affect the biofunctions of CAMKK2 gene and PNLIPRP1 gene, and increased risk of diabetes in these family members through the previously mentioned signaling pathways, which have not been discussed in the previous documents. The rs778701848 mutation and rs2305205 mutation were identified as the main causes for the familial diabetes in this study, suggesting that CAMKK2 gene and PNLIPRP1 gene might be promising and novel targets for treating diabetes. Therefore, the mutations at rs2305205 of PNLIPRP1 gene and rs778701848 of CAMKK2 gene may be related to the occurrence of diabetes in six type 2 diabetic family members.

There are still some shortcomings in this study. Firstly, peripheral blood of the non-diabetic relatives was not collected, which would be more appropriate for clarifying the discovered genome exon. In a following study, we would compare the data in this study with the diabetic patients without a family history or healthy individuals. Secondly, only one family was included in this study, therefore, relevant studies should be carried out to explore diabetes in families. Thirdly, the correlation between PNLIPRP1 gene and CAMKK2 gene in the diabetes patients was not conducted in this study, which is critical for providing the theoretical basis for early prevention and treatment of diabetes. Fourthly, sample size of this study was relatively small (involving only 3 generations and 7 family members). Fifthly, the clinical issues of the diabetes patients, including duration of diabetes, sex, pharmacological treatment for diabetes, overweight/obesity, were not analyzed, which are important for designing a therapeutic strategy. Finally, for the genes identified in this study, we did not determine which gene belongs to a family cluster and which gene belongs to the diabetes disease its own. The analyses for the previously mentioned genes might be critical for the treatment of diabetes.

In conclusion, by conducting whole exome sequencing and bioinformatic data analysis, we found that all exons of diabetic patients in this family demonstrated an rs2305205 mutation in PNLIPRP1 gene and an rs778701848 mutation in CAMKK2 gene. The rs2305205 mutation and rs778701848 mutation might promote occurrence of type 2 diabetes through reducing sensitivity of peripheral tissues to insulin and reducing insulin secretion.

**Funding**

The study was supported by grants from the National Natural Science Fund of China (Grant No. 81160104, 30760087and 82160165), the Yunnan provincial science and technology department-Kunming medical University applied basic research joint project [Grant No. 2018FE001 (~043)], the Training plan for medical discipline leaders of Yunnan provincial health and family planning commission (Grant No. D-2017039) and The young and middle-aged academic and technical leaders reserve talented person project in yunnan province (Grant NO. 202105AC160093).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


