MAHIDOL UNIVERSITY
Wisdom of the Land

Final Report

Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Exome Sequencing

By

Watip Tangjittipokin

June 2016

Contract No. TRG5780113

Final Report

Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Exome Sequencing

การศึกษาอณูพันธุศาสตร์ของโรคเบาหวานชนิดที่ 2 ในครอบครัวชาวไทยที่มีการถ่ายทอดแบบ autosomal dominant โดยใช้เทคนิค Exome Sequencing

Watip Tangjittipokin

Department of Immunology

Faculty of Medicine Siriraj Hospital

Mahidol University

This project granted by the Thailand Research Fund

Acknowledgment

This research project studies on the Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Exome Sequencing (Project code: TRG5780113). The aim of this project is to study the pathogenesis of familial type 2 diabetes in Thai populations.

The achievement of this project is attributable to an extensive support from my mentors Associate Professor Nattachet Plengvidhya and Professor. Dr. Pa-thai Yenchitsomanus who kindly helped me along the way in conducting this work. I am grateful to the Head of Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, for laboratory facility.

I deeply thank to Assistant Professor Dr. Mayuree Homsanit, Department of Prevention and Social Medicine, Siriraj Hospital, for her valuable advices and guidance on the sample collections.

Watip Tangjittipokin

Principle investigator

Abstract

Project Code: TRG5780113

Project Title: Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Exome Sequencing

Investigator: Watip Tangjittipokin, Ph.D.

Department of Immunology, Siriraj hospital, Mahidol University

E-mail: watip.tan@mahidol.edu

Project Period: 2 year

Abstract

Diabetes is a chronic disease and a major global public health problem including Thailand. Currently, prevalence of diabetes is increasing worldwide. Diabetes is a major cause of illness and premature death. It can cause complications of eyes, kidneys, nerves, heart and brain. Thus, studying its pathogenesis may reduce risk factors and provide early treatment that should help diminishing burden of the disease.

Type 2 Diabetes (T2D) is a metabolic disorder arising from a complex interaction between genes and environment. However, the genetics basis of T2D is not fully understood due to its intricate nature. At present, the effort to find genes associated with T2D is heading toward more promising direction. The search for rare variants that cause abnormal proteins functions (rare variants-common, disease hypothesis). Recent advance in nucleic acid sequencing technology using next generation sequencing (NGS) techniques focusing on non-synonymous variants in the coding regions or exons have considerable success in identification of causative genes of several diseases with high prevalence in the family. Exome sequencing was used to determine particular variations in the exons of each gene on each chromosome that are transmitted with diabetes in extended Thai family with autosomal dominant-inherited T2D.

The aim of this study was to investigate the genetic background of familial T2D using next generation sequencing. The study was carried out in a Thai family with 3 generations affected with diabetes. This family comprised of 27 members, of which 19 had diabetes. To identify the susceptibility variants, exome sequencings were performed in 2 affected family members. We hypothesized that the familial clustering of diabetes is caused by non-synonymous variant that segregated with diabetes. A novel Pro30Leu of PRMT8 was identified and co-segregated with diabetes. This variation was not detected among 400 non-diabetic controls. Thus, Pro30Leu of PRMT8 could be considered as candidate mutation.

Exome sequencing is a valuable tool for revealing new mutations. This is the first evidence that a nonsynonymous variant of PRMT8 may contribute to familial form of T2D. Its impact on glucose homeostasis will be subjected of further investigation.

Keywords: Type 2 diabetes (T2D), Exome sequencing, PRMT8

บทคัดย่อ

รหัสโครงการ: TRG5780113

ชื่อโครงการ: การศึกษาอณูพันธุศาสตร์ของโรคเบาหวานชนิดที่ 2 ในครอบครัวชาวไทยที่มีการถ่ายทอดแบบ autosomal dominant โดยใช้เทคนิค Exome Sequencing

ชื่อนักวิจัย: อ.ดร. วทิพย์ ตั้งจิตติโภคิน

ภาควิชาวิทยาภูมิคุ้มกัน คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

E-mail: watip.tan@mahidol.edu

ระยะเวลาโครงการ: 2 ปี

โรคเบาหวานเป็นโรคเรื้อรังและเป็นปัญหาสาธารณสุขที่สำคัญทั่วโลกรวมถึงประเทศไทย ปัจจุบันความชุกของโรคเบาหวานเพิ่มขึ้นทั่วโลก โรคเบาหวานเป็นสาเหตุสำคัญของการเจ็บป่วยและเสียชีวิตก่อนวัยอันควร ทำให้เกิดภาวะแทรกซ้อนของตา ไต เส้นประสาท หัวใจและสมอง ดังนั้นการศึกษาพยาธิกำเนิดของโรคเบาหวานอาจลดปัจจัยเสี่ยงและการรักษาของโรค

โรคเบาหวานชนิดที่ 2 เกิดจากการความผิดปกติของยีนร่วมกับสิ่งแวดล้อม ในปัจจุบันสาเหตุทางพันธุกรรมยังไม่ทราบสาเหตุที่ชัดเจน เนื่องจากลักษณะที่ซับซ้อนของโรค ในปัจจุบันมีความพยายามที่จะหายีนที่เกี่ยวข้องกับโรคเบาหวานชนิดที่ 2 โดยค้นหาความผิดปกติที่พบความถี่น้อย และก่อนให้เกิดความผิดปกติของการทำงานของโปรตีน ปัจจุบันเทคโนโลยีในการหาลำดับเบสได้พัฒนาขึ้น โดยใช้เทคนิค next generation sequencing (NGS) โดยเทคนิคนี้มุ่งเน้นหาลำดับเบสที่ผิดปกติในส่วนของ exons

จุดมุ่งหมายของการศึกษาครั้งนี้คือการตรวจหาความผิดปกติทางพันธุกรรมของครอบครัวของผู้ป่วยโรคเบาหวานชนิดที่ 2 โดยใช้เทคนิค exome sequencing โดยการศึกษานี้ได้ทำการศึกษาในครอบครัวไทยที่เป็นโรคเบาหวานชนิดที่ 2 มีสมาชิกในครอบครัวเป็นโรคถึง 3 รุ่น โดยประกอบด้วยสมาชิกจำนวน 27 ราย, เป็นโรคเบาหวาน 19 ราย คณะผู้วิจัยได้ทำ exome sequencing ในสมาชิกในครอบครัวที่เป็นโรคเบาหวานชนิดที่ 2 จำนวน 2 ราย และพบว่าความผิดปกติชนิด Pro30Leu ที่ยีน *PRMT8* อาจเป็นสาเหตุในการทำให้เกิดโรคเบาหวานชนิดที่ 2 ในครอบครัว ทั้งนี้ความผิดปกตินี้ไม่พบในอาสาสมัครปกติจำนวน 400 ราย

เทคนิค exome sequencing นับเป็นเครื่องมือที่มีประโยชน์ในการค้นหาความผิดปกติของลำดับเบสที่อาจเป็นสาเหตุของการเกิดโรค และการค้นพบความผิดปกติชนิด nonsynonymous ในยีน *PRMT8* อาจมีความสัมพันธ์กับการเกิดโรคเบาหวานชนิดที่ 2 ดังนั้นการศึกษาหน้าที่และผลกระทบของยีนจึงควรได้รับการศึกษาต่อไป

คำหลัก : โรคเบาหวานชนิดที่ 2, exome sequencing, PRMT8

Executive Summary

Project title (ไทย) การศึกษาอณูพันธุศาสตร์ของโรคเบาหวานชนิดที่ 2 ในครอบครัวชาวไทยที่มีการถ่ายทอดแบบ autosomal dominant โดยใช้เทคนิค Exome Sequencing

(English) Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Exome Sequencing

Principle investigator: Watip Tangjittipokin, Ph.D.

Office: Department of Immunology,
Faculty of Medicine Siriraj Hospital, Mahidol University
Tel 02-419-6660

E-mail: watip.tan@mahidol.edu

Objective

To identify novel genes responsible for autosomal dominant type 2 diabetes in Thai families by next generation sequencing

Methodology

In this study, probands and their relatives from selected families were analyzed for novel Type 2 Diabetes (T2D) genes by exome sequencing. All protein-coding regions of the genome were sequenced. After applying certain filtering methods putative causal variants were identified. These variants were tested for segregation with diabetes in the families. Once possible pathogenic genes have been identified, additional genotype in relatives and non-diabetic controls were performed. All strategies are included: (1) Subject recruitment: Patient and Family selection, control, (2) Blood collection (3) Exome sequencing (4) Identification of putative variants (5) In silico protein-function prediction (6) Segregation analysis in T2D family (7) Additional genotyping of possible pathogenic genes(s) in non-diabetic controls (8) Statistical analysis

Output/outcome

To identify the susceptibility variants, exome sequencings were performed in 2 affected family members. The familial clustering of diabetes is caused by variant that segregated with diabetes. A novel Pro30Leu of PRMT8 was identified and co-segregated with diabetes. This variation was not detected among 400 non-diabetic controls. Thus, Pro30Leu of *PRMT8* could be considered as candidate mutation. Exome sequencing is a valuable tool for revealing new mutations. This is the first evidence that a nonsynonymous variant of *PRMT8* may contribute to familial form of T2D. Its impact on glucose homeostasis will be subjected of further investigation. Furthermore, studies of the impact of these variations on clinical characteristics and biochemical parameters of the patients provided more insight into the impact on human glucose homeostasis. It would facilitate more appropriate measures for prevention and treatment of diabetes.

Watip Tangjittipokin
Principle Investigator

Objective

To identify novel genes responsible for autosomal dominant type 2 diabetes in Thai families by next generation sequencing

Introduction

Diabetes Mellitus (DM) is a complex metabolic disorder characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both. It is one of the most common chronic diseases and its prevalence continues to rise significantly worldwide. There are substantial evidences showing that it is epidemic in many developing and newly industrialized nations. In most developed countries, diabetes is the fourth or fifth leading cause of death. International Diabetes Federation (IDF) showed that the number of adult diabetic patients in 2011 was 336 million and it will increase to 552 million in 2030 and most of them have type 2 diabetes. Therefore, substantial burden to public-health system in many countries including Thailand is inevitable. In Thailand, the most recent survey was conducted in 2000 and the prevalence of diabetes in Thai adults was estimated at 9.6% (2.4 millions) and 5.4% (1.4 millions) had impaired fasting glucose.

Diabetes is a major cause of illness and premature death. The chronic complications are involving but not limit to the eyes, kidneys, nerves, heart and brain. The costs of treatment are enormous and create a very strong burden to national health budget. Thus, studying pathogenesis of diabetes may help reducing risk factors, delaying onset of the disease, providing early treatment and lessening the impact on health care cost. Countless evidences have showed that genetic factors are important in pathogenesis of diabetes especially T2D which is by far the most common form of all. Recently, genome-wide association studies (GWAS) have identified genes or regions of the genome associated with the disease. This method is widely used because it is a comprehensive study of all areas of the genome. Since 2007, more than 40 genetic loci linked to T2D have been discovered. However, there are certain limitations of this approach. Variations revealed by GWAS could explain less than 20% of disease inheritability. Moreover, common single nucleotide polymorphisms (SNPs, allele frequency greater than or equal to 0.05) were used in GWAS derived from Caucasians, 80% of them lied in introns, which is not directly related to protein synthesis. In addition, requirement of very large sample size and high cost of microarray genotyping make this strategy inappropriate to study genetics of T2D in Thais especially in patients with clear evidence of disease transmission in the family.

At present, the effort to find genes associated with T2D is heading toward more promising direction. The search for rare variants that cause abnormal proteins functions (rare

variants-common. disease hypothesis). Recent advance in nucleic acid sequencing technology using next generation sequencing (NGS) techniques focusing on non-synonymous variants in the coding regions or exons have considerably successful in identification of causative genes of several diseases with high prevalence in the family. We will use exome sequencing to determine particular variations in the exons of each gene on each chromosome that are transmitted with diabetes in Thai family with autosomal dominant-inherited type 2 diabetes. The knowledge gained from this study will provide novel information regarding pathogenesis of T2D in Thais. It may lead to development of new therapeutic strategies, better disease screening and even prevention of T2D.

American Diabetes Association (ADA) has classified diabetes into 4 major groups including type 1 diabetes, type 2 diabetes, gestational diabetes and other specific types of diabetes. Type 1 diabetes (T1D) and type 2 diabetes (T2D) which are major forms of diabetes. T1D is mainly caused by autoimmune-mediated destruction of beta cells and usually occurs in children. T2D is the most common subtypes, accounting for 80-90% of cases diabetes. It is caused by an inability of beta cells to adequately compensate for insulin resistance and usually occurs in middle age or older people. Gestational diabetes mellitus (GDM), a less frequent subgroup of DM, is explained as any degree of glucose intolerance with onset or first recognition during pregnancy. Other specific types of DM account for only 1-2% of cases.

T2D results from interaction between environmental factors and strong hereditary components. Estimation of heritability of T2D range from 20%-80% comes from a variety of population, family, and twin-based studies. The risk of developing T2D is 40% for individuals who have one parent with T2D and 70% if both parents are affected. First degree relatives of individuals with T2D are about 3 times more likely to develop the disease than those without a positive family history. The concordance rates in monozygotic twins and in dizygotic twins are approximately 70% and 20%-30% respectively. The observed familial risk is higher when studies are restricted to parents in the 35-60 year age range, indicating the greater role of environmental factors in those who develop diabetes late in life.

Very few T2D risk genes were identified using candidate gene and linkage-based approaches. Even though, the genome-wide association studies (GWAS) have uncovered genetic loci (including several that were not previously known to play any role in T2D) associated with T2D. Taken together, the loci discovered to date explain only a small proportion of the observed heritability. Since then, the strategy to identify T2D genes has

turned to the concept of “rare variant – common disease” hypothesis. One attractive way to uncover this rare variant is to study family in which several members are affected with diabetes. Since the possibility that the causative variant is aggregated in the family is higher. Moreover, if the mode of disease inheritance can be ascertained, the likelihood of unearthing the variant of interest is encouraging. This notion has proved to be extremely successful in studying an autosomal dominant monogenic form of T2D, or monogenic diabetes of the young or MODY (previously referred to as maturity onset diabetes of the young). Up till now, at least 13 subtypes of MODY have been uncovered. Since then, there is a breakthrough in understanding beta-cell function, development and neogenesis. New drug for treatment of diabetes has been developed. However, there is also a family with diabetes transmitted in autosomal dominant fashion but the onset of the disease is greater than MODY (usually after 40 years). The affected members are usually obese or overweight with features of insulin resistance while MODY patients are generally lean and have insulin secretion defect. This discrepancy in clinical phenotypes may imply that different genes are responsible for diabetes in this particular family. Recently, the advance in high-throughput sequence capture methods and next-generation sequencing technology (NGS) offer new opportunities for Mendelian disorder research. The high-throughput sequence capture methods are able to separate the collected exons from human genome in a more efficient and cost-effective way than traditional PCR-based methods by reducing cost and time for designing PCR-primer and amplification of target of interest. These sequence capture methods are able to target the interested genomic regions more than ten megabases and also enable the enrichment of exome in a single experiment. These methods are commercially marketed, for example the NimbleGen Sequence Capture technology (http://www.nimblegen.com) and Agilent SureSelect Target Enrichment technology (http://www.home.agilent.com). High-throughput sequencing data are produced by NGS technologies using various strategies depending on a combination of template preparation, sequencing and imaging, alignment of genome and assembly method. Many platforms for massively parallel DNA sequencing are commercially available. The high-throughput sequence capture method linked with the high-throughput sequencing data produced by NGS technologies guarantees an adequate depth of sequencing coverage to accurately detect the variants in the exome or targeted regions. Currently, exome sequencing has been successfully identify the causal variants in several Mendelian disorders, such as *MYH3* in Freeman-Sheldon syndrome, *MLL2* in Kabuki syndrome, *DHODH* in Miller syndrome and KCNJ11 as the thirteenth MODY gene, etc.

Therefore, identification of genes causing autosomal dominant T2D in Thai family by using exome sequencing is a major goal in this study.

There are 2 major steps to identify variants by NGS including *i)* sample preparation and *ii)* sequencing and analysis. In the first part, Target Enrichment technology will be used to capture the exons from genomic DNA. In the second part, Illumina platforms were used to sequencing DNA to identify putative variants. Once possible pathogenic mutations are identified, it will be examined for segregation with diabetes in the family. Impact on protein function will be investigated further in future study. Siriraj Diabetes Research Group (SiDRG) has actively been participating in molecular genetics study of diabetes in Thais during the past 13 years. MODY 9 which is caused by mutation of *PAX4* was discovered by our group. In addition, we have replicated the T2D GWAS loci in our large cohort of Thai T2D patients using genetic case-control association study. We are enthusiastically using exome sequencing exploring for novel MODY gene. We are certain that this approach would primarily generate the information of novel genes responsible for autosomal dominant T2D in Thai population base on our extensive experience. Further studies in genetics and molecular biology will promote better understanding of mechanisms underlying the pathogenesis of the diabetes, which may be useful for clinical implication and treatment of the disease.

Research methodology

1. Subjects

In this study, probands and their relatives from selected families will be analyzed for novel T2D genes by exome sequencing. All protein-coding regions of the genome will be sequenced. After applying certain filtering methods putative causal variants will be identified. These variants will be tested for segregation with diabetes in the families. Once possible pathogenic genes are identified, additional genotype in relatives and non-diabetic controls will be performed.

1.1 Patients and Families

Type 2 diabetes probands and their relative were recruited at the Diabetic clinic of Siriraj Hospital, Mahidol University, Bangkok, Thailand, according to following criteria:

- The proband and least or one first degree relative diagnosed with T2D.
- Two or more generations were affected by diabetes.
- No history of diabetic ketoacidosis (DKA).
- Anti-glutamic acid decarboxylase (GAD) antibody which is a marker for type 1 diabetes is negative.

1.2 Non-diabetic controls

Four hundred of non-diabetic subjects were recruited from health checkup facility, Department of Prevention and Social Medicine, Siriraj Hospital, Mahidol, University, Bangkok, Thailand according to the following criteria:

- Age greater than 40 years
- No family history of diabetes in first-degree relatives.
- Not in end stage of hepatic and renal diseases.
- No autoimmune disease
- No hypertension.
- Not in terminal state of cancer.
- Not receiving immunosuppressive or immunostimulant drugs.
- Not receiving drugs that effect blood glucose level or lipid metabolism.
- Fasting plasma glucose (FPG) less than 100 mg/dl.
- Glycosylate hemoglobin (HbA1c) less than or equal to 5.6 %
- 2-hr plasma glucose less than 140 mg/dl during an OGTT

All subjects in this study were informed the purpose and extent of the study before signing a consent form of willing to participate prior to being enrollment in this study. The informed-consent procedures have been approved by Ethic Committee, Faculty of Medicine Siriraj Hospital, Mahidol University. Certificate of Approval number is Si 107/2009. Relevant history, physical examination, anthropometric measurement, pedigree information and blood samples were collected.

2. Experimental design

In this study, exome sequencing was performed to identify the causative gene responsible for T2D in Thais family. The proband and relatives from selected T2D family were collected peripheral blood and extracted DNA from WBCs by using standard phenol/chloroform method. DNA samples from the proband and affected individual underwent exome sequencing. All protein-coding regions of the genome were sequenced. Certain filtering methods were used to identify candidate variants. Then, selected variants were tested for segregation with diabetes in selected T2D family and Sanger sequencing confirmed these variants. Once the segregation was established, additional genotyping of these variants in and non-diabetic control and other T2D probands were performed. Then, novel causative variants were identified. Work flow of this study was shown in Figure 1.

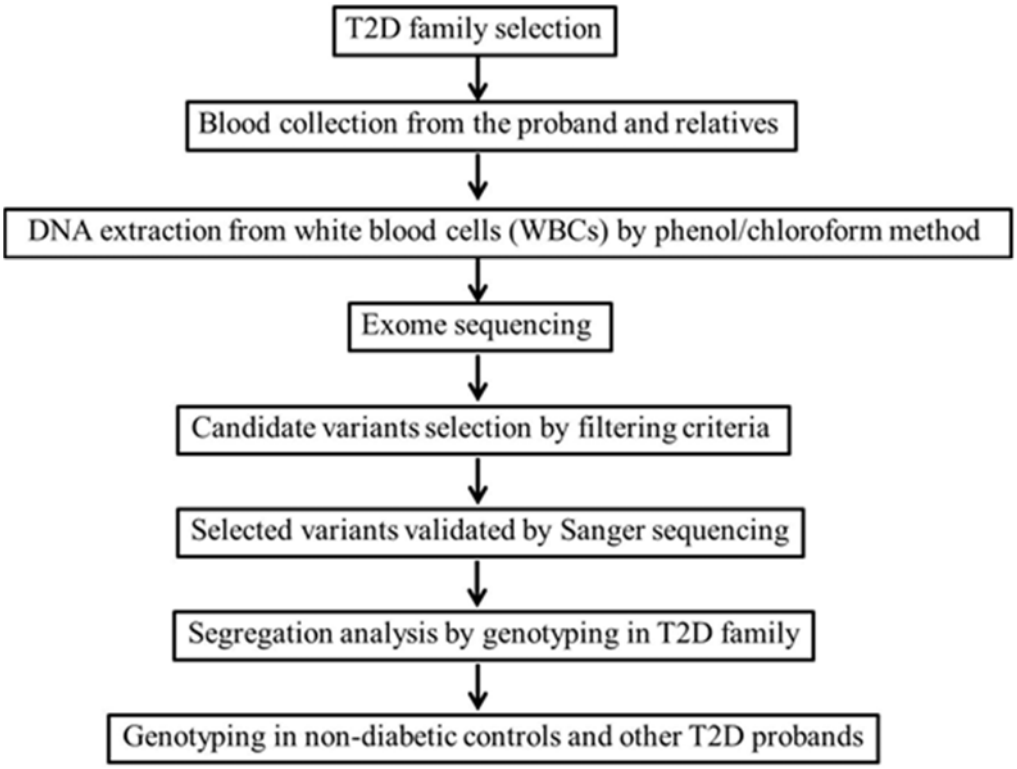

Figure 1 Work flow of experimental design

3. Blood samples collection and laboratory assays

Blood samples were obtained from subjects by venipuncture. Approximately 20 ml of peripheral venous blood was collected from each subject into two sterile tubes. Roughly 10 ml of peripheral venous blood was filled into a first sterile tube to measure plasma glucose, HbA1C, serum total cholesterol, LDL cholesterol, HDL-cholesterol, triglyceride, creatinine, anti-GAD antibody. Moreover, 10 ml of peripheral venous blood was loaded into another sterile tube containing 20 ul of 20% EDTA as anticoagulant for DNA preparation by separation of white blood cells (WBCs) from red blood cells (RBCs). RBCs were lysed by RBC lysis. Genomic DNA were prepared from WBCs by standard phenol/chloroform method and preserved at -20C.

4. White blood cell (WBC) preparation

WBCs were harvested from EDTA-blood sample, for performing DNA extraction. The blood sample was centrifuge at 3,000 rpm, 40C for 5 min. Plasma was discarded and whole cells were washed with sterile PBS two time. Whole cells were collected by centrifugation as above. Three volume of 1XRBC lysis buffer was added into whole cells to lyse red blood cells (RBCs) and centrifuged at 3,000 rpm, 40C for 10 min. Subsequently, the RBC lysate was discarded and WBCs were saved. This RBC lysate step was repeated 2-3 time until the lysate was clear. Then, WBCs were washed with PBS and pelleted by centrifugation. WBCs were stored in a 15-ml screw-cap tube at -200C until used for genomic DNA extraction.

5. Genomic DNA extraction

Standard phenol/chloroform method was used to extract genomic DNA from WBCs for this study. WBC pellet was resuspended by using TE20-5 buffer 4 ml, 10% SDS 200 ul and 2 mg/ml proteinase K 200 ul and incubated at 37 °C for overnight. Subsequently, 2 ml of saturated phenol and chloroform/isoamyl alcohol (24:1) was added into resuspended WBC to separate digested protein from DNA and gently mixed until the whole mixture was appeared like milk. The bottom organic phase was removed by centrifugation at 2,500 rpm for 10 minutes. This phenol/chloroform/isoamyl alcohol extraction step was performed 2 times. Then, 4 ml of chloroform/isoamyl alcohol (24:1) was added into the aqueous phase, gently mixed and centrifuged at 2,500 rpm for 10 minutes. Organic phase was removed. This chloroform/isoamyl alcohol extraction step was also performed 2 times. Then, 1/10 volume of 4 M NaCl and 2 volume of cold absolute ethanol were added and gently mixed to

precipitate DNA. DNA was pelleted by centrifugation at 2,500 rpm for 10minutes. After supernatant was aspirated, pelleted DNA was dried at room temperature. Pelleted DNA was resuspended again by adding sterile water. DNA concentration was measured by using NonaDrop-1000 spectrophotometer (NanoDrop Technologies, DE, USA). DNA sample were labeled sample ID and kept at -80℃ for long term storage.

6. Identification of causative gene causing T2D in Thai family by exome sequencing

6.1 Genomic DNA re-precipitation

DNA samples were diluted to 100 ng/ul by using TE buffer. Then 1/10 volume of 3M sodium acetate (pH 5.2) and 2 to 2.5 absolute ethanol were added into each DNA sample tube and gently mixed to precipitate DNA. Precipitated DNA samples were put on ice or at -20 ℃ for more than 20 minutes and pelleted by centrifugation at 3,000 rpm for 10-15 minutes 4 ℃. After supernatant was decanted, 1 ml of 70% ethanol was added into DNA samples and centrifuged at 3,000 rpm for 10-15 minutes 4 ℃. Supernatant was aspirated and pelleted DNA was dried. Then, pelleted DNA was resuspended again in an appropriate volume of 10 mM TE buffer (pH 7.5). DNA concentration was measured by using NonaDrop-1000 spectrophotometer (NanoDrop Technologies, DE, USA).

6.2 Exome sequencing

DNA samples were performed exome sequencing, executed by Axeq Technologies that divided three parts including:

1) Targeted capture

Consensus Coding DNA Sequence (CCDS) or RefSeq exons from 3 ug of genomic DNA were captured by using the Agilent SureSelect Human All Exon Kit, following the manufacture's protocols (Agilent, Santa Clara, CA, USA).Briefly, DNA was shorn by using the Covaris System. Shorn DNA size is around 150 to 200 bp. Then, fragmented DNAs were purified using AMPure XP beads and the Agilent 2100 Bioanalyzer was used to assess the quality of fragmentation and purification. The ends of fragmented DNAs were modified for target enrichment and adaptors were ligated to the fragments. Size of ligated DNAs is around 250 to 275 bp. The ligated DNAs were purified using the AMPure XP beads, amplified by PCR and captured by hybridization to the biotinylated RNA library baits. Bound genomic DNAs were captured with streptavidin coated magnetic Dynabeads. Then, captured DNAs library were purified by using the AMPure XP beads and re-amplified to add index tags. The amplified indexed libraries were purified again using AMPure XP beads and assessed quality

by Agilent 2100 Bioanalyzer High Sensitivity DNA Assay. The size of DNA libraries is approximately 300 to 400 bp. Capture workflow was shown in Figure 2. The whole-exome DNA libraries were sequenced by using IlluminaHiSeq 2000 Sequencer.

2) Massive parallel sequencing and imaging

This part was performed by Illumina/Solexa platforms. Specific adapter ligated DNA libraries, which were prepared from previously step, were loaded on the surface of a flow cell. DNA polymerase was added to amplify DNA and produce multiple DNA clusters. Each cluster was generated from a single template DNA of ligated DNA libraries. Each template cluster provides free ends to hybridize with a universal sequencing primer. Then, a flow cell containing millions of unique clusters was loaded into the HiSeq 2000 for sequencing and imaging. Each cycle of sequencing step, all four nucleotides were added simultaneously into the flow cell and loaded together with DNA polymerase to incorporate a single nucleotide into the oligo-primed cluster fragments. After the addition of single nucleotide, DNA synthesis was terminated and unincorporated nucleotides were washed. Imaging was then conducted to determine the identity of the incorporated nucleotide. The four color were detected by total internal reflection fluorescence (TIRF) imaging using two lasers. Then, the terminating group and fluorescent dye were removed. Additionally, washing was performed before starting the next cycle. The number of cycles was determined by user-defined instrument settings.

3) Sequence Reads Mapping and Variant Calling

Sequence reads were mapped to the reference human genome (UCSC NCBI37/hg19) by using the Burroughs-Wheeler Aligner (BWA). Variants detection was done by SAM tools. Detected variants were filtered to fit a quality and depth of variants. Then, the variants were reported.

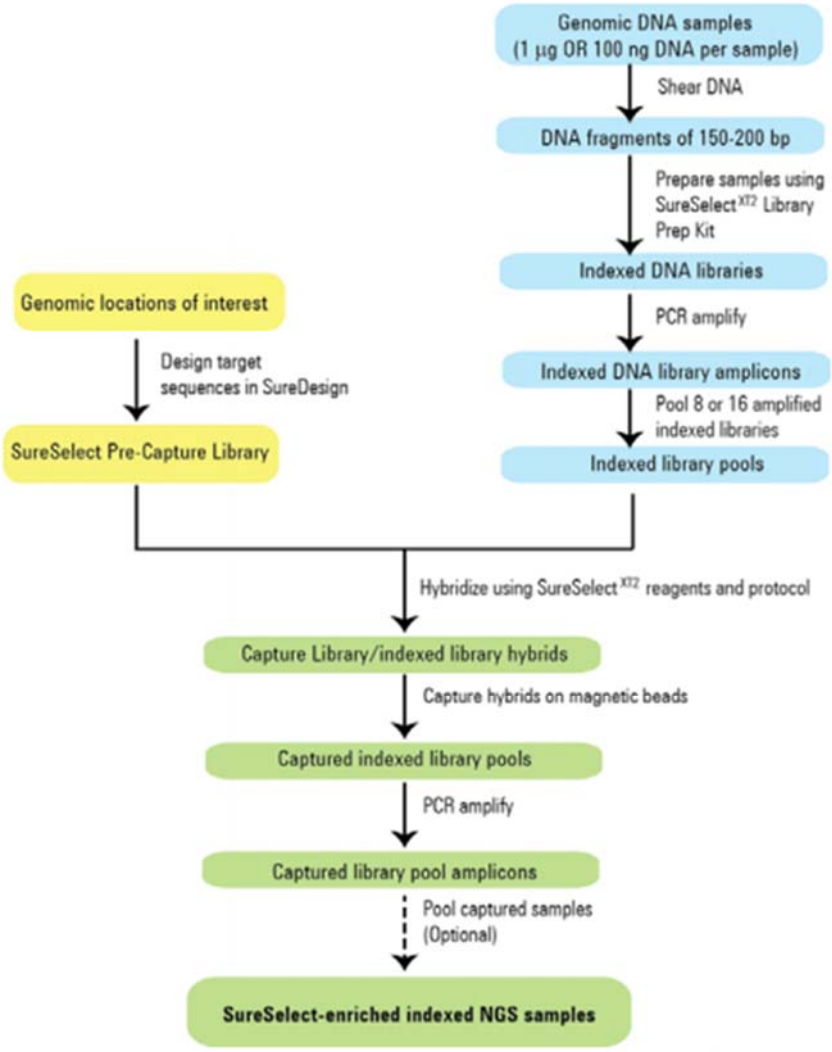

Figure 2 Overall sequencing capture workflow

(From: Agilent SureSelect Human All Exon Kit protocol)

7. Candidate variants selection

Approximately 20,000 coding variants were identified from exome sequencing. Thus, certain filtering methods must be applied to select candidate variants as shown in figure 2. Variants which were identified in only affected and were not present in 16 in-house control subjects. To distinguish potentially pathogenic variants from other variants, functional variants (non-synonymous, nonsense, located in the canonical splice sites, or coding indels) have been focused. As the hereditary mode in the family is autosomal-dominant, the candidate variants were expected to be heterozygous variants with minor allele frequency (MAF) <1% from 1000 Genome Project (available in the public domain at http://www.1000genomes. org/), and Exome Sequencing Project (ESP) database (available in the public domain at http://evs.gs.washington.edu/EVS/) and dbSNP135 database (available in the public domain at http://www.ncbi.nlm.nih.gov/SNP/) were included for analysis. Then, these variants were studied further for effects on protein expression and function by in silico programs, including PolyPhen2, SIFT, Mutation Taster and VarioWatch, Provean and SNP&GO were shown Table 1 to help select proper candidate variants and those showing deleterious effect on protein prediction in at least 4 out of 6 programs were selected for testing segregation with diabetes within selected T2D family.

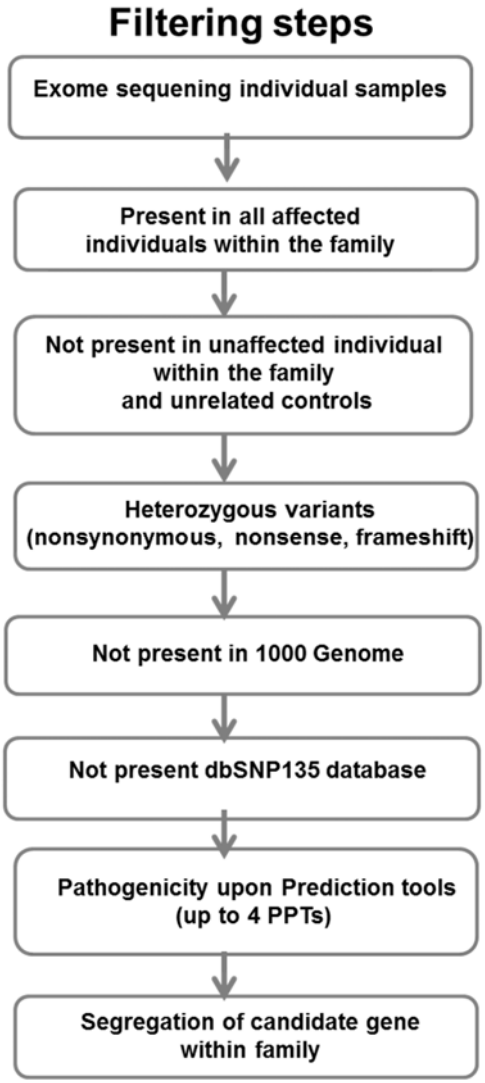

Figure 3 Filtering criterion steps for selection candidate variants.

Table 1 *In silico* prediction program for impact of non-synonymous variants on protein function.

Program	Website	Method	Parameter used	Prediction
SIFT	http://sift.jcvi.org/	Conservation of protein homologues	Evolutionary conservation	Tolerated, Damaging
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	Naive Bayes classifier	Evolutionary conservation+ structural effects	Benign, Probably damaging
Mutation Taster	http://www.mutationtaster.org/	Naive Bayes classifier	Evolutionary conservation	Disease causing, Polymorphism
VarioWatch	http://genepipe.ncgm.sinica.edu.tw/variowatch/main.do	Conservation of protein homologues	Evolutionary conservation	High, Medium, Low
PROVEAN	http://provean.jcvi.org/index.php	Conservation of protein homologues	Evolutionary conservation+ structural effects	Deleterious, Neutral
SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go/	Support vector machine	Evolutionary conservation+ Local sequence+gene ontology score	Neutral, Disease

8. Segregation analysis of selected variants

Candidate variants showing deleterious effect on protein, prediction in at least 4 out of 6 programs were chosen for testing segregation with diabetes within selected T2D family by using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) method.

8.1 Polymerase chain reaction (PCR)

Forward and reverse oligonucleotide primers were designed to amplify DNA fragments containing candidate variant. Each DNA fragment was amplified by using genomic DNA as a template by PCR method. The PCR were optimized of each DNA fragment. The total volume of PCR was performed in 12.5 ul. PCR reaction was done in the 200 ul microtube. PCR mixture contained 1.25 ul of 10X PCR buffer, 1.25 ul of 2mM dNTP, 0.375 ul of 50 mM MgCl2, 0.625 ul of 10 pmol/ml of each F- and R primers, 0.05 ul of Immolase DNA polymerase, 5.825 ul of sterile distilled water, and 2.5 ul of 25 ng/ul of DNA template. List of the primer sequences and proper PCR condition was shown in Table 2-4.

8.2 Detection of PCR product

The PCR products were examined by agarose gel electrophoresis using 1.5% agarose gel. One microliter of 6X gel-loading buffer were mixed with 5 ul of PCR product and the mixture was then load into one well of agarose gel. Three microliters of 100 bp DNA Ladder were loaded into one well of agarose gel as a standard-size markers. Electrophoresis was run at 120 volts for approximately 15 minutes. The agarose gel was stained in 2 ug/ml ethidium bromide solution, and PCR products were detected using Gel documentation.

Table 2 Primers for amplification of genes and their RFLP condition that were predicted by prediction programs (positive 6 in 6 programs)

Gene	Primer name	Primer sequence (5'→3')	Tm (°C)	Product size (bp)	Restriction Enzyme	PCR product (bp) after digestion	
						Major Allele	Minor allele
MYMO3	MYMO3-F MYMO3-Rm	CCTGACTCCCAGCCCAGA CCACCTTCCCCAGGTGACTA	58	208	MaeI	189+19	208
GJA5	GJA5-F GJA5-R	CCAGGCTGAATCGTATCACAC TGGATGGATGGATCAGTGG	56	312	DdeI	312	55+103+154
SPTA1	SPTA1-F SPTA1-Rm	AGCAGAAAAGGTGAGAAAAGG TTGGCAGATGATGAAGATTACTA	58	347	MaeI	22+138+187	138+209
CACNA1S	CACNA1S-F CACNA1S-Rm	TGCTGGGCACCAGAAACAAAGAC CCTGGGCATGCAGCTCCTTG	61	348	StyI	19+329	348
IRX1	IRX1-F IRX1-R	CAAGATGACCCTCACGCAGG GCCAGCGACCAGATCTTGG	58	459	HpyCH4IV	140+319	70+140+249
USP6NL	USP6NL-F USP6NL-Rm	TTCCCGTCCTCAGCATCC TCAAATATCAGGAAGGAGTTGGTG	56	200	NlaIV	22+178	200
SPINT1	SPINT1-F SPINT1-Rm	GTGCCTACAGTTATTGTCCA TAACAACCACCATAGGTAAAC	49	306	HPaII	22+284	306
MYH4	MYH4-F MYH4-Rm	ATTCTGCTTGGGAAGCCTTTCCTG TTGTGCGGTGCATCATCCCTA	61	215	MaeI	215	20+195
PNPLA6	PNPLA6-Fm PNPLA6-R	CCACAGGATGTGGGCTACCTGTAC GCAGAGGTCTGAGAGCTCAGCAA	58	252	Bsp1407I	252	21+231
NTN5	NTN5-Fm NTN5-R	CCTCCCTCAGGCCCAGAGTCCAGG GTGGGGGTGTTTGTGAGCGGCG	58	200	FnuDII	19+181	200

Table 3 Primers for amplification of genes and their RFLP condition that were predicted by prediction programs (positive 5 in 6 programs)

Gene	Primer	Nucleotide sequence (5'→3')	Tm (°C)	Product size (bp)	Restriction Enzyme	PCR product (bp) after digestion	
						Major Allele	Minor allele
FAM131C	FAM131C-Fm FAM131C-R	GGTGCAGCTCCTCTTCCT GGCATAGGGTGGGTGAGT	54	238	PstI	238	22+216
DTX4	DTX4-F DTX4-R	CTGCTGGAAAGGAGGAGAGTAGGT TTTCTACTCTTGCTGACGCGGG	58	201	Cac8I	18+183	201
GLT1D1	GLT1D1-F GLT1D1-R	GTGAAGCATGAAGTCACTG CTCAATGCTCCTGTTCC	54	242	BsrI	22+220	242
FLT1	FLT1-F FLT1-R	GTTTTCTTCTCGATGACTTTGA TGAGCAGCTGGAGGGTAG	54	292	AciI	119+173	292
PCK2	PCK2-F PCK2-R	CACCCCTTAGATGGGACA GAGAAGGAGTTACAATCACCGT	59	239	AciI	178+61	239
ABCA9	ABCA9-Fm ABCA9-R	AGCTCTCCAAGGAGCAGCA TTGTACCACCTCTGATATAAGGCA	54	183	PvuII	21+162	183
RBBP8	RBBP8-F RBBP8-Rm	ATCAATCATCAGCATCACACAG CTTTTTGGACGAGGCCAAG	54	188	StyI	18+170	188
EEF2	EEF2-F EEF2-Rm	ACTGTCTCCGTCTGCACG CTTCCTCATCAACCTCATTGAAT	54	226	BsmI	226	23+203
P2RY11	P2RY11-Fm P2RY11-R	CTATGGGGAGGCCGCGTGTC GGTGGGAGAAGCTGAGTGTGGGC	68	221	Tsp45I	221	18+203
LAMA5	LAMA5-F LAMA5-R	CAGTACAGCTCGGGGAAAGCCTC AGCGCTGCCGGAGCTCG	68	300	BseNI	89+211	89+103+108
FAT1	FAT1-Fm FAT1-R	ACCACTGTAACAAGATATTCTGTA TGTTTGTCAACCTTCCCT	52	226	RsaI	226	24+202

Table 4 Primers for amplification of genes and their RFLP condition that were predicted by prediction programs (positive 4 in 6 programs)

Gene	Primer name	Primer sequence (5'→3')	Ta (°C)	Product size (bp)	Restriction Enzyme	PCR product (bp) after digestion	
						Major Allele	Minor allele
ACAP3	ACAP3-F ACAP3-R	GTGGGCGCCAGGGACTT CGTCTCACGCTGTGCCTC	61	243	BstUI	186+ 20+27	206+27
CCDC18	CCDC18-F CCDC18-R	GAGTTTTCATATCCTGCCT TGTTTCTAAACTACTTAATTTGCT	58	272	HindIII	148+92+ 30	178+92
TRIM45	TRIM45-F TRIM45-R	GATGCTCCTCAATGGCCTTAAT CCTGTGTCCTGTTCACCCTG	59	308	BanII	157+ 151	308
SRBD1	SRBD1-F SRBD1-Rm	CCCAAGAAAGTTACAAGTA AAATCAAGACTGAGACATAAC	51	260	MaeII	260	260+239 +21
FZD5	FZD5-Fm FZD5-R	GGCGCGGATGGGTCAGTCC GAGGAGCGGAGCGCTGCCA	64	228	Hpy188I	209+19	228
GJA10	GJA10-F GJA10-Rm	CCAGATGGAGAATCCAGAT AACCCATAGAGAATATATTGGACT	56	195	Tsp4CI	195	172+23
PLEKHA8	PLEKHA8-F PLEKHA8-R	TAGAAAAAGATGGTAATAGTATCC TTAAAAGTTTCCACAGAGACA	58	270	SmaI	186+84	270
PRMT8	PRMT8-Fm PRMT8-R	CTCAGGTGAACAGCCCGC ACTCAGGAACACTGGGAGCACT	57	316	MbiI	316	297+19
TGM1	TGM1-F TGM1-R	GCCCTTATCATTAGCTTCCTATC GCTGTGGCTGCTGTTCAT	59	283	HaeIII	186+127 +58	186+98
C14orf43	C14orf43-F C14orf43-R	GCCAGTGATGTAGGAACAGG GCAGCAACCCCAGGACT	61	296	Tsp4CI	250+46	46+ 86+
C15orf42	C15orf42-Fm C15orf42-R	AGTCCCCTGAAAAAGGACAT TAGCCCTCATCCTGAATTTC	54	259	NlaIII	237+22	259
PGLYRP2	PGLYRP2-F PGLYRP2-R	CTGGCTCAGCAAGTGGCTGA AGATGTCCAAGCTTCCTTGCCA		316	Hpy188I	141+70+ 105	211+105
PEPD	PEPD-F PEPD-R	ACAGCACTGTTTGGTCTGAT ACTGTGGAAGAGATTGAAGCA		221	AciI	221	172+49
HIPK4	HIPK4-F HIPK4-R	GCTCTGAGGAGCAGTTCA GCTCAGCAGACACTAATACACT	58	320	SmaI	194+126	320
FAM71E2	FAM71E2-F FAM71E2-R	TGGCGGTGCTGGAAATGATG CCCCGACAGAACTCACACCCA	58	172	MboI	130+42	101+42+ 29
EVC2	EVC2-F EVC2-R	GTCTTCTGTCTGAAGTGTCCCT AAGCCCTCCCTGAGTCAC	58	231	HpaII	186+ 127+58	186+98
CENPE	CENPE-Fm CENPE-R	GCACTTCTTTCCTCAAACACT CATCCCAGGCAGGACTG	59	264	TspRI	205+34+ 25	230+34

8.3 Restriction fragment length polymorphism (RFLP)

PCR products were digested by specific restriction enzymes and incubated at proper time and temperature depending on restriction enzyme. In electrophoresis step, two glass plates were used to prepare the polyacrylamide gel cassette. The glass plates were clean by 70% ethanol before use. Two glass plates were assembled by padding with two spacers. The gel was prepared by mixing 2.5 ml of 40% polyacrylamide (49:1), 2 ml of 5X Tris-borate-EDTA (TBE) buffer, 0.5 ml glycerol, 4.93 ml of deionized water, 70 ul of 10% ammonium persulfate, and 4 ul of tetramethylethylenediamine (TEMED). Then, mixing solution was loaded into space between the glass plates until completely filling the space and comb was then insert into the opening between the glass plates. Once the gel was polymerized, the comb can be gently removed and then each wells of polyacrylamide gel were washed by deionized water. Subsequently, gel cassette was placed into the electrophoresis running chamber and 1X TBE buffer was added into the chamber. Then 5 ul of PCR products were mixed with 1 ul of 6X gel-loading buffer and then loaded into each well (one well/sample). One microliters of 100 bp DNA Ladder were loaded into one well of acrylamide gel as a standard-size markers. The electrophoresis running condition was set at 110 Volt, 400 mA for 2 h 30 min.

8.4 Silver staining

Fifty milliliters of 40% methanol and 160 mM HNO3 were used to fix acrylamide gel for 10 minutes and 6 minutes, respectively. Then, acrylamide gel was washed with deionized water twice, 5 minutes each. Subsequently, 50 ml of silver stain solution (0.2% W/V AgNO3) was added to stain the acrylamide gel for 20 minutes in shaker. The acrylamide gel was then washed twice again with deionized water, 5 minutes each. Fifty milliliters of developer (3% W/V Na_2CO_3 and formaldehyde) were added and the gel was shaken gently. Then, 70 ml of stop solution (5% W/V citric acid) were added and the gel was shaken gently until the DNA bands were appeared clearly. After shaking for 5 minutes, the acrylamide gel was washed with deionized water. Finally, the acrylamide gel was dried on a sheet of glass paper at room temperature for 2 days.

9. Validation of selected variants by Sanger sequencing

Sanger sequencing of PCR products was performed to validate each variants. DNA sequences were examined by using BioEdit version 7.0.9.0 or Chromas version1.44.

10. Additional genotyping in non-diabetic and other T2D subjects

Segregated variation was analyzed further in non-diabetic controls (n=400) and in other T2D probands (n=1000) by using PCR-RFLP method to ascertain that it is a rare variant and to explore its role in other T2D probands.

11. In silico analysis

The putative variants that passed filtration criteria were tested by considering the impact of each variant on protein function. Six programs software: Polyphen-2 SIFT, VarioWatch, Mutation Taster, Provean and SNP&GO to determine possible changes in the protein structure that might affect the phenotype. Clustal X1.83 software was used to compare the human candidate protein sequence with the orthologs from *Pan troglodytes*, *Macaca mulatta*, *Mus musculus*, *Canis lupus*, *Bos Taurus, Gallus gallus*, *Rattus norvegicus*, *Danio rerio* and *Xenopus tropicalis* and to examine evolutionary conservation and structural prediction for this protein.

12. Statistical analysis

Comparisons of clinical characteristics of diabetic patients carrying variant and non-diabetic controls not carrying the same variant were analyzed by SPSS software (version 17; SPSS Inc., Chicago, IL). The significant level was determined at P-value less than 0.05, by t-test or Mann–Whitney U-test.

Results

1. Type 2 Diabetes Family selection

To study genetics of Thai type 2 diabetes family, the members of T2D were recruited. According information, the pedigree of T2D family is a three-generation with autosomal dominant inheritance as illustrated in Figure 4. Peripheral blood from twenty seven family member were collected including nine with diabetes, ten with increased risk of diabetes and eight with non-diabetic subjects. Detail clinical information of family members was shown in Table 5. The proband woman, 83 year old (I-1) presented diabetes with age 61 year old. The other affected members of the family had 45 to 78 years old at diagnosis.

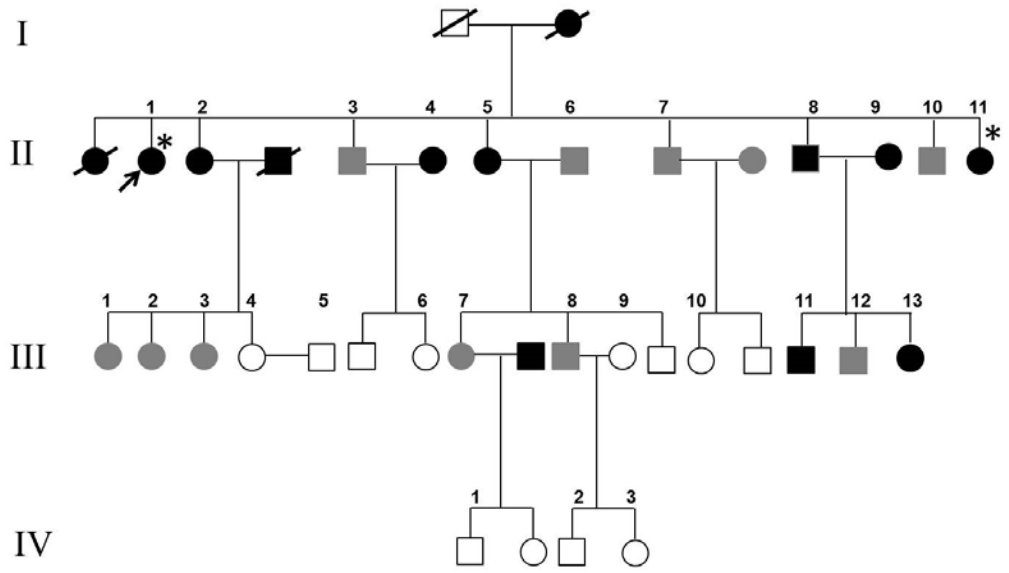

Figure 4 **Pedigree of type 2 diabetes family.**

Roman number on the left side indicated the pedigree generation. Number above represented individual ID within generation. Squares and circles represented by males and females, respectively. The black, gray and open symbols represented diabetes, increased risk of diabetes: HbA1c ≥5.7%-6.4%, (impaired fasting glucose (IFG) and non-diabetic subjects, respectively. Asterisk indicated sample which were investigated by exome sequencing. Proband was indicated by arrow.

Table 5 Clinical characteristics of members in type 2 diabetes family studies.

Gene No	ID	DM Status	Age (year)	Age Dx (year)	BMI	Waist	Sys BP	Dias BP	FPG	HbA1c	TC	TG	LDL	HDL
I	1	Affected	83	61	21.85	71	150	80.00	272.0	8.8	194.0	126.0	126.0	42.0
I	2	Affected	83	70	20.24	77	110	86	123	6.6	183	91	101.8	63
I	3	Affected	81	81	22.49	89	139	75	94	6.1	191	73	117.4	59
I	4	Affected	70	70	39.45	116	155	89	135	6.6	178	98	104.4	54
I	5	Affected	78	78	21.50	78	126	63	89	6.5	177	52	62.6	104
I	6	Affected	77	77	24.30	99	149	85	94	5.7	177	79	104.2	57
I	7	Affected	76	76	24.73	96	136	69	104	6	145	85	72	56
I	8	Affected	76	50	16.14	64	125	71	150	7.3	93	48	42.4	41
I	9	Affected	75	60	22.35	92	138	56	113	7	156	123	72.4	59
I	10	Affected	72	72	17.26	64	161	95	80	5.9	211	75	135	61
I	11	Affected	62	55	21.11	80	132	78						
II	1	Affected	54	54	29.48	91	158	117	89	6.4	335	123	253.4	57
II	2	Affected	53	53	21.94	84	129	86	81	5.9	180	77	87.6	77
II	3	Affected	50	50	28.07	95	106	68	86	6.1	237	94	138.2	80
II	4	Unaffected	44	44	30.85	102	135	95	91	5.6	219	88	144.4	57
II	5	Unaffected	44	44	31.83	111	133	88	87	5.6				
II	6	Unaffected	49	49	18.71	77	125	82	77	5.4	192	44	126.2	57
II	7	Affected	50	50	21.42	85	157	102	116	5.9	254	345	136	49
II	8	Affected	48	48	26.37	92	113	84	109	5.8	232	191	155.8	38
II	9	Unaffected	46	46	27.47	83	117	76	93	5.5	280	80	193	71
II	10	Unaffected	38	38	18.02	62	93	55	75	5.6	241	47	151.6	80
II	11	Affected	52	45	30.48	100	145	88	142	7.3	233	91	151.8	63
II	12	Affected	50	50	22.13	79	137	81	98	5.8	194	91	114.8	61
II	13	Affected	46	46	25.65	84	103	73	106	6.6	189	56	116.8	61
III	1	Unaffected	16	16	20.39	82	112	67	79	5.2	157	58	86.4	49
III	2	Unaffected	17	17	21.12	82	104	77	89	5.5				
III	3	Unaffected	10	10	16.46	59	100	63	81	5.6	195	74	109.2	71

2. Identification of causative gene causing T2D by exome sequencing

Two DNA samples from this family including proband (II-1) and her younger sister (II-11) were performed for exome sequencing. The average of 97% (more than 10X) of target regions are covered with mean read depth of 89%. Approximately 69,083 per individual were identified and of 19,912 were coding variants. The exome sequencing characteristic of two DNA samples were summarized in Table 6. Exome sequencing standard analyses include chromosome name, chromosome start position, chromosome end position, reference allele, alternative allele, heterozygosity status, SNPs quality score calculated from Phred-scaled likelihood, total read depth and read depth of alternative allele were reported in *_SNP_Indel.xls file. Moreover, advanced analyses comprise region of SNP such as exonic splicing, 5UTR or 3UTR, gene symbols, the functional consequences of the variant such as nonsynonymous SNPs, synonymous SNPs or frameshift insertion, annotation, dbSNP135 and allele frequency were also reported in *_SNP_Indel.xls file. The contents in*_SNP_Indel.xls file are shown in Figure 4.

Table 6 Identified variants through the exome sequencing analyses

Exome sequencing information	II-1	II-11
Total reads	67,805,160	70,944,168
Total yield (bp)	6,848,321,160	7,165,360,968
Read length (bp)	101	101
Target regions (bp)	51,189,318	51,189,318
Average throughput depth of target regions	133.8	140
Initial mappable reads (mapped to human genome)	67,608,070	70,759,488
% Initial mappable reads (out of total reads)	99.70%	99.70%
Non-redundant reads (de-duplicated by Picard tools)	65,258,288	68,158,202
% Non-redundant reads (out of initial mappable reads)	96.50%	96.30%
Non-redundant unique reads (uniquely mapped to human genome)	64,177,915	67,049,624
% Non-redundant unique reads (out of non-redundant reads)	98.30%	98.40%
On-target reads (mapped to target regions)	51,962,697	54,260,253
% On-target reads (out of non-redundant unique reads)	81.00%	80.90%
% Coverage of target regions (more than 1X)	99.00%	98.90%
Number of on-target genotypes (more than 1X)	50,658,515	50,624,296
% Coverage of target regions (more than 10X)	97.80%	97.60%
Number of on-target genotypes (more than 10X)	50,056,978	49,982,062
Mean read depth of target regions	87.4	91.2
Number of SNPs	68,984	69,182
Number of coding SNPs	19,919	19,912
Number of synonymous SNPs	10,507	10,511
Number of nonsynonymous SNPs	8,937	8,892
Number of Indels	6,659	6,702
Number of coding Indels	469	467

#chr_name	chr_start	chr_end	ref_base	alt_base	hom_het	snp_quality	tot_depth	alt_depth	region	gene	change	annotation	dbSNP135_full	dbSNP135_common	1000G_2010Nov_allele_freq
chr01	762273	762273	G	A	hom	176	101	101	ncRNA_exonic	LINC00115	.	.	rs3115849	rs3115849	0.555
chr01	866319	866319	G	A	hom	101	15	15	intronic	SAMD11	.	.	rs9988021	rs9988021	0.947
chr01	876499	876499	A	G	hom	126	15	15	intronic	SAMD11	.	.	rs4372192	rs4372192	0.894
chr01	877715	877715	C	G	hom	182	37	37	intronic	SAMD11	.	.	rs6605066	rs6605066	0.903
chr01	877831	877831	T	C	hom	214	38	38	exonic	SAMD11	nonsynon	SAMD11:NM	rs6672356	.	1.000
chr01	878021	878021	C	T	het	10.4	10	2	exonic	SAMD11	nonsynon	SAMD11:NM	.	.	.
chr01	879317	879317	C	T	het	225	129	59	exonic	SAMD11	synonym	SAMD11:NM	rs7523549	rs7523549	0.077
chr01	879676	879676	G	A	hom	15.4	5	5	UTR3	NOC2L,SA	.	.	rs6605067	rs6605067	0.878
chr01	879687	879687	T	C	hom	25.3	5	5	UTR3	NOC2L,SA	.	.	rs2839	rs2839	0.910
chr01	880238	880238	A	G	hom	222	71	71	intronic	NOC2L	.	.	rs3748592	rs3748592	0.924
chr01	880390	880390	C	A	het	19.1	6	3	intronic	NOC2L	.	.	rs3748593	rs3748593	0.076
chr01	881627	881627	G	A	het	48	32	15	exonic	NOC2L	synonym	NOC2L:NM_	rs2272757	rs2272757	0.441
chr01	883625	883625	A	G	hom	165	33	33	intronic	NOC2L	.	.	rs4970378	rs4970378	1.000
chr01	884091	884091	C	G	het	15.1	60	16	intronic	NOC2L	.	.	rs7522415	rs7522415	0.610
chr01	884101	884101	A	C	het	39	42	27	intronic	NOC2L	.	.	rs4970455	.	0.200
chr01	886788	886788	G	A	het	6.2	13	5	intronic	NOC2L	.	.	rs10465242	rs10465242	0.717
chr01	887560	887560	A	C	hom	169	39	38	intronic	NOC2L	.	.	rs3748595	rs3748595	0.901
chr01	888639	888639	T	C	hom	222	179	179	exonic	NOC2L	synonym	NOC2L:NM_	rs3748596	rs3748596	0.918
chr01	888659	888659	T	C	hom	222	189	189	exonic	NOC2L	nonsynon	NOC2L:NM_	rs3748597	rs3748597	0.917

Figure 5 Example of exome sequencing output for analysis *_SNP_Indel.xls file

3. Candidate variants selection

Since all variants were approximately 68,583 variants which were identified in 2 affected individual (I-1 and I-11), the filtering criterion steps were performed to identify candidate variants as shown in Table 7. Variants that were shared by two affected individuals and were not present in sixteen in-house control subjects, showing (68,583 and 25,388 variants), respectively. Under assumption autosomal dominant inheritance, so the analysis focused on heterozygous variants (showing 11,922 variants) with possible protein altering effects (nonsense, nonsynonymous, and frameshift indel). The variants that had an allele frequency >1% in either the 1000 Genomes database or SNP135 database were exclude, resulting in (2,280 and 140 variants), respectively. Subsequently, 140 variants were predicted to be damaging by six different bioinformatics algorithms (Polyphen, Sift, Matation taster, Variowatch, Provean and SNP&GO). Finally, the possible candidate variants were passed filtration criteria to examine the segregation within family P285, including five of frameshift, forty of non-synonymous variants as shown in Table 8-11.

Table 7 **Filtering criterion for identification of causative variants.**

Filtering criterion	Number of variants following each filtering criterion
Total variants were identified by exome sequencing in proband and affected members	68,583
Excluded variants sharing in 16 in house controls	25,388
Selected Heterozygous variants (nonsense, nonsynonymous and frameshift indel)	11,922
Exclude variants identifying in 1000 Genome database (MAF> 1%)	2,280
Excluded variants identifying in dbSNP135 (MAF> 1%)	140
Frameshift	5
Predicted impact on protein function by in siligo programs, including Polyphen2, Sift, Matation taster, Variowatch, Provean and SNP&GO	11[a]
	11[b]
	18[c]

[a] represented candidate variants that displayed 6 out of 6 programs.

[b] represented candidate variants that displayed 5 out of 6 programs.

[c] represented candidate variants that displayed 4 out of 6 programs.

Table 8 **List of genes that showed frameshift variants.**

Type	Gene	Chromosome	Position	Variation
frameshift	ANKRD36	2	97877470	CT/-
frameshift	ZRANB3	2	136148370	ACTT/-
frameshift	OR6B2	2	240969009	CA/-
frameshift	OR2C1	16	3406758	T/-
frameshift	FAM129C	19	17653040	AGCACCGTTTGGC/-

Table 9 List of genes that passed the prediction of impact on protein function by in *silico* programs (6 out of 6 programs).

Gene	Chr.	Position	Ref.base	Alt.base	Novel/rs.no	Prediction programs					
						Polyphen2	SIFT	Variowatch	Mutation taster	Provean	SNP&GO
MYOM3	chr01	24392484	T	C	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
GJA5	chr01	147231325	C	T	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
SPTA1	chr01	158639198	C	A	Novel	possibly damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
CACNA1S	chr01	201046130	C	G	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
IRX1	chr05	3599632	G	C	Novel	possibly damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
USP6NL	chr10	11505473	G	A	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
SPINT1	chr15	41148176	C	T	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
MYH4	chr17	10359148	T	C	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
PNPLA6	chr19	7625552	G	A	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
NTN5	chr19	49167929	G	A	rs376906018	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease
CBR3	chr21	37518391	A	T	Novel	probably damaging	Damaging	Damaging	disease_causing	Deleterious	Disease

Table 10 List of genes that passed the prediction of impact on protein function by in *silico* programs (5 out of 6 programs).

Gene	Chr.	Position	Ref.base	Alt.base	Novel/rs.no	Prediction programs					
						Polyphen2	SIFT	Variowatch	Mutation taster	Provean	SNP&GO
FAM131C	chr01	16386043	C	G	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
PCDH15	chr10	55945018	G	C	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
DTX4	chr11	58956712	C	G	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
GLT1D1	chr12	129442166	G	A	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	N/A
FLT1	chr13	28913359	G	A	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
PCK2	chr14	24567423	G	A	Novel	possibly damaging	Damaging	High	disease_causing	Deleterious	Neutral
ABCA9	chr17	66972125	C	A	Novel	possibly damaging	Damaging	High	polymorphism	Deleterious	Disease
RBBP8	chr18	20606129	C	G	Novel	possibly damaging	Damaging	High	disease_causing	Deleterious	Neutral
EEF2	chr19	3983194	G	C	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
P2RY11	chr19	10224597	G	A	Novel	probably damaging	Damaging	High	polymorphism	Deleterious	Disease
LAMA5	chr20	60902423	G	A	Novel	probably damaging	Damaging	High	disease_causing	Deleterious	Neutral
FAT1	chr04	187541300	T	C	Novel	possibly damaging	Damaging	High	disease_causing	Deleterious	N/A

Table 11 List of genes that passed the prediction of impact on protein function by in *silico* programs (4 out of 6 programs).

Gene	Chr.	Position	Ref.base	Alt.base	Novel/rs.no	Prediction programs					
						Polyphen2	SIFT	Variowatch	Mutation taster	Provean	SNP&GO
ACAP3	chr01	1229056	C	T	Novel	probably damaging	Damaging	High	polymorphism	Deleterious	Neutral
ATP13A2	chr01	17318282	G	A	rs200393545	possibly damaging	Damaging	High	polymorphism	Deleterious	Neutral
CCDC18	chr01	93677666	T	C	Novel	N/A	Damaging	High	disease_causing	Deleterious	Neutral
TRIM45	chr01	117661157	G	T	Novel	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
SRBD1	chr02	45826667	G	C	rs369884817	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
FZD5	chr02	208633457	G	A	rs575951396	possibly damaging	Damaging	High	polymorphism	Neutral	Neutral
GJA10	chr06	90604705	T	C	rs375382627	benign	Tolerated	High	disease_causing	Deleterious	Disease
PLEKHA8	chr07	30100511	C	G	Novel	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
GLB1L3	chr11	134179637	C	G	Novel	benign	Damaging	High	polymorphism	Deleterious	Disease
PRMT8	chr12	3649785	C	T	Novel	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
TGM1	chr14	24731290	G	A	Novel	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
C14orf43	chr14	74205656	C	G	rs540101166	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
C15orf42	chr15	90161441	G	A	Novel	possibly damaging	Damaging	High	disease_causing	Neutral	Neutral
GADD45GIP1	chr19	13067794	G	A	rs200567073	probably damaging	Damaging	High	polymorphism	Deleterious	Neutral
PGLYRP2	chr19	15586715	G	A	Novel	probably damaging	Damaging	High	polymorphism	Deleterious	Neutral
PEPD	chr19	33878269	G	C	rs 370067875	possibly damaging	Damaging	High	disease_causing	Neutral	Neutral
HIPK4	chr19	40885538	G	A	Novel	probably damaging	Damaging	High	disease_causing	Neutral	Neutral
FAM71E2	chr19	55872113	G	A	Novel	probably damaging	Damaging	High	polymorphism	Deleterious	Neutral
PTGIS	chr20	48140635	A	G	Novel	possibly damaging	Tolerated	High	disease_causing	Deleterious	Neutral
EVC2	chr04	5687147	C	G	rs200393545	probably damaging	Damaging	High	polymorphism	Deleterious	Neutral
CENPE	chr04	104044227	G	C	Novel	probably damaging	Damaging	High	polymorphism	Neutral	Disease

4. Segregation of candidate variants with diabetes in family

Selected candidate variants were passed prediction of impact on protein function (Table 8-11) at least 4 out of 6 programs and frameshift to chosen for testing the segregation with diabetes in this family by using PCR-RFLP method following validated each variant by Sanger sequencing (Figure 6-80). All of variants were not segregated with diabetes status excepted for *PRMT8*:c.C89T:p.P30L to display a partial segregation with the disease in this family. This variant presented in the proband (I-1) and segregated to fifteen affected (II-2, II-3, II-5, II-7, II-8, II-9, II-10, II-11, III-1, III-2, III-3, III-7, III-8, III-10,III-11 and III-13) and three unaffected individuals (III-4,III-9 and IV-1) (Figure 75-76). In addition, this variant was not observed in any 400 non-diabetic controls, demonstrating that it is rare variant for general population (Table 12). However, the other 1000 T2D patients were not detected.

Conservation of amino acid alignment further evaluated *PRMT8*:c.C89T:p.P30L in other *PRMT8* orthologs, including *Homo sapient, Pan troglodytes*, *Macaca mulatta*, *Mus musculus*, *Canis lupus*, *Bos Taurus, Gallus gallus*, *Rattus norvegicus*, *Danio rerio* and *Xenopus tropicalis* as shown in Figure 81. This variant is fully conserved in mammalian and vertebrate *PRMT8* orthologs.

5. Analysis of clinical characteristic of diabetic patients and non-diabetic family members

Clinical characteristics including age, age at diagnosis, BMI, waist circumference , waist hip ratio, systolic blood pressure, diastolic blood pressure, FPG, HbA1c, total cholesterol, triglyceride, LDL and HDL did not show significant difference between affected and unaffected groups as shown in Table 13.

Analysis of impact of variant on clinical characteristic was also investigated. There were 19 affected family members including fourteen of them carried heterozygous variant (C/T) while five members carried homozygous variant (C/C) of *PRMT8*. All of clinical characteristics were not different among affected who carried C/T genotype or C/C genotype. For unaffected members, there were no difference between the groups were observed (Table14).

6. Analysis of frameshift variants

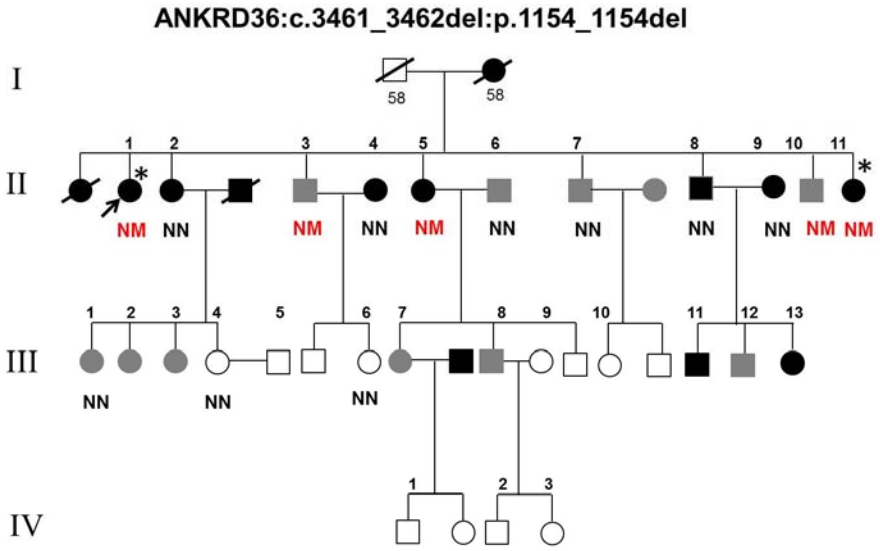

Figure 6 Segregation of ANKRD36:c3461_3462del:p.1154_1154del with diabetes in family P285

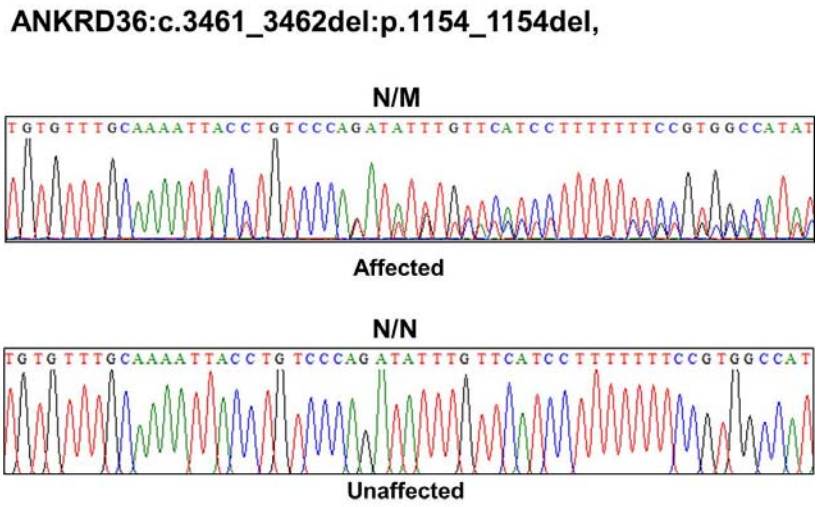

Figure 7 Sequence chromatogram of ANKRD36:c3461_3462del:p.1154_1154del was confirmed homozygous variant (N/N) in affected (upper panel) and heterozygous variant (N/M) in unaffected (lower panel).

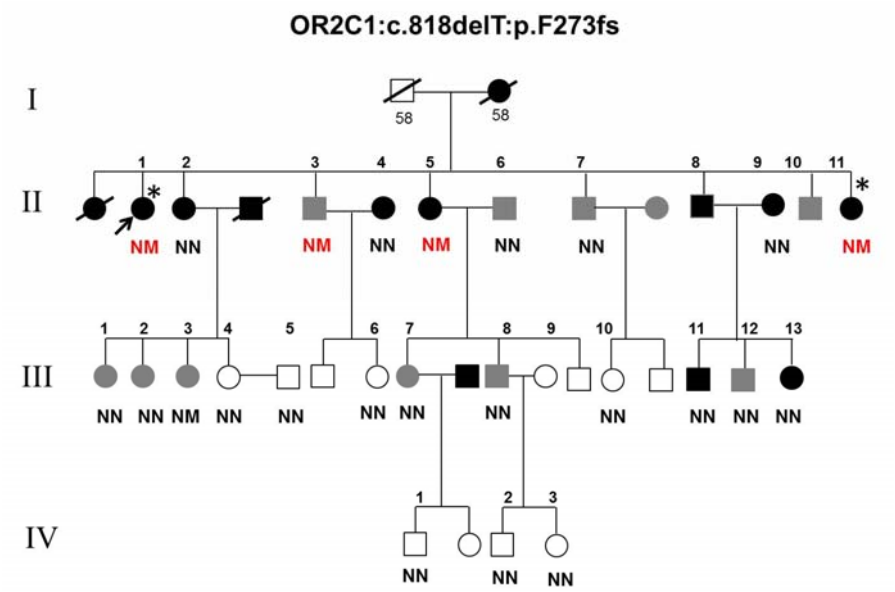

Figure 8 **Segregation of OR2C1:c.818delT:p.F273fs with diabetes in family P285.**

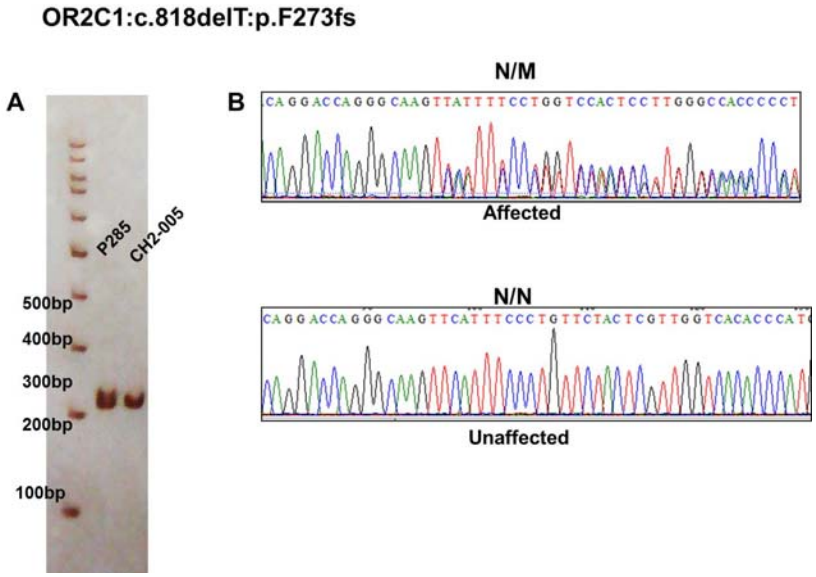

Figure 9 **Sequence chromatogram result of OR2C1:c.818delT:p.F273fs.**

Polyacrylamide gel electrophoresis of PCR product of heteroduplex pattern for homozygous variant (N/N) in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

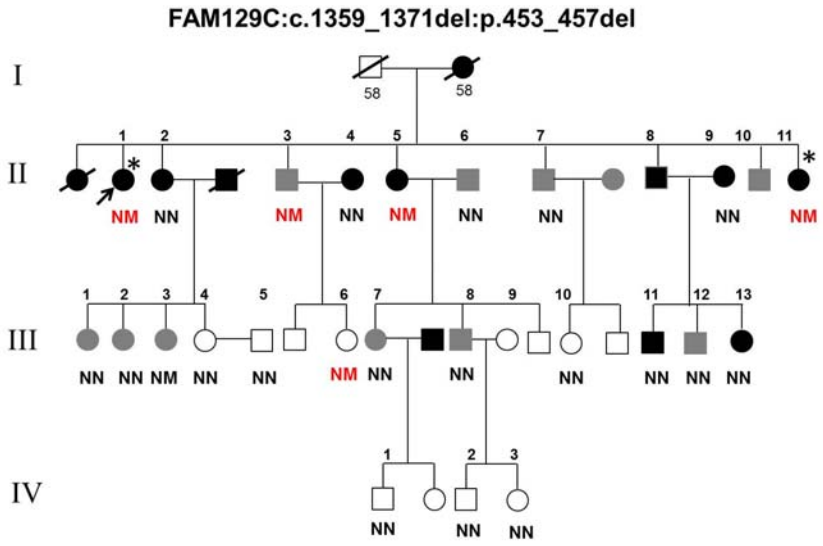

Figure 10 **Segregation of FAM129C:c.1359_1371del:p.453_457del with diabetes in family P285.**

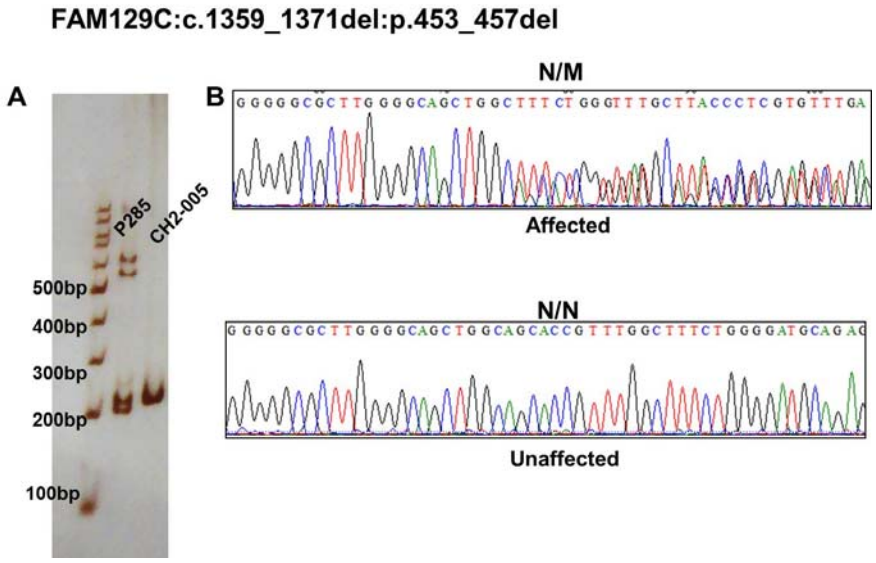

Figure 11 **Sequence chromatogram result ofFAM129C:c.1359_1371del:p.453_457del.** Polyacrylamide gel electrophoresis of PCR product of heteroduplex pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B)

.

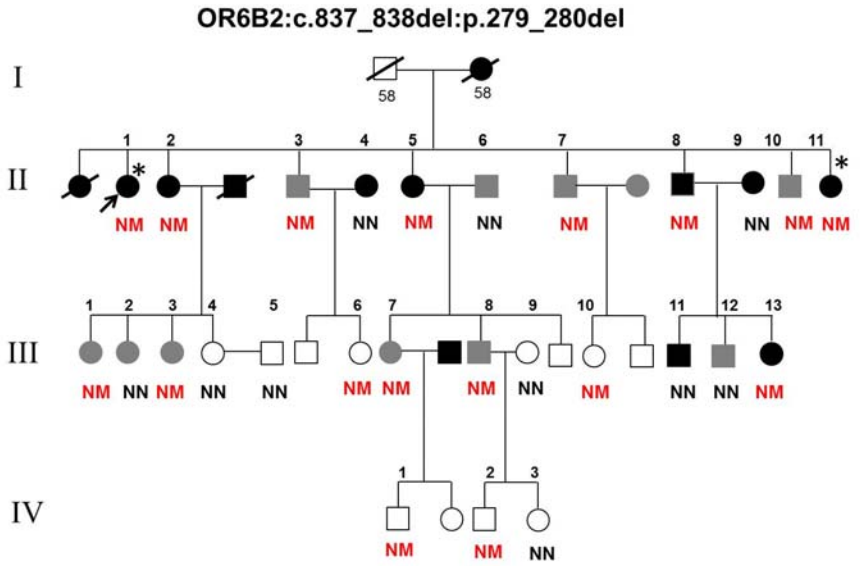

Figure 12 **Segregation of OR6B2:c837_838del:p.279_280del with diabetes in family P285.**

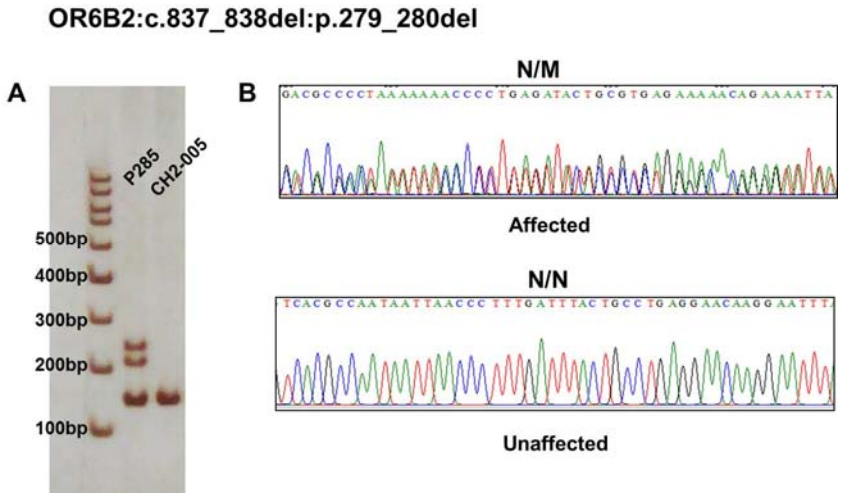

Figure 13 **Sequence chromatogram result of OR6B2:c837_838del:p.279_280del**
Polyacrylamide gel electrophoresis of PCR product of heteroduplex pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

1) Analysis of variants showed positive in silico program 6 out of 6 programs

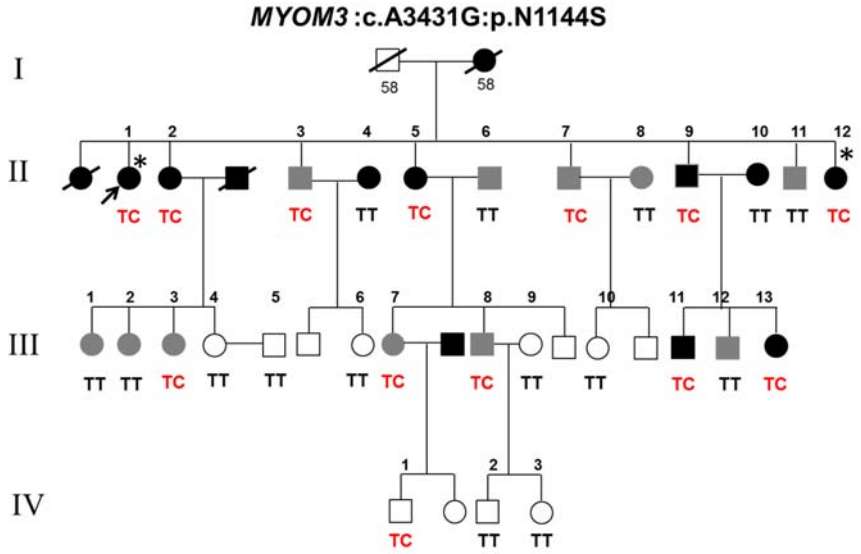

Figure 14 Segregation of MYOM3:c.A3431G:p.N1144S with diabetes in family P285.

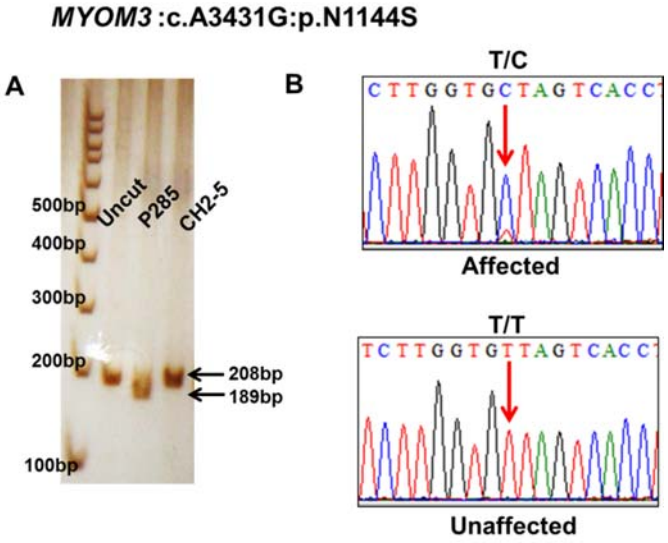

Figure 15 Sequence chromatogram result of MYOM3:c.A3431G:p.N1144S

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

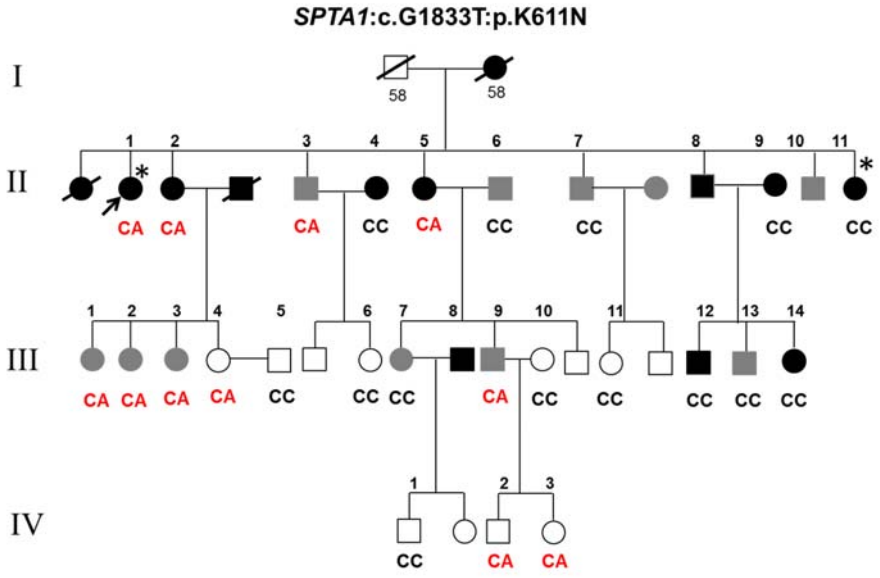

Figure 16 Segregation of SPTA1:c.G1833T:p.K611N with diabetes in family P285.

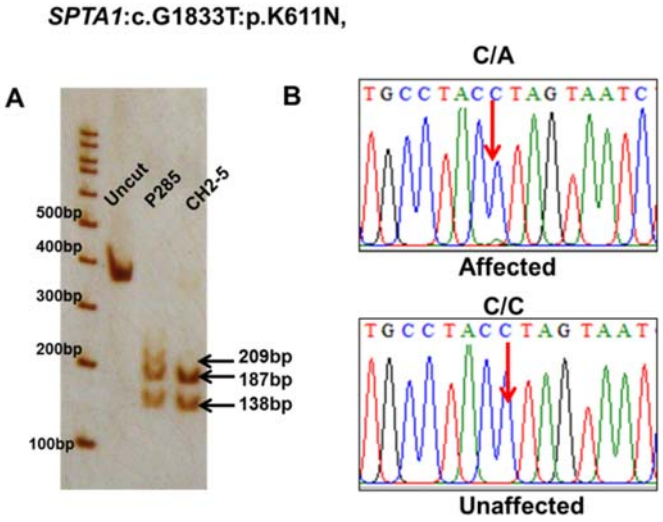

Figure 17 Sequence chromatogram result of SPTA1:c.G1833T:p.K611N

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

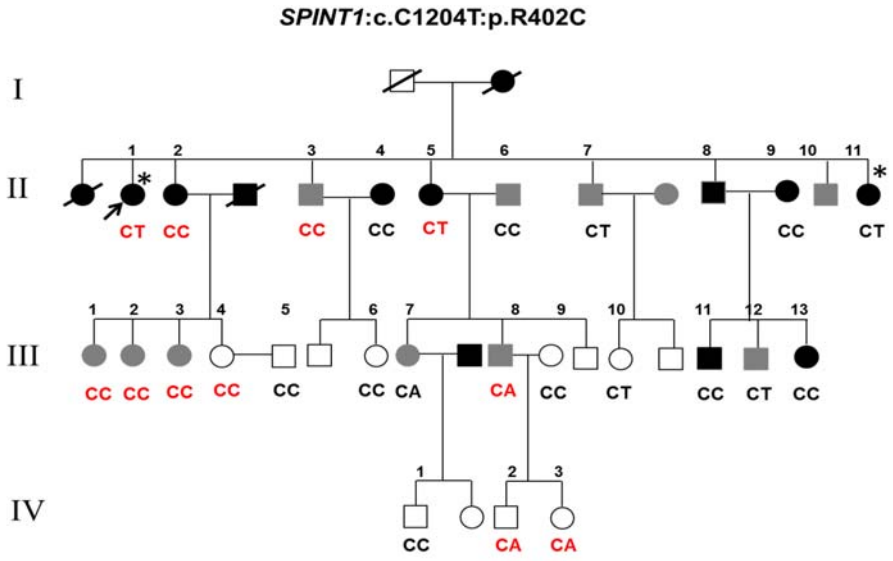

Figure 18 Segregation of SPINT1:c.C1204T:p.R402C with diabetes in family P285.

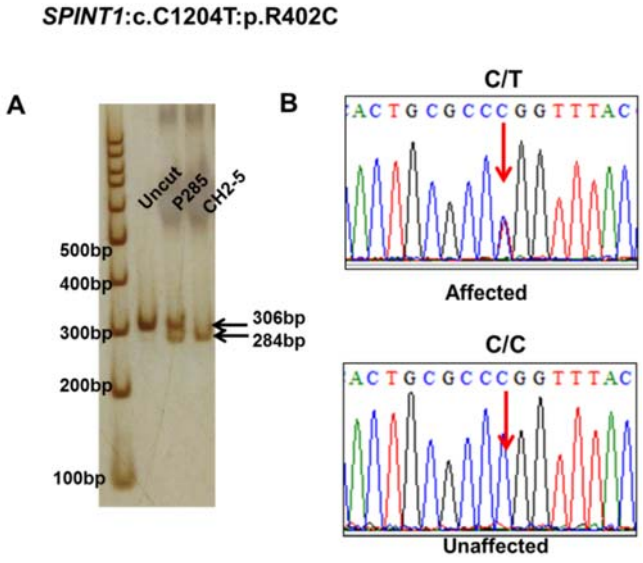

Figure 19 Sequence chromatogram result of SPTA1:c.G1833T:p.K611N

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

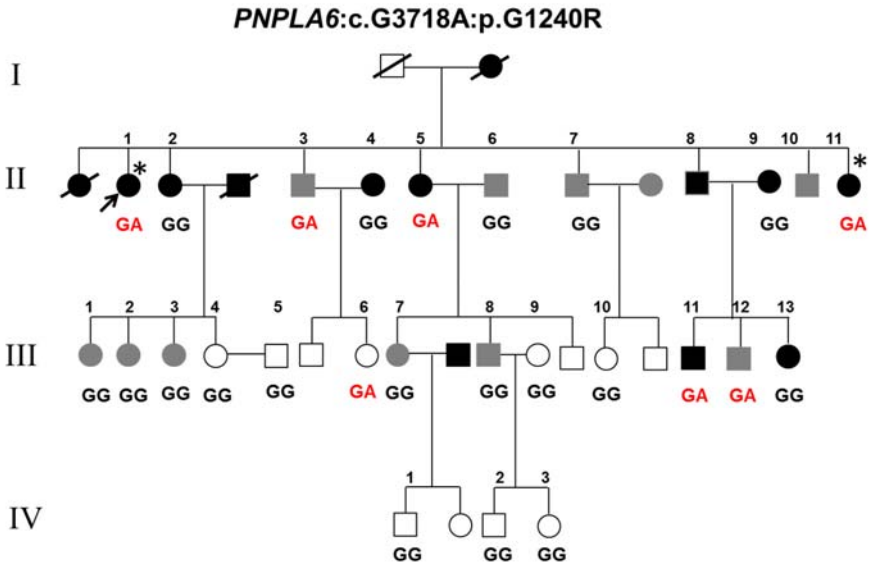

Figure 20 Segregation of PNPLA6:c.G3718A:p.G1240R with diabetes in family P285.

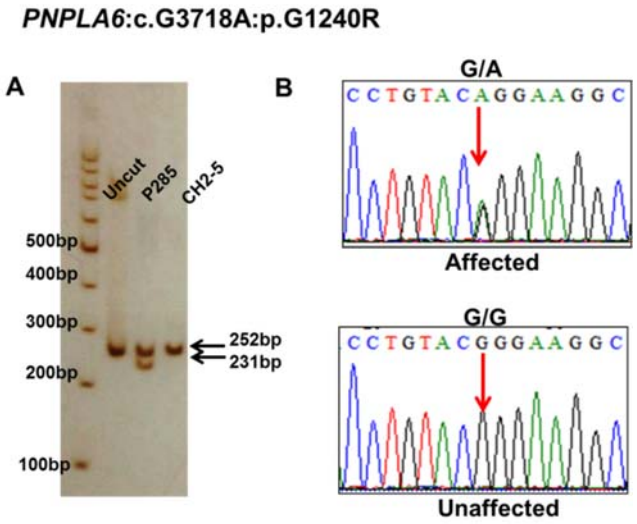

Figure 21 Sequence chromatogram result of PNPLA6:c.G3718A:p.G1240R. Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

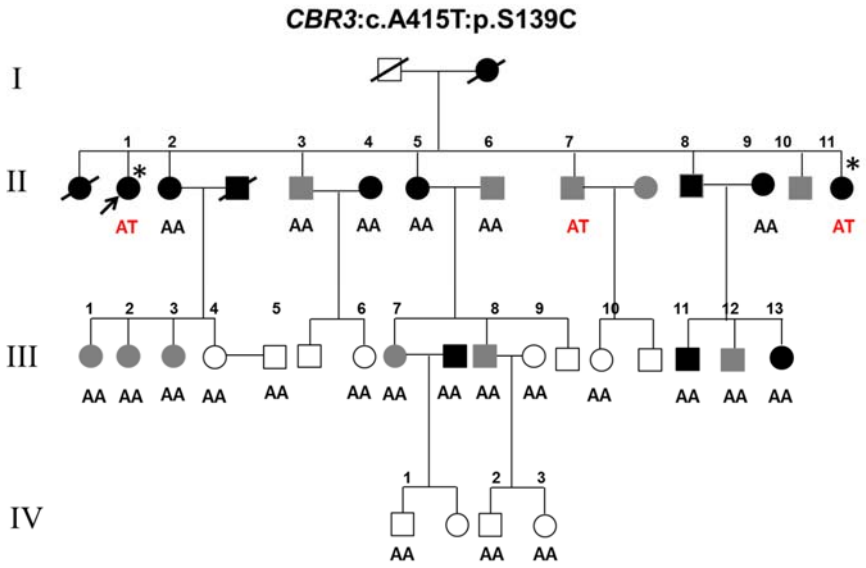

Figure 22 Segregation of CBR3:c.A415T:p.S139C with diabetes in family P285.

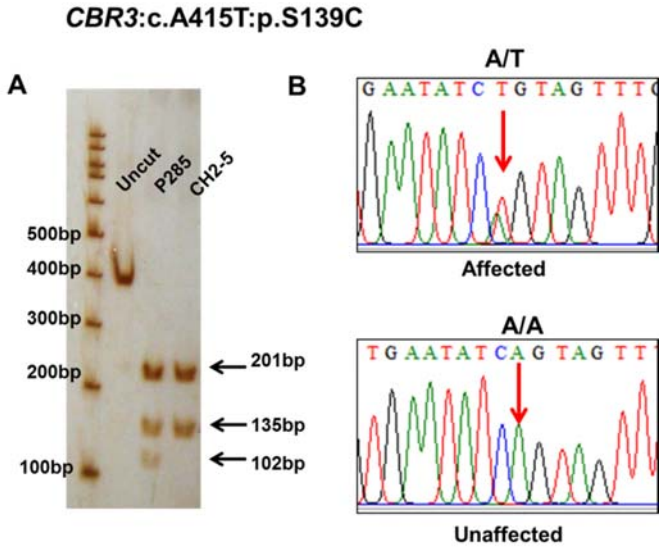

Figure 23 Sequence chromatogram result of CBR3:c.A415T:p.S139C.

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

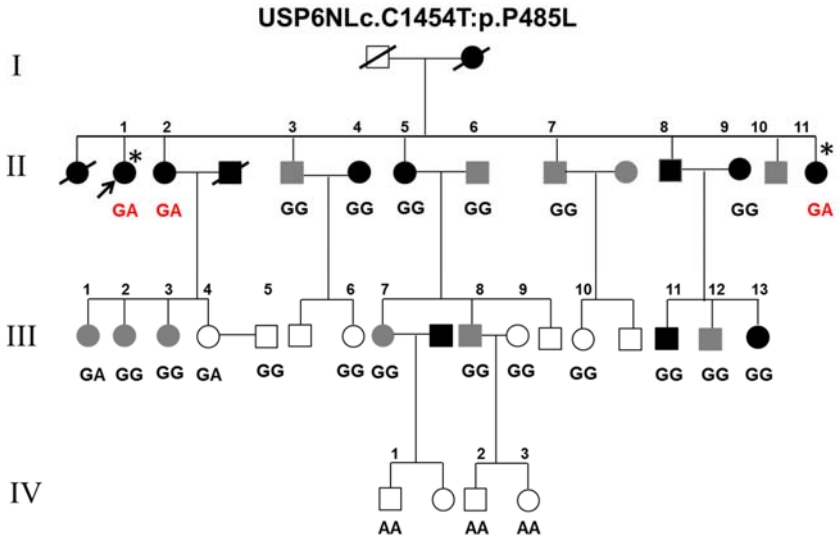

Figure 24 **Segregation of USP6NL:c.C1454T.p.P485L with diabetes in family P285.**

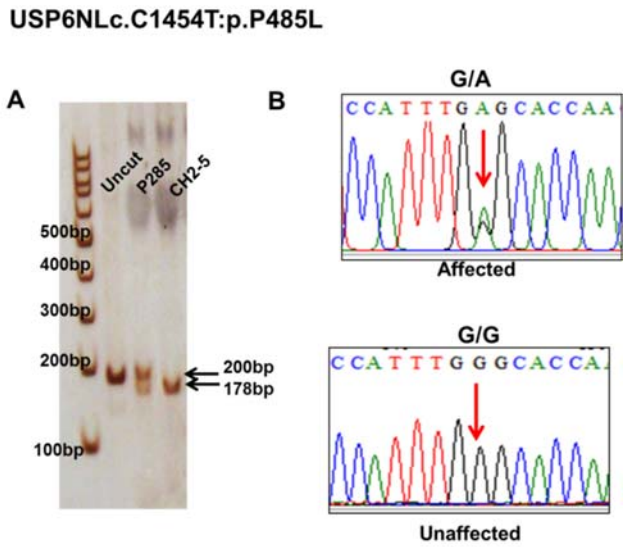

Figure 25 **Sequence chromatogram result of USP6NL:c.C1454T.p.P485L**

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

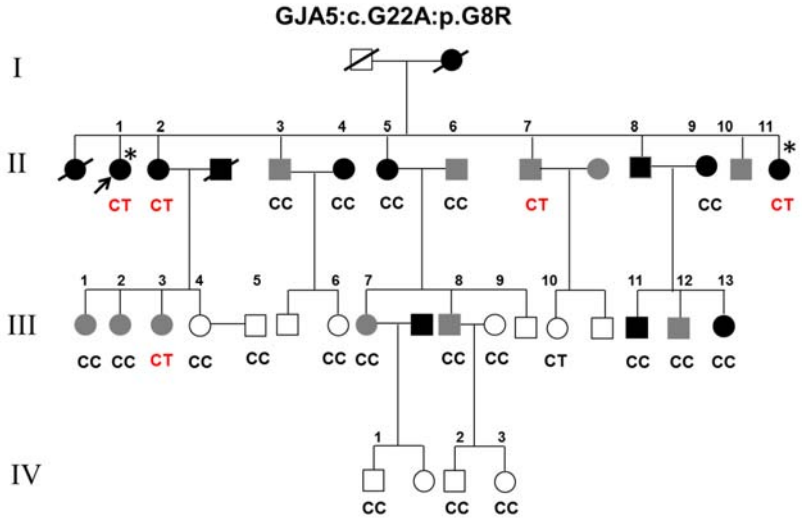

Figure 26 **Segregation of GJA5:c.G22A:p.G8R with diabetes in family P285.**

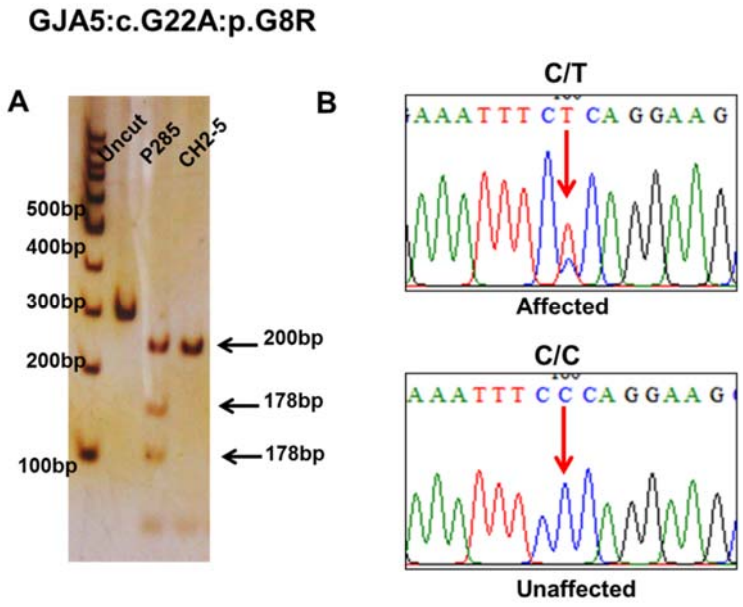

Figure 27 **Sequence chromatogram result of GJA5:c.G22A:p.G8R**

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

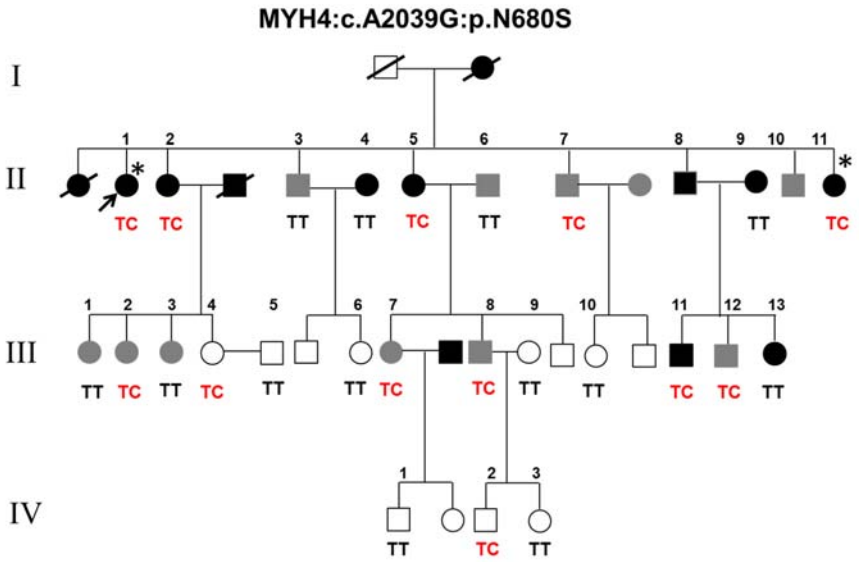

Figure 28 **Segregation of MYH4:c.A2039G:p.N680S with diabetes in family P285.**

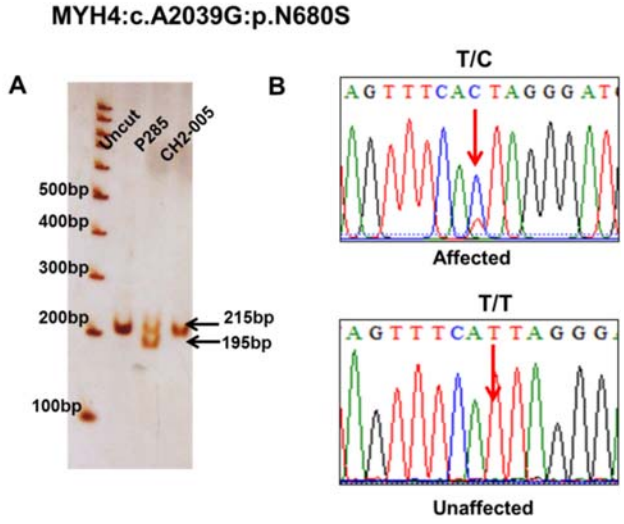

Figure 29 **Sequence chromatogram result of MYH4:c.A2039G:p.N680S.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

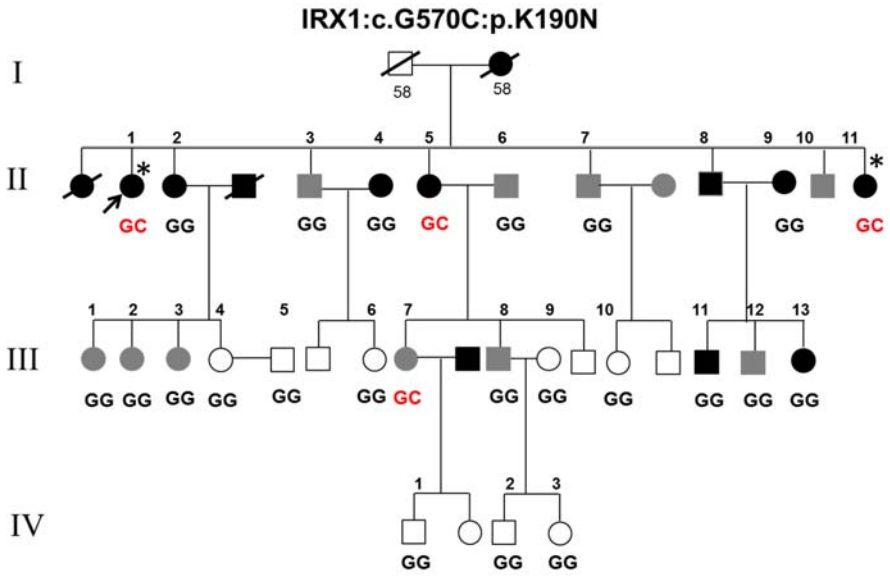

Figure 29 Segregation of IRX1:c.G570C:p.K190N with diabetes in family P285.

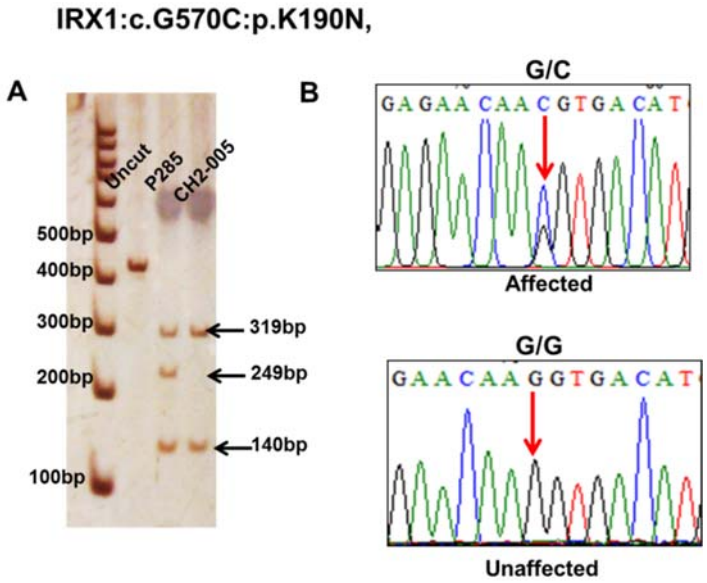

Figure 30 Sequence chromatogram result of IRX1:c.G570C:p.K190N.

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

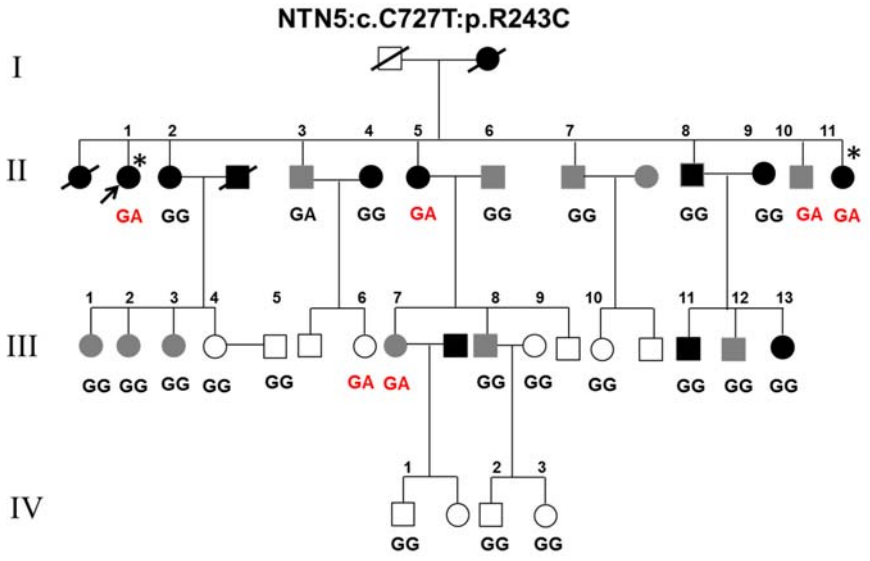

Figure 31 Segregation of NTN5:c.C727T:p.R243C with diabetes in family P285.

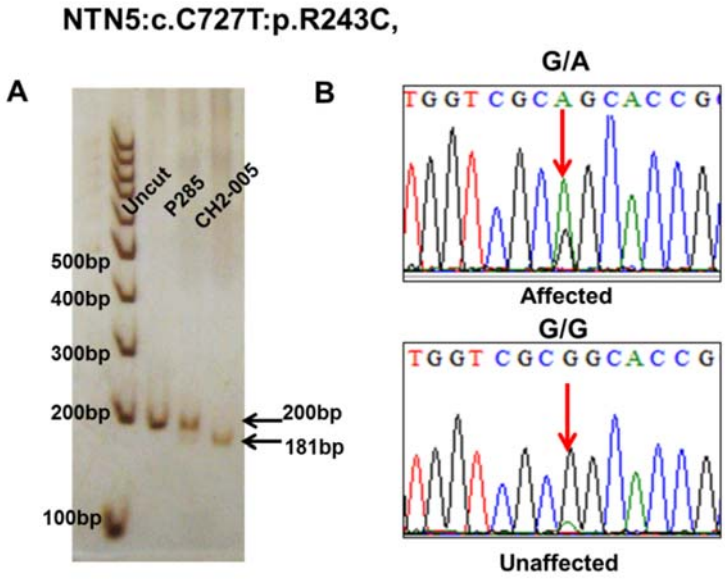

Figure 32 Sequence chromatogram result of NTN5:c.C727T:p.R243C

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

2) Analysis of variants showed positive in silico program showed 5 out of 6 programs

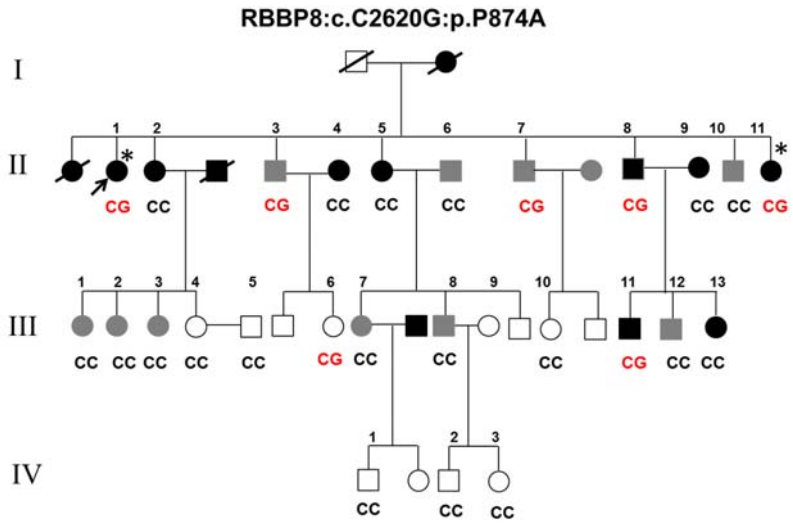

Figure 33 Segregation of RBBP8:c.C2620G:p.P874A with diabetes in family P285.

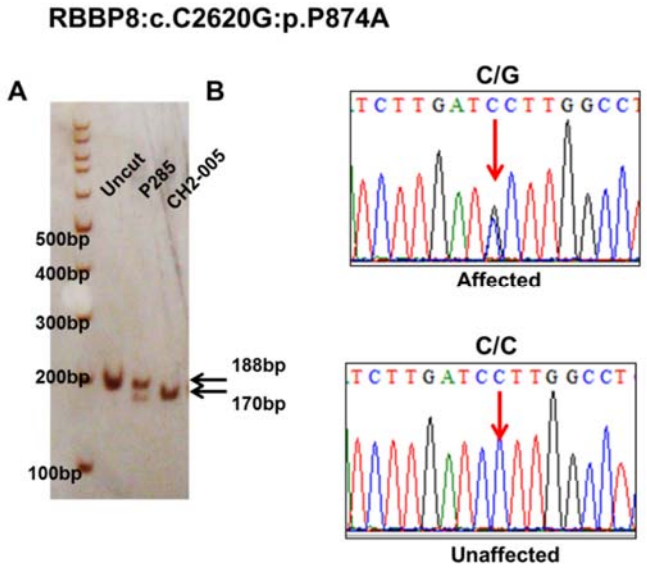

Figure 34 **Sequence chromatogram result of RBBP8:c.C2620G:p.P874A.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

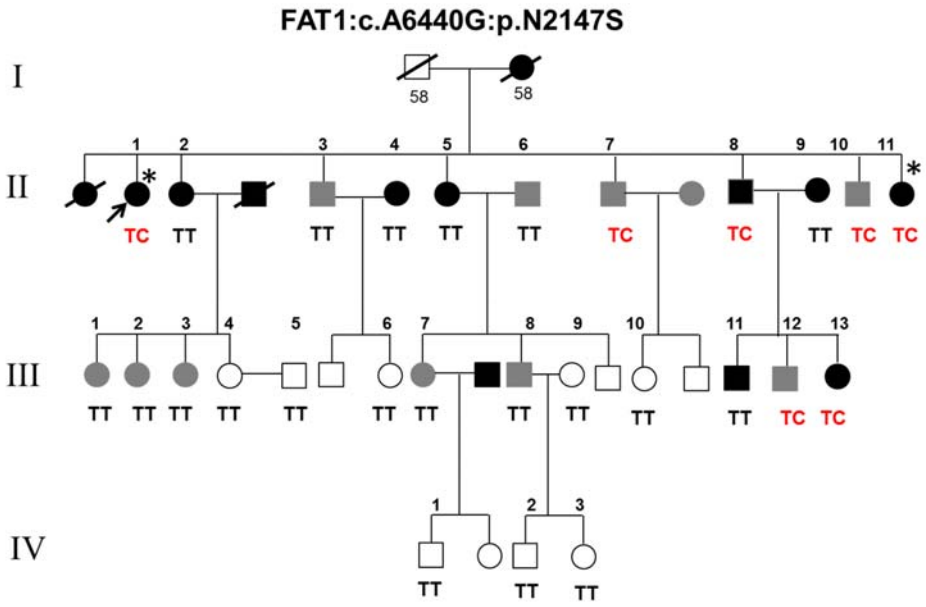

Figure 35 Segregation of FAT1:c.A6440G:p.N2147S with diabetes in family P285.

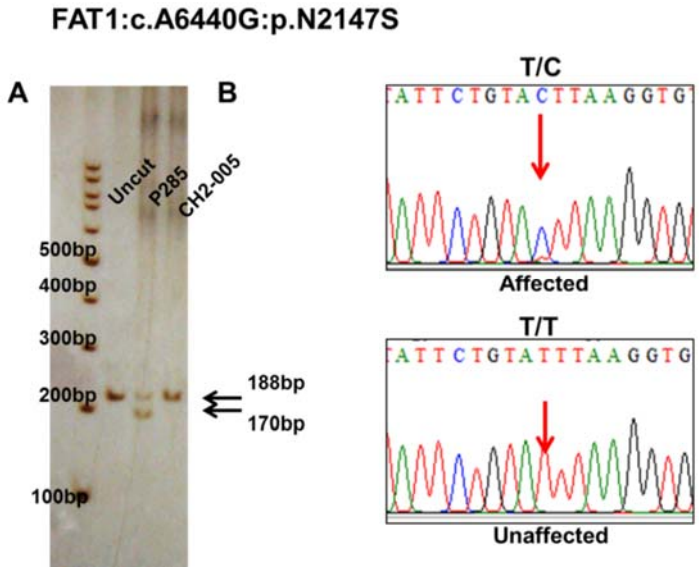

Figure 36 **Sequence chromatogram result of FAT1:c.A6440G:p.N2147S**

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

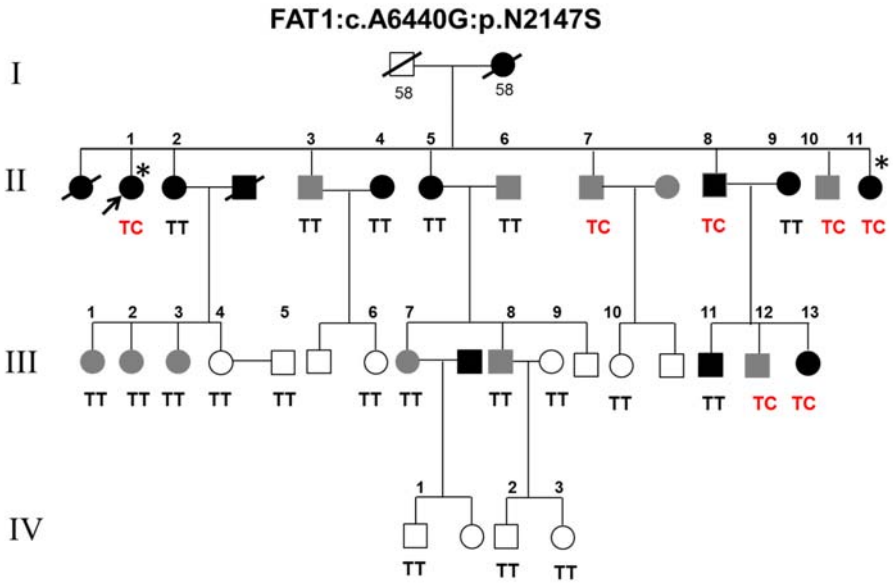

Figure 37 Segregation of FAT1:c.A6440G:p.N2147S with diabetes in family P285.

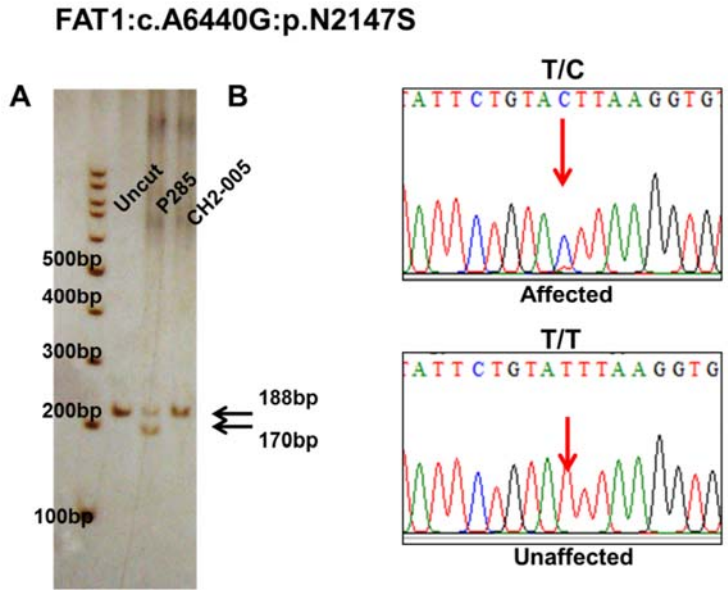

Figure 38 Sequence chromatogram result of FAT1:c.A6440G:p.N2147S

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

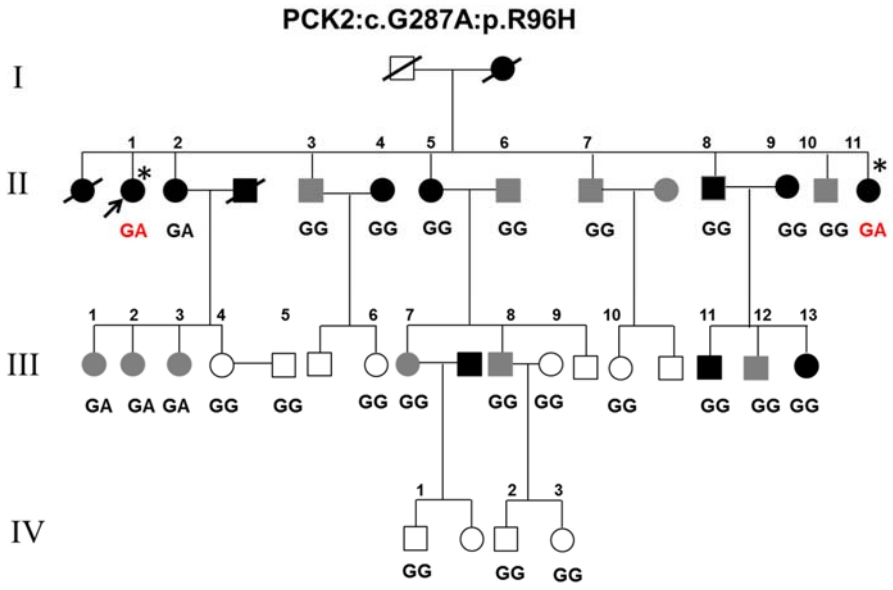

Figure 39 Segregation of PCK2:c.G287A:p.R96Hwith diabetes in family P285.

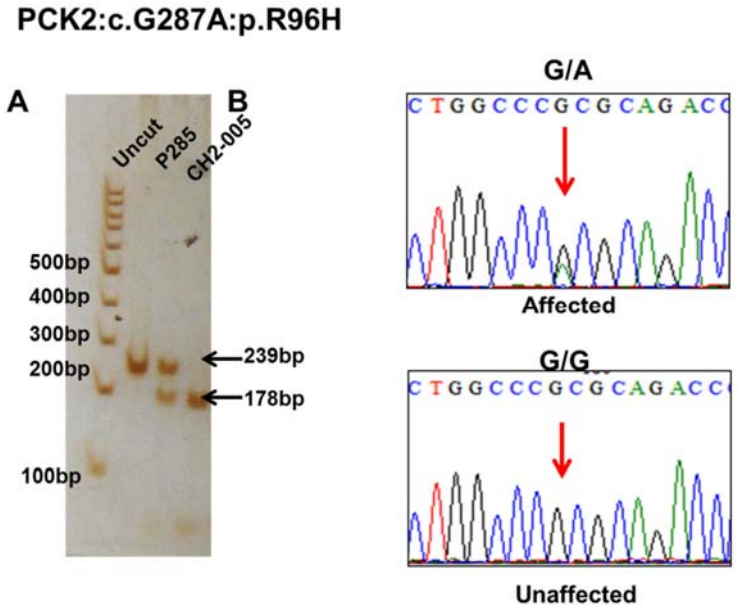

Figure 40 Sequence chromatogram result of PCK2:c.G287A:p.R96H

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) and subsequently confirmed by Sanger sequencing (B).

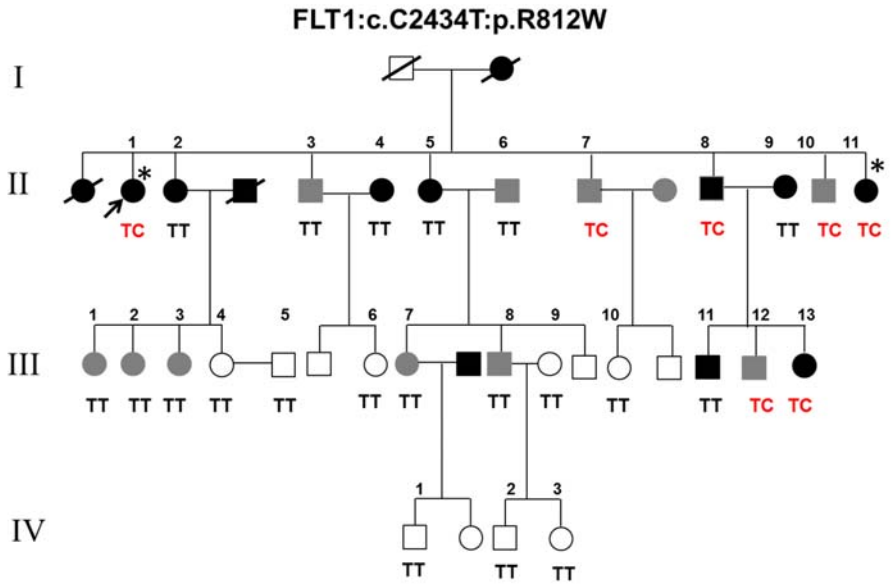

Figure 41 Segregation of FLT1:c.C2434T:p.R812W with diabetes in family P285.

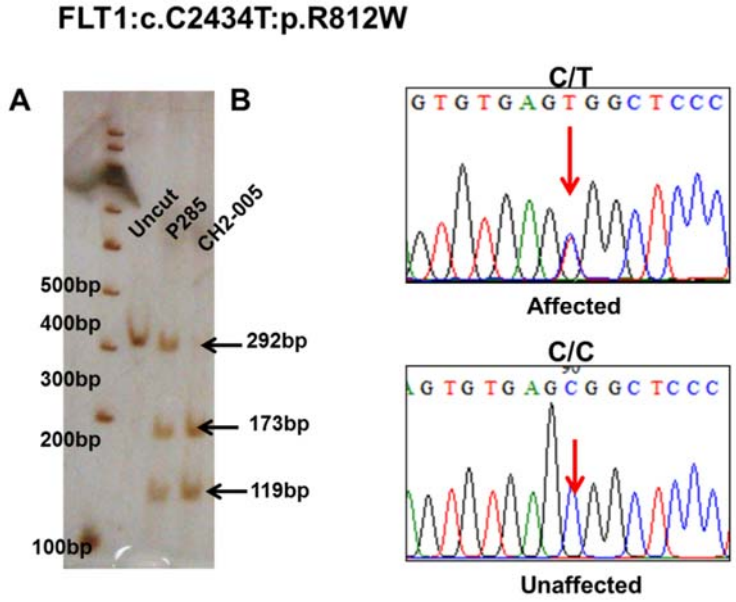

Figure 42 Sequence chromatogram result of FLT1:c.C2434T:p.R812W

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) and subsequently confirmed by Sanger sequencing (B).

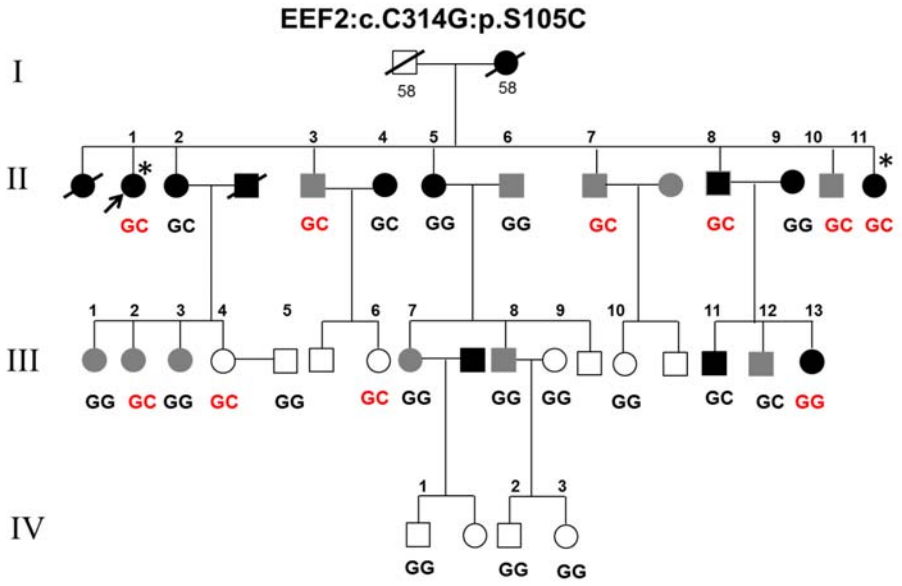

Figure 43 Segregation of EEF2:c.C314G:p.S105C with diabetes in family P285.

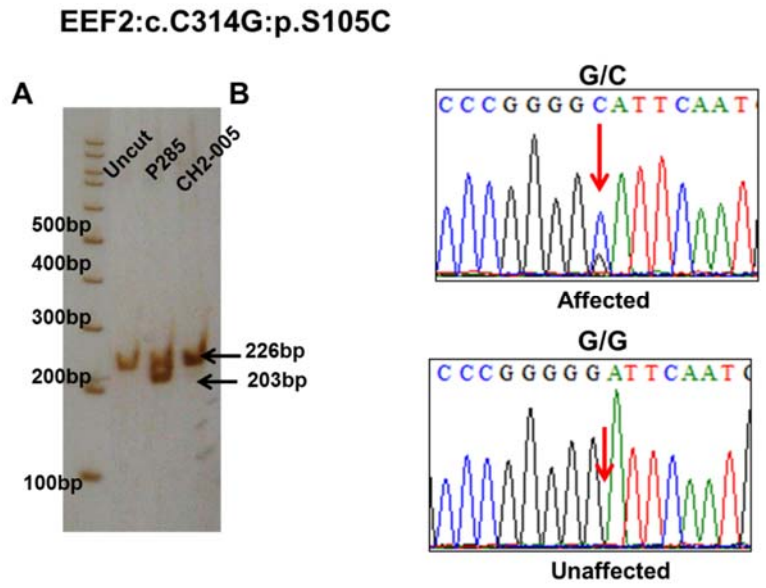

Figure 44 Sequence chromatogram result of EEF2:c.C314G:p.S105C

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

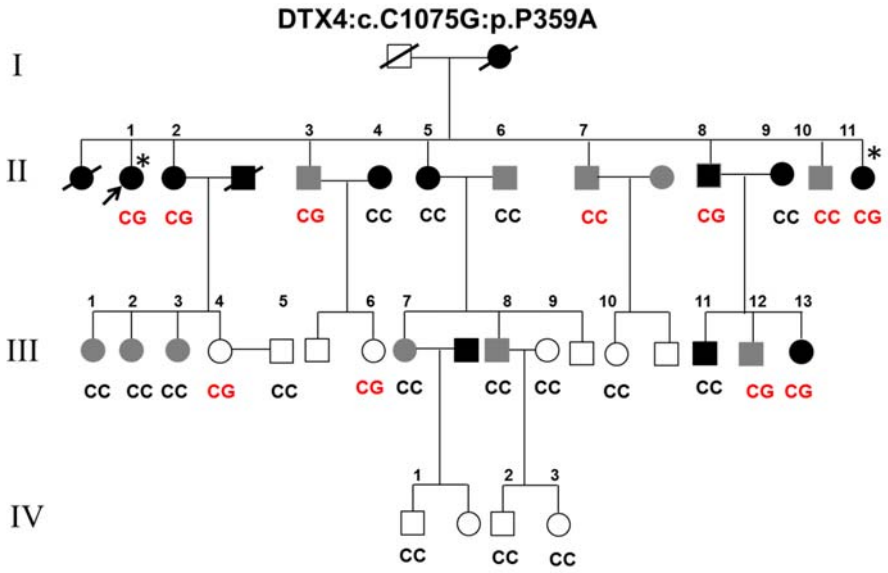

Figure 45 Segregation of DTX4:c.C1075G:p.P359A with diabetes in family P285.

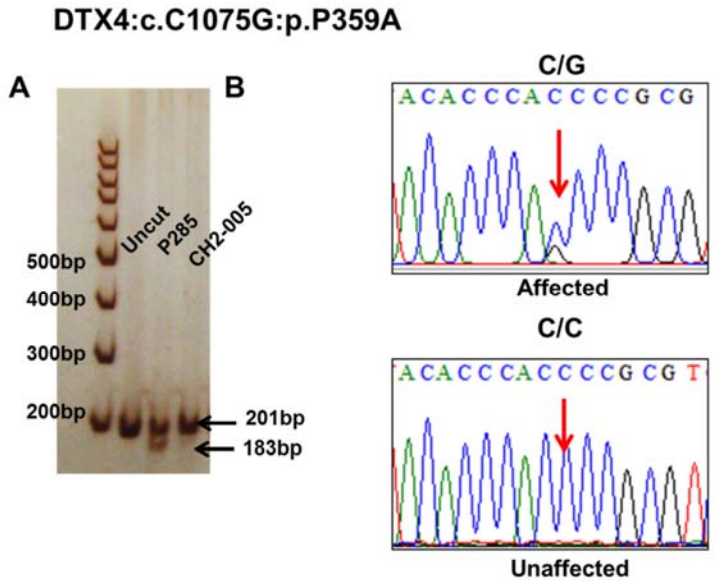

Figure 46 Sequence chromatogram result of DTX4:c.C1075G:p.P359A

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

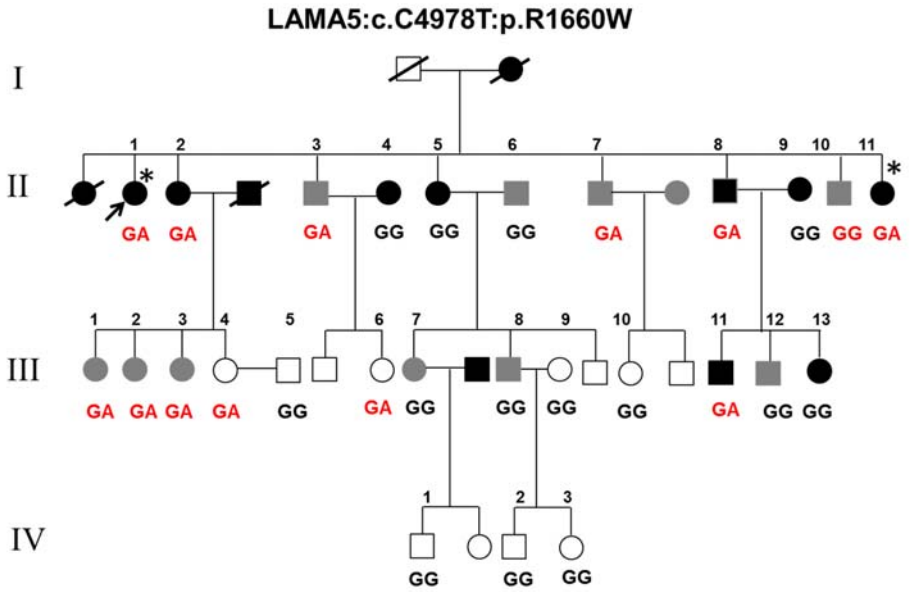

Figure 47 Segregation of LAMA5:c.C4978T:p.R1660W with diabetes in family P285.

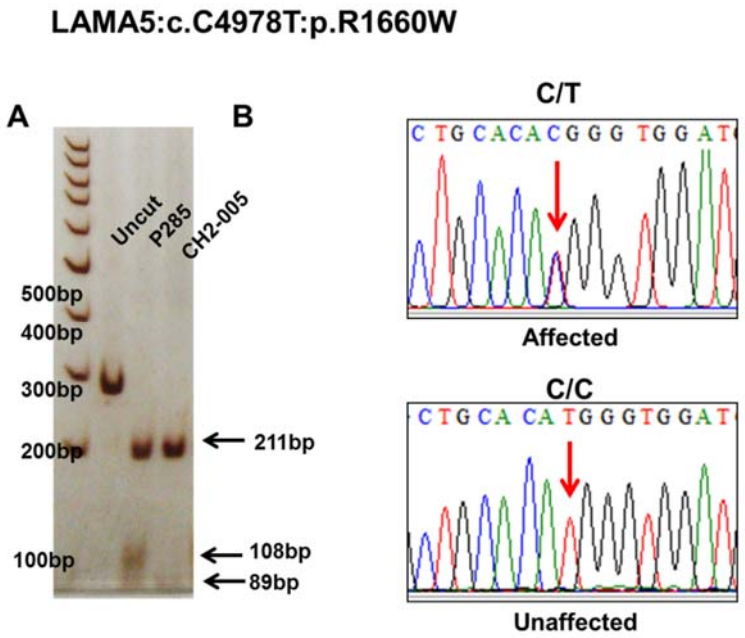

Figure 48 **Sequence chromatogram result of LAMA5:c.C4978T:p.R1660W.**
Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

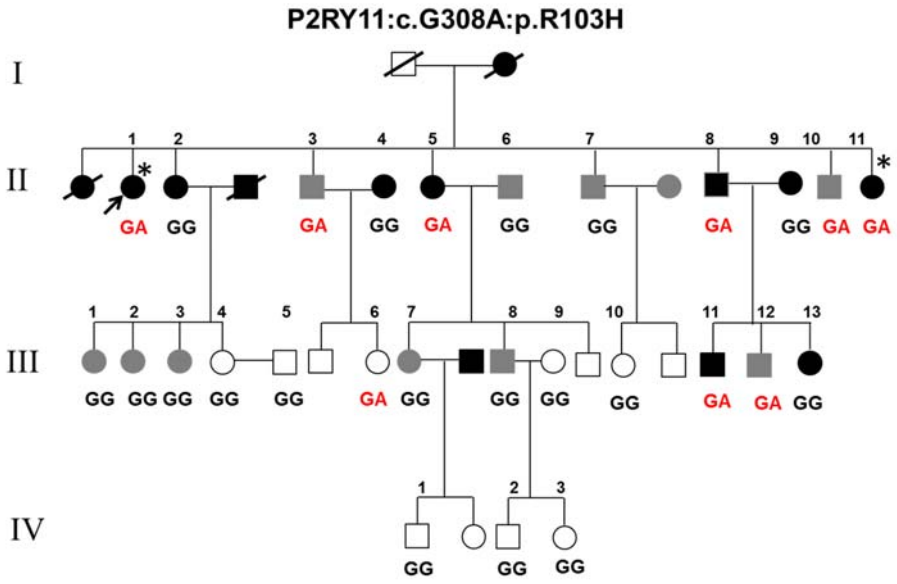

Figure 49 Segregation of P2RY11:c.G308A:p.R103H with diabetes in family P285.

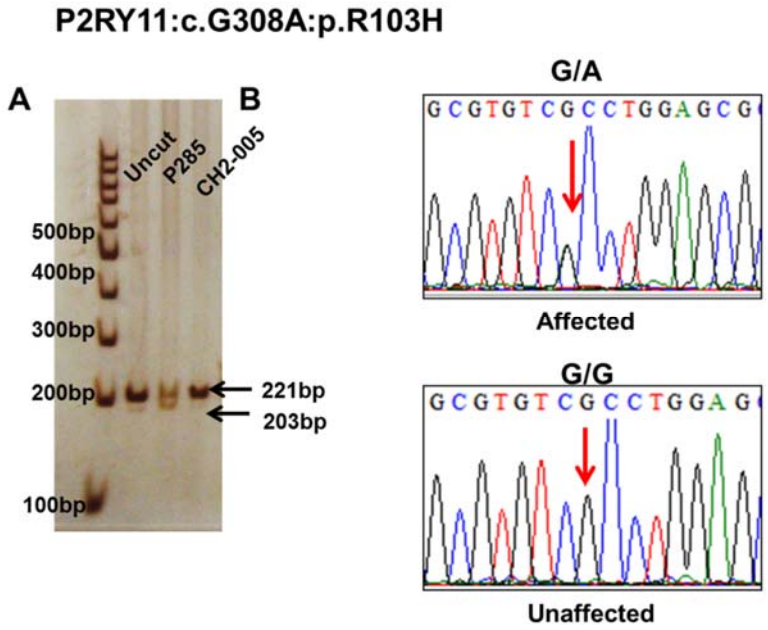

Figure 50 **Sequence chromatogram result of P2RY11:c.G308A:p.R103H.**
Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

3) Analysis of variants showed positive in silico program showed 4 out of 6 programs

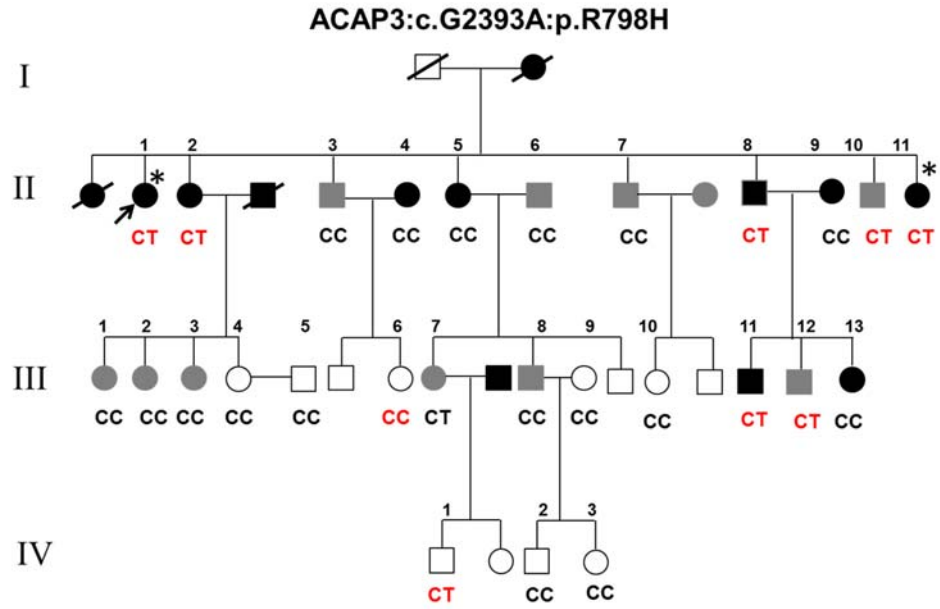

Figure 51 Segregation of ACAP3:c.G2393A:p.R798H with diabetes in family P285.

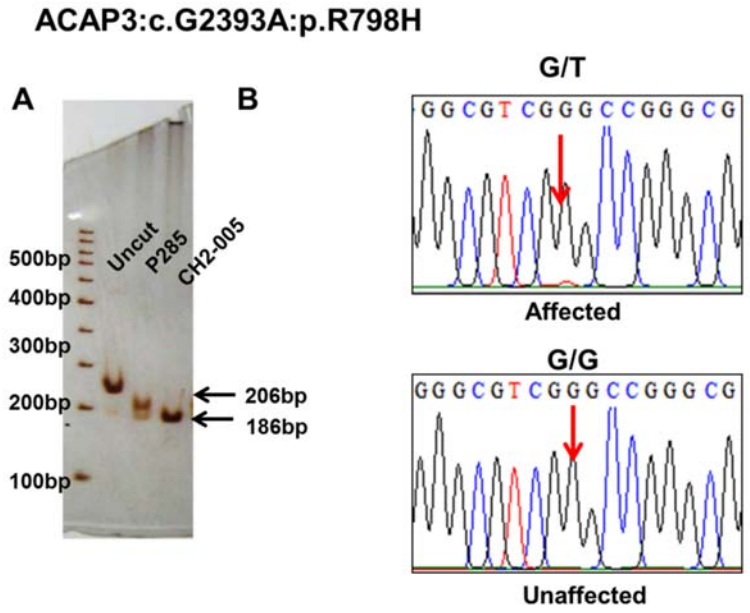

Figure 52 Sequence chromatogram result of ACAP3:c.G2393A:p.R798H
Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

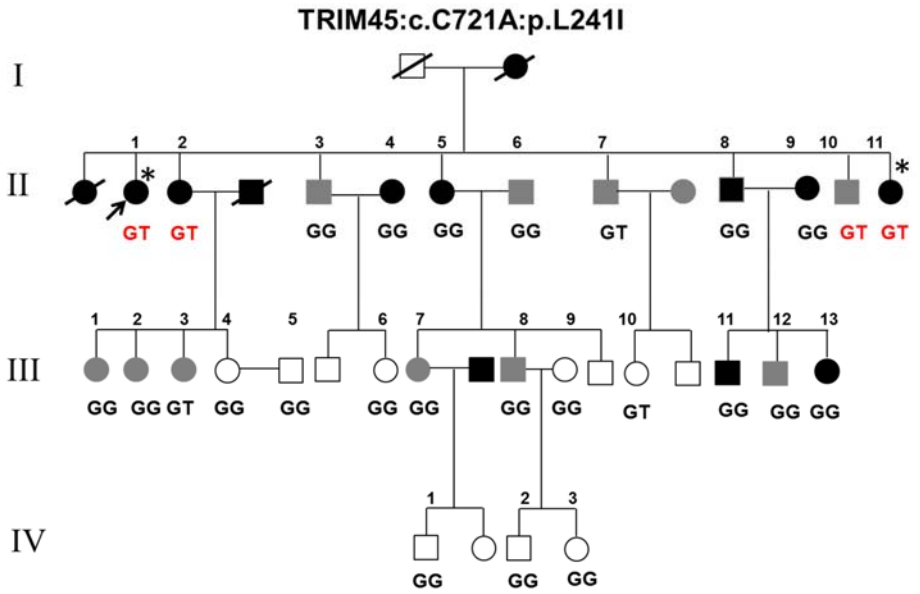

Figure 53 Segregation of TRIM45:c.C721A:p.L241with diabetes in family P285.

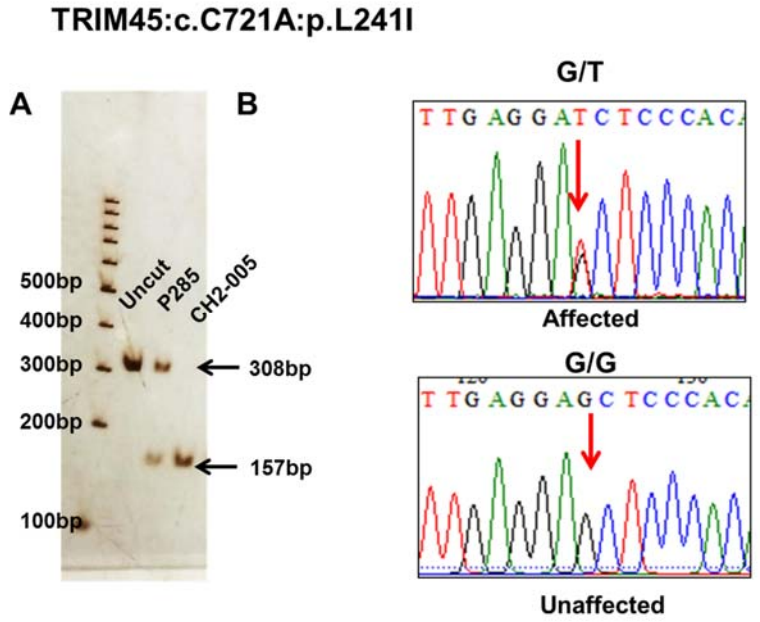

Figure 54 **Sequence chromatogram result of TRIM45:c.C721A:p.L241**

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

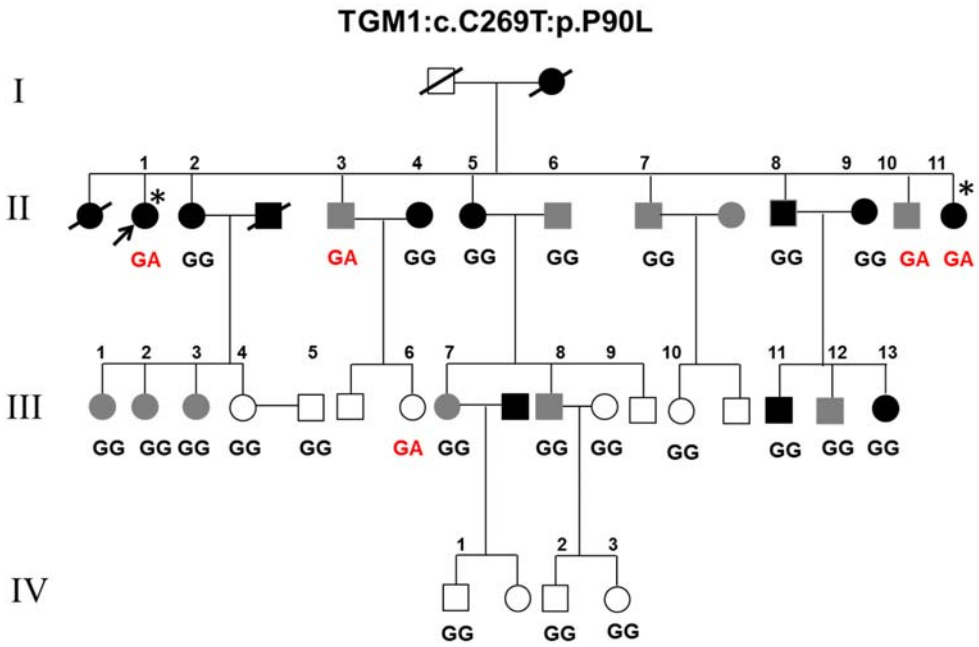

Figure 55 Segregation of TGM1:c.C269T:p.P90L with diabetes in family P285.

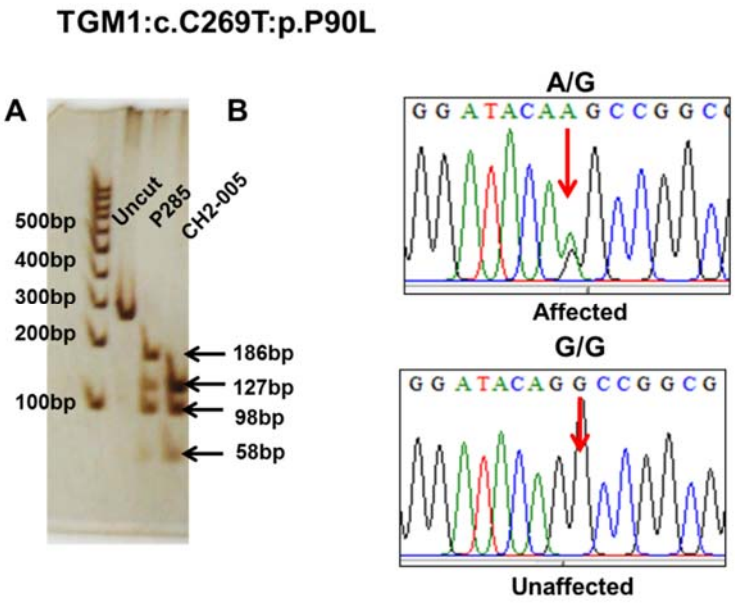

Figure 56 Sequence chromatogram result of TGM1:c.C269T:p.P90L

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B)

.

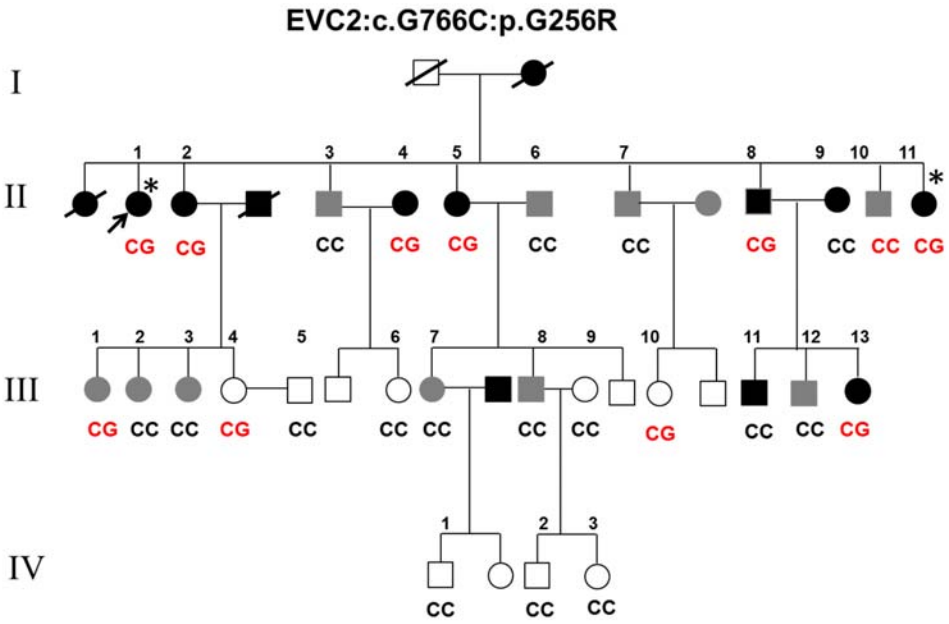

Figure 57 Segregation of EVC2:c.G766C:p.G256R with diabetes in family P285.

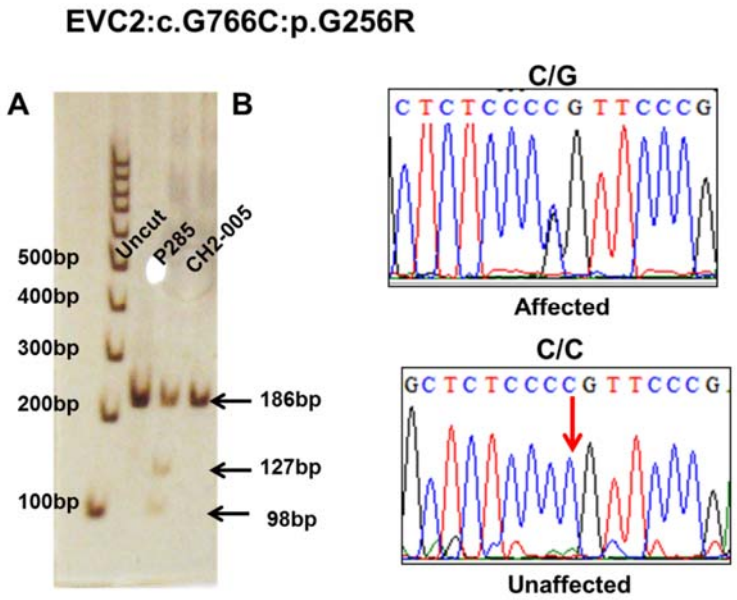

Figure 58 **Sequence chromatogram result of EVC2:c.G766C:p.G256R.**
Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

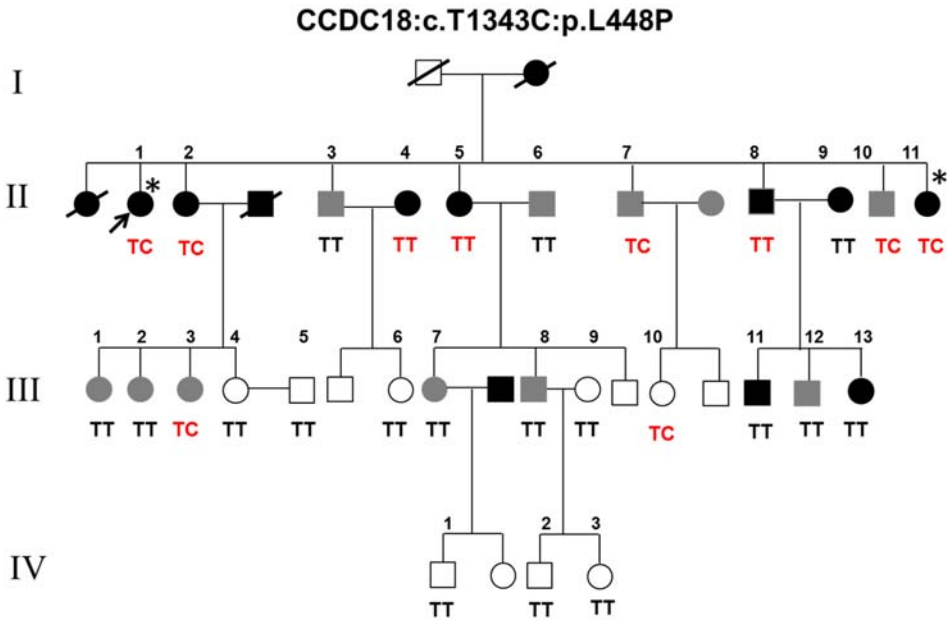

Figure 59 Segregation CCDC18:c.T1343C:p.L448P with diabetes in family P285.

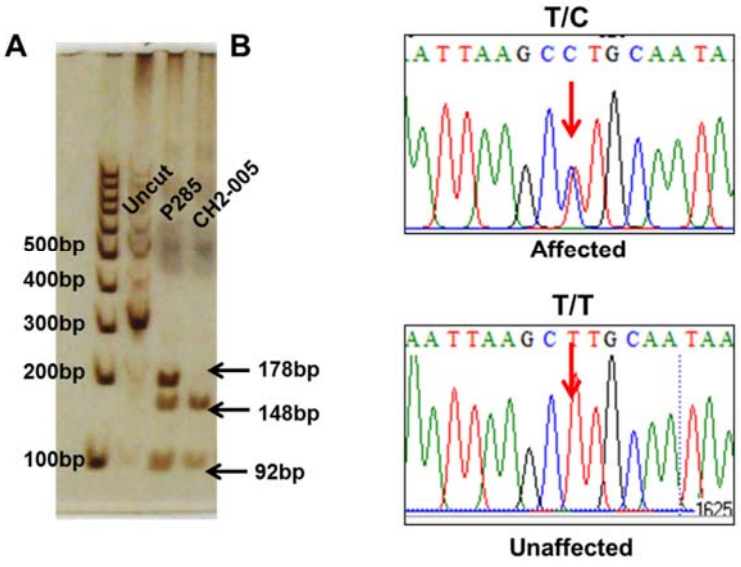

Figure 60 **Sequence chromatogram result of EVC2:c.G766C:p.G256R.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

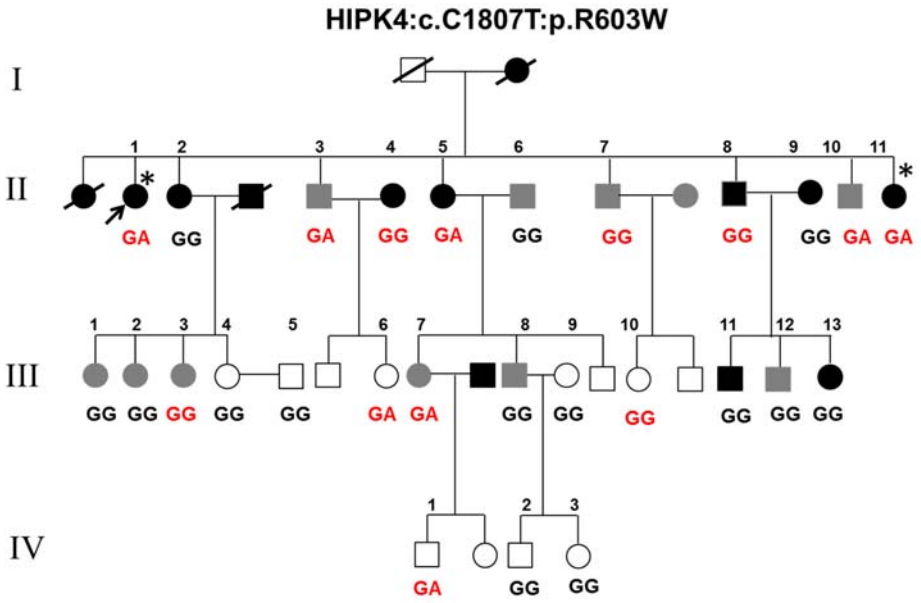

Figure 61 Segregation HIPK4:c.C1807T:p.R603W with diabetes in family P285.

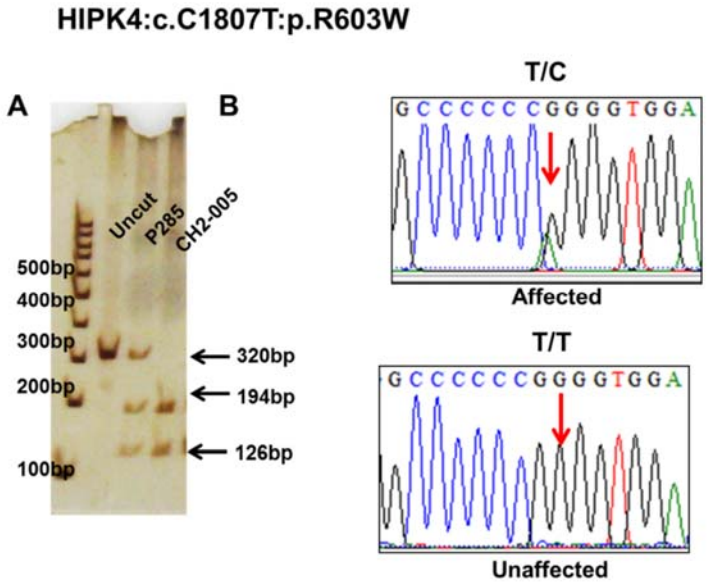

Figure 62 **Sequence chromatogram result of HIPK4:c.C1807T:p.R603W.**
Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

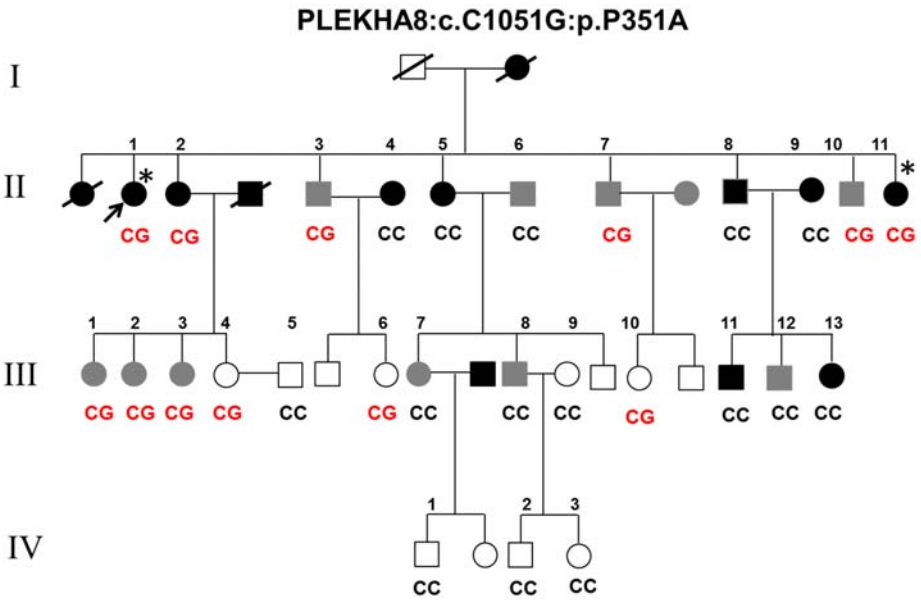

Figure 63 Segregation of PLEKHA8:c.C1051G:p.P351A with diabetes in family P285.

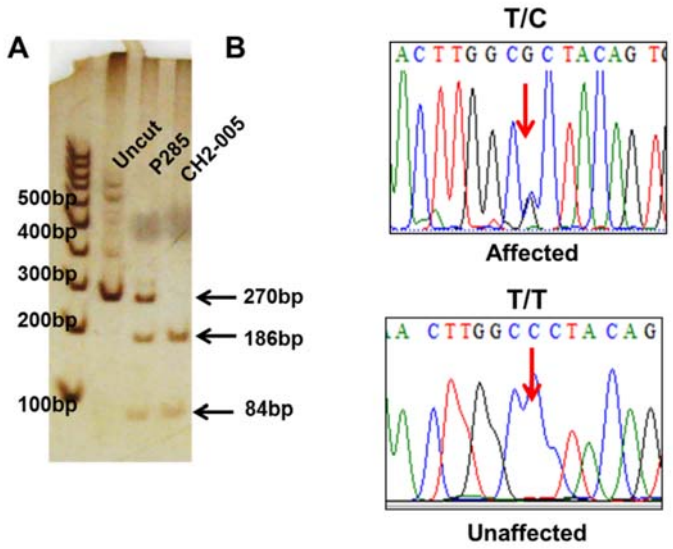

Figure 64 Sequence chromatogram result of PLEKHA8:c.C1051G:p.P351A Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

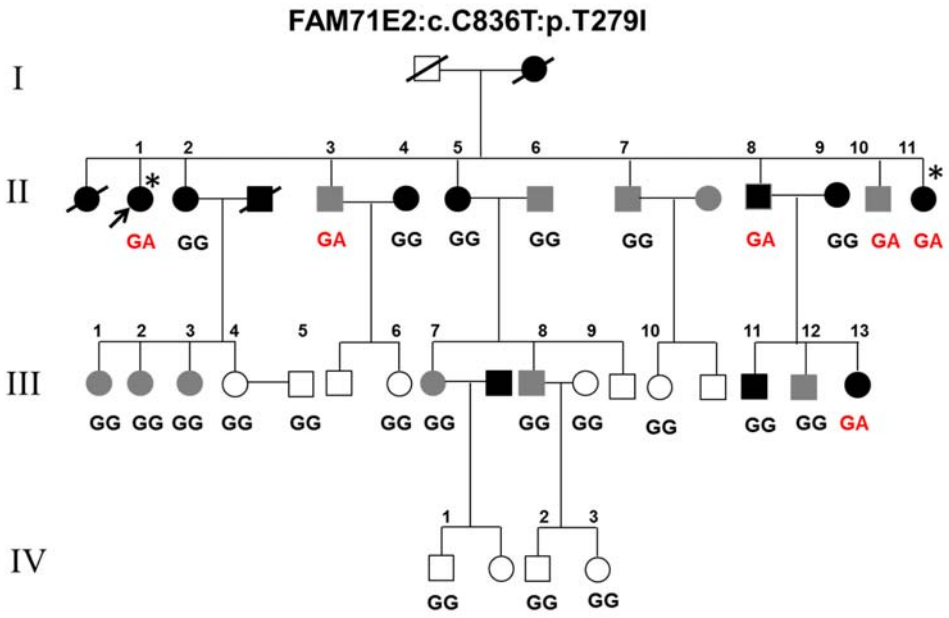

Figure 65 Segregation of FAM71E2:c.C836T:p.T279I with diabetes in family P285.

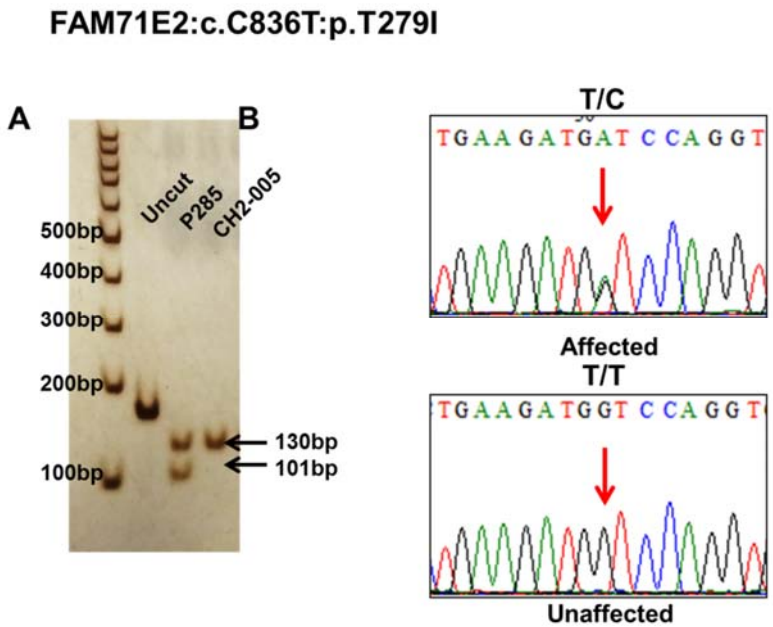

Figure 66 Sequence chromatogram result of FAM71E2:c.C836T:p.T279I Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

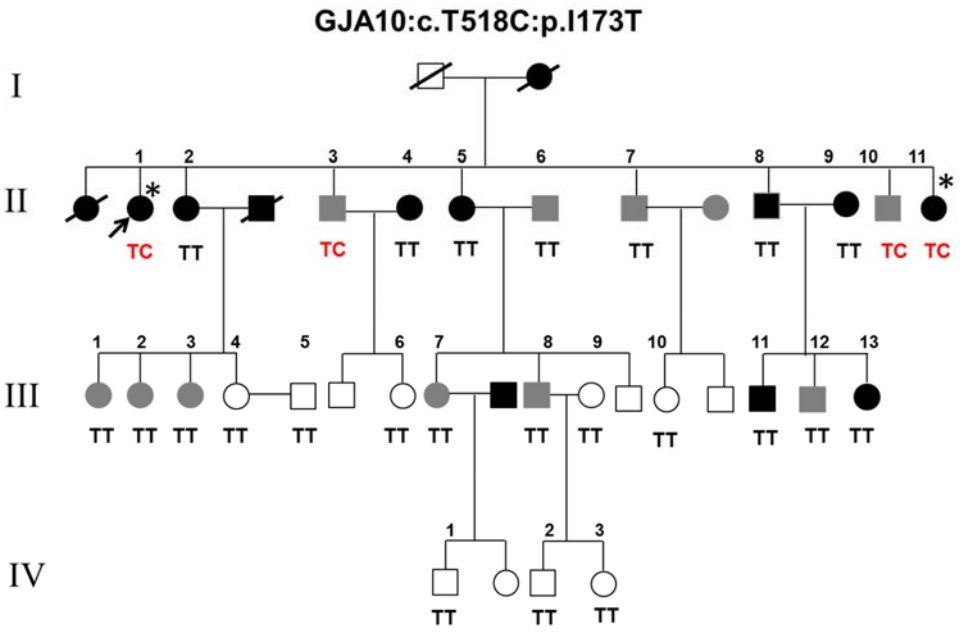

Figure 67 Segregation of GJA10:c.T518C:p.I173T with diabetes in family P285.

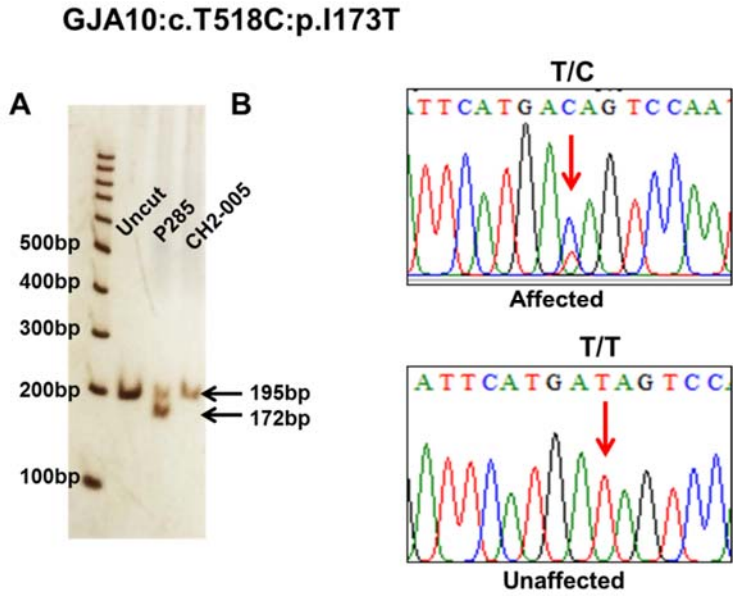

Figure 68 **Sequence chromatogram result of PLEKHA8:c.C1051G:p.P351A** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

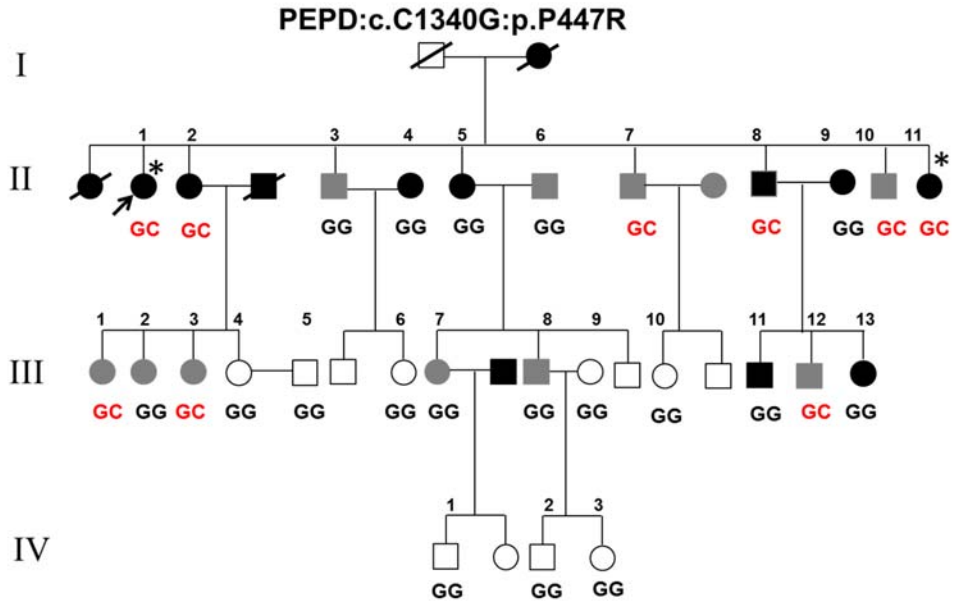

Figure 69 Segregation of PEPD:c.C1340G:p.P447R with diabetes in family P285.

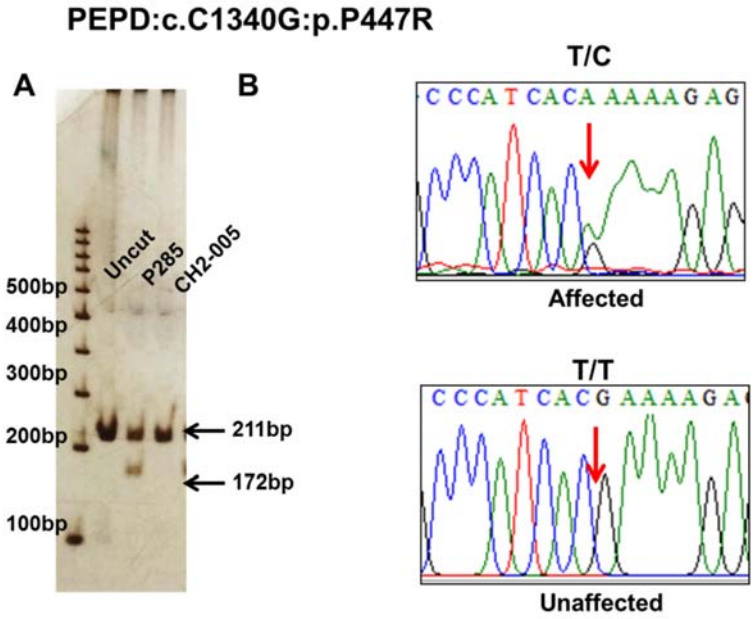

Figure 70 Sequence chromatogram result of PEPD:c.C1340G:p.P447R Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

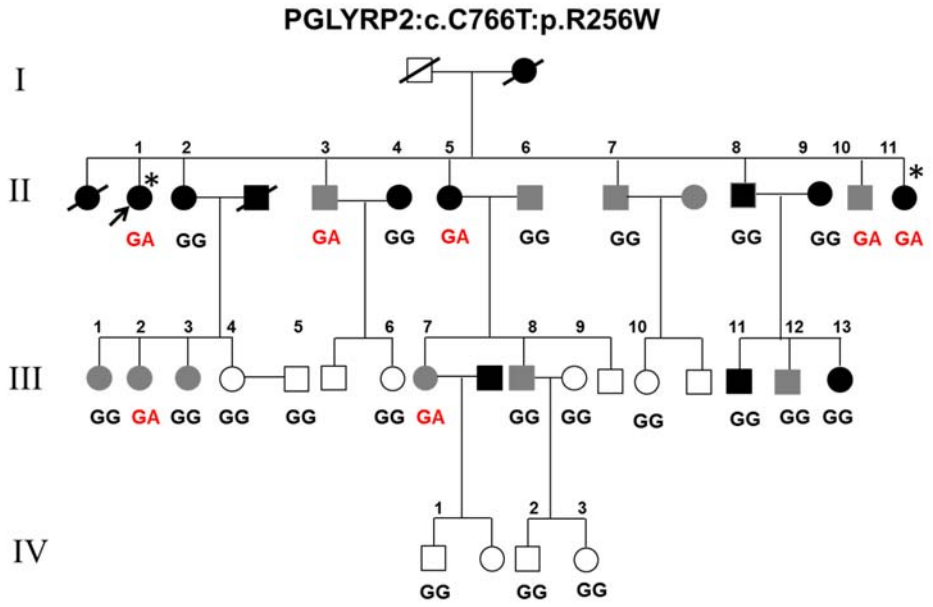

Figure 71 Segregation of PGLYRP2:c.C766T:p.R256W with diabetes in family P285.

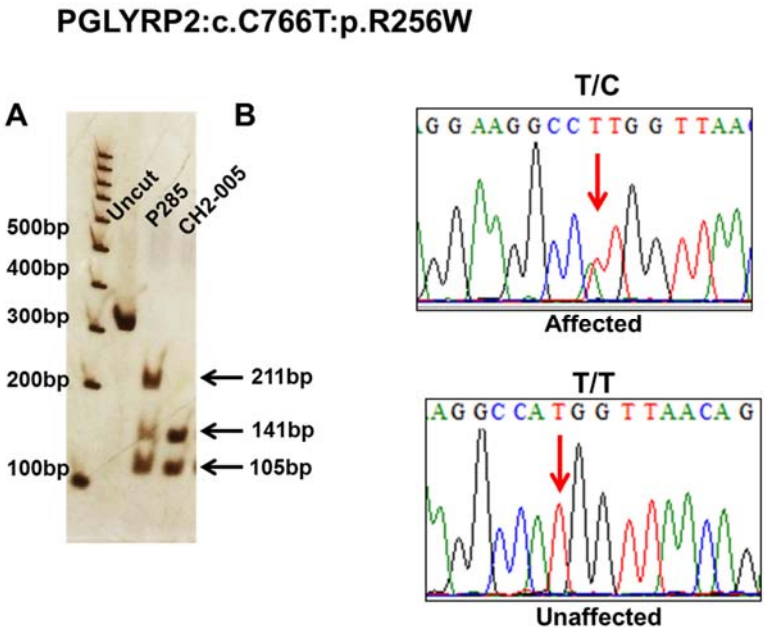

Figure 72 **Sequence chromatogram result of PGLYRP2:c.C766T:p.R256W.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

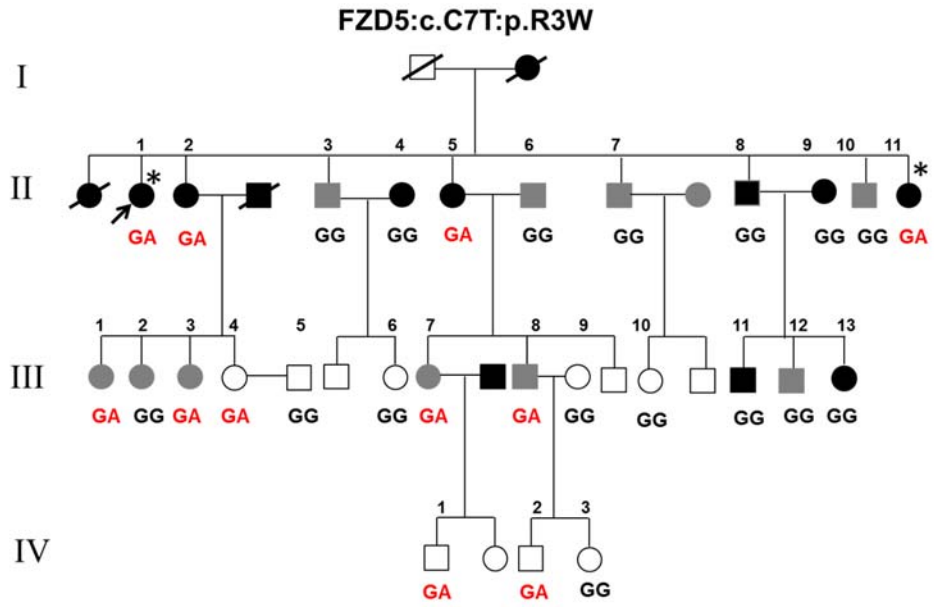

Figure 73 Segregation of FZD5:c.C7T:p.R3Wwith diabetes in family P285.

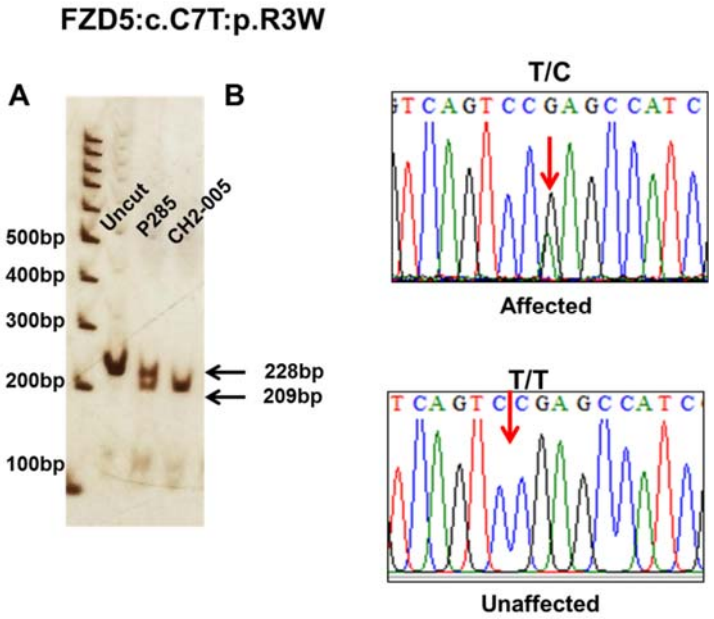

Figure 74 **Sequence chromatogram result of PEPD:c.C1340G:p.P447R.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

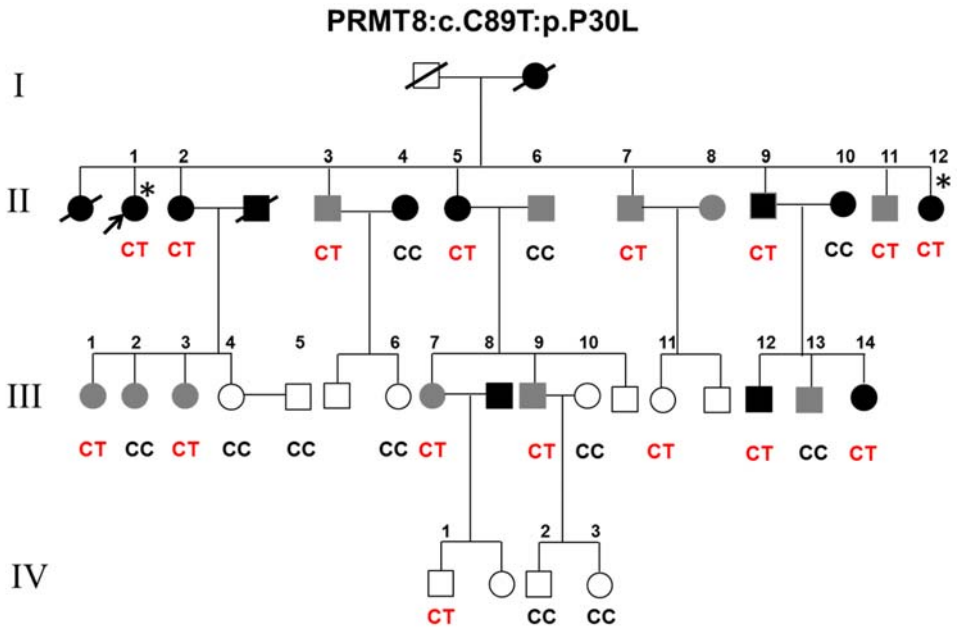

Figure 75 Segregation of PRMT8:c.C89T:p.P30L with diabetes in family P285

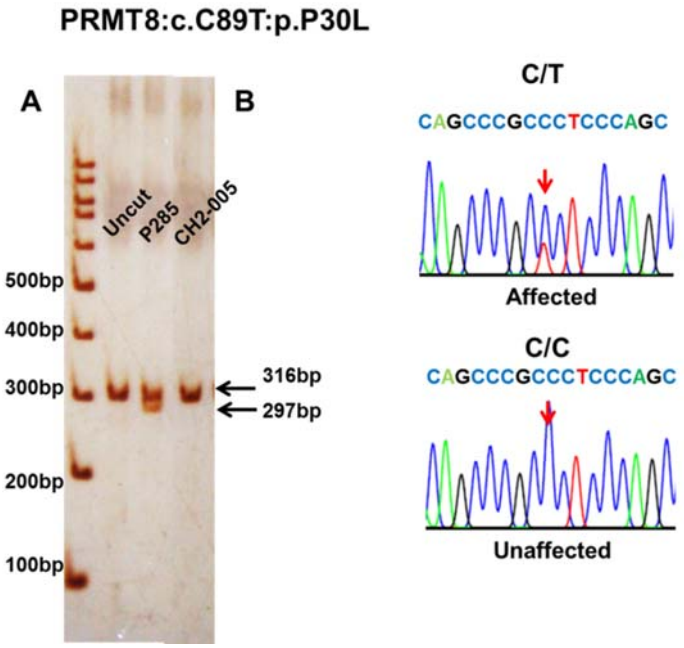

Figure 76 **Sequence chromatogram result of PEPD:c.C1340G:p.P447R.** Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

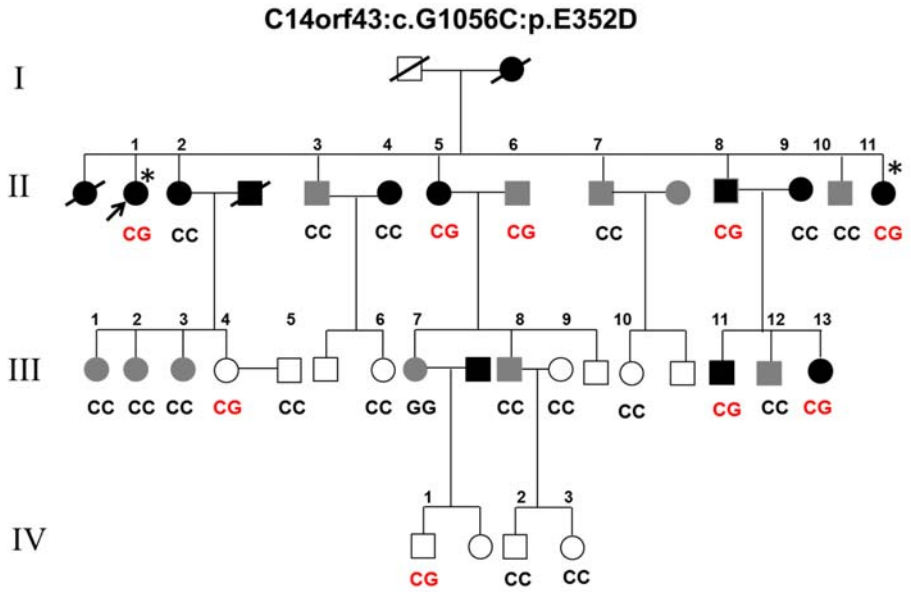

Figure 77 Segregation of C14orf43:c.G1056C:p.E352D with diabetes in family P285

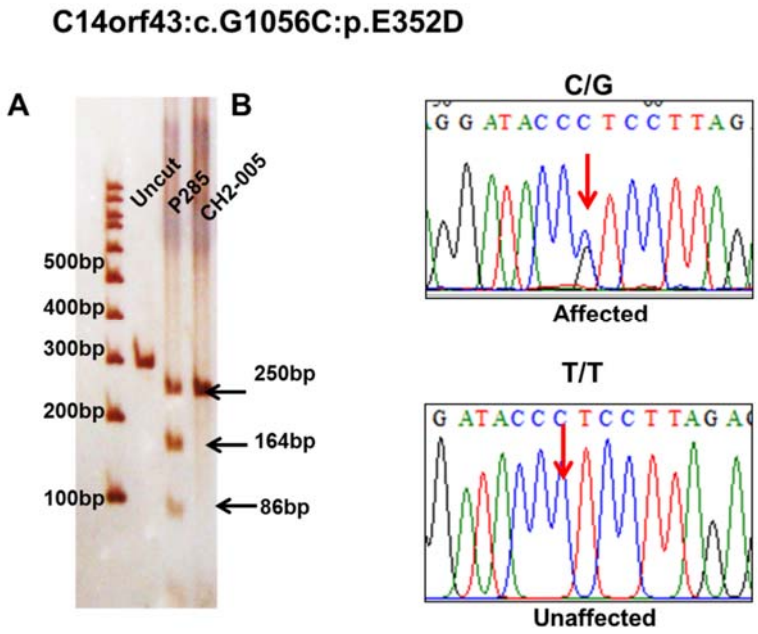

Figure 78 Sequence chromatogram result of C14orf43:c.G1056C:p.E352D.

Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

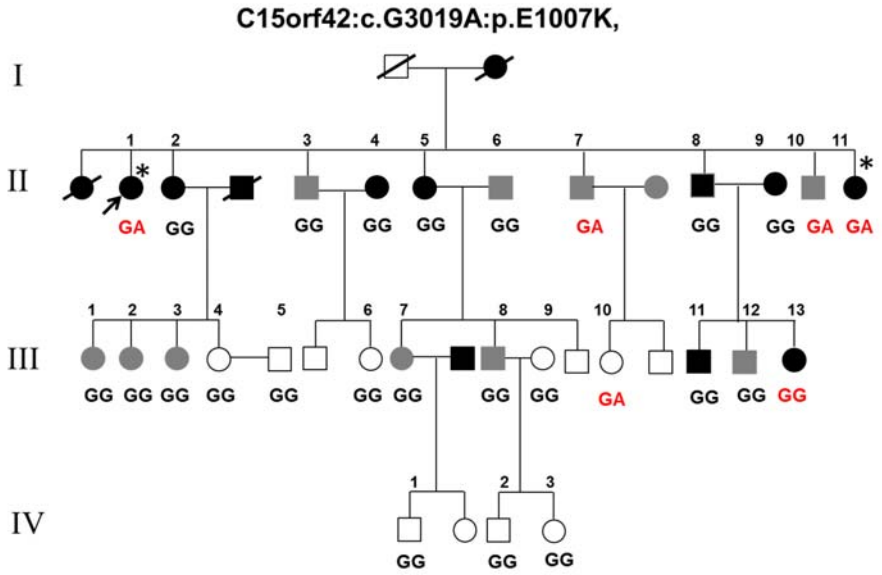

Figure 79 Segregation of C15orf42:c.G3019A:p.E1007K with diabetes in family P285

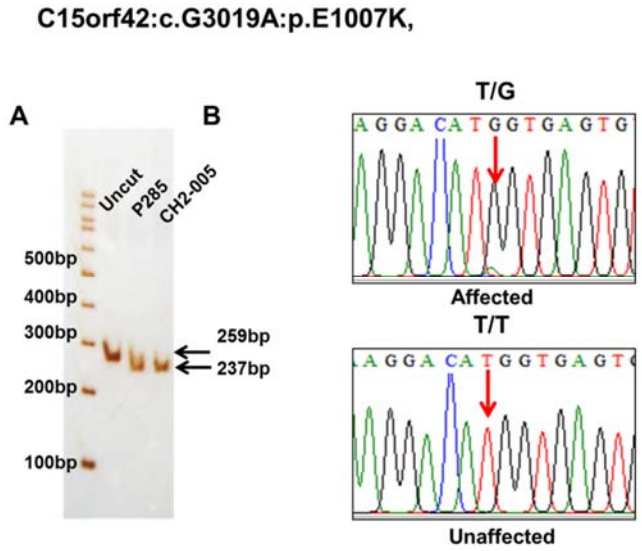

Figure 80 Sequence chromatogram result of C15orf42:c.G3019A:p.E1007K. Polyacrylamide gel electrophoresis of PCR product of RFLP pattern for homozygous variant in unaffected sample (CH2-005) and heterozygous variant in affected sample (P285) (A) and confirmed by Sanger sequencing (B).

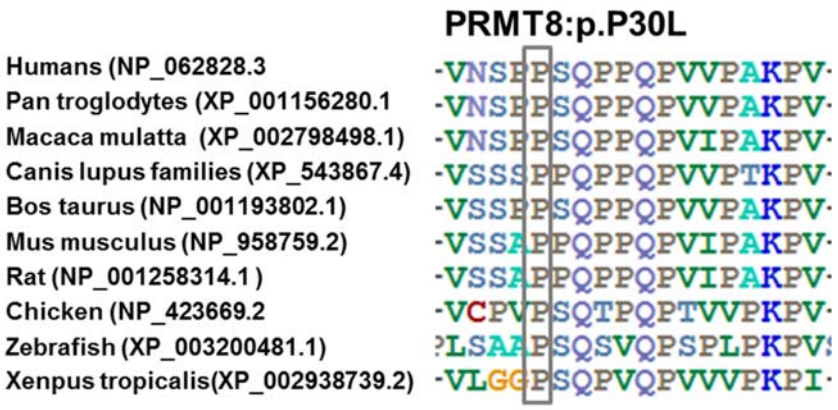

Figure 81 **The alignment of the amino acid sequences of the PRMT8 among different species.**

The Proline (P) at position 30 (p.P30, indicated in the box) is highly conserved among different species.

Table 12 Genotyping results of *PRMT8:*c.C89T:p.P30L in 400 non-diabetic controls and 1000 T2D probands.

Gene	Subjects	Genotype frequency		
PRMT8:c.C89T:p.P30L		C/C	C/T	T/T
	T2D (n=1000)	1000	0	0
	Non-diabetic controls (n=400)	400	0	0

Table 13 **Clinical parameters between affected and unaffected members in P285 type 2 diabetes family.**

Characteristics	Affected member	Unaffected member	*P* value
Number (n)	19	8	
Age (years)	65.0±13.7	33.0±15.8	**<0.0001**
Age at diagnosis (years)	60.3±12.3	-	
BMI (kg/m^2)	23.1±6.0	24.0±5.2	0.686
Waist circumrance (cm)	86.1±12.8	82.2500±17.70996	NS
Waist hip ratio	.89±.07	.83±.06	NS
Systolic BP (mmHg)	135.2±18.0	114.8750±15.4	**0.01**
Diastolic BP(mmHg)	81.3±14.1	75.3±13.2	NS
FPG (mg/dl)	115.6±44.0	84.0±6.84	NS
HbA1c (%)	6.4±.77	5.5±.14	**<0.0001**
Total Cholesterol(mg/dl)	197.7±50.6	214.0±42.9	NS
LDL(mg/dl)	116.2±46.3	135.1±36.9	NS
HDL(mg/dl)	60.1±15.3	64.1±11.6	NS

Table 14 **Clinical parameters between affected and unaffected members who carried genotype with C/T or C/C genotypes in P285 family.**

Characteristics	Affected member		*P* value	Unaffected member		*P* value
	Carried C/T	Carried C/C		Carried C/T	Carried C/C	
Number (n)	14	5		2	6	ND
Age (years)	65.07±14.58	65.00±12.63	0.273	27 ±15.55	35±16.89	0.578
Age at diagnosis (years)	65.00 ± 14.57	62.00± 14.57	0.435	ND	ND	ND
BMI (kg/m^2)	23.34±4.30	26.03±4.30	0.253	19.20±1.67	24.40±6.51	0.328
Waist circumference (cm)	83.28±11.51	94.00±14.47	0.687	72±14.14	85.66±18.52	0.385
Waist hip ratio	0.88±0.06	83.28±11.51	0.687	0.80±0.120	0.84±0.058	0.558
Systolic BP (mmHg)	132.92±19.87	141.60±10.33	0.102	102.50±13.43	119±14.68	0.212
Diastolic BP(mmHg)	82.07±14.79	79.40±13.39	0.667	61.00±8.48	80.16±11.01	0.069
FPG (mg/dl)	120.00±50.36	104.22±0.656	0.374	77.00±2.82	86.33±6.15	0.093
HbA1c (%)	6.56±0.83	6.20±0.57	0.701	5.40±0.28	5.53±0.08	0.279
Total Cholesterol(mg/dl)	205.69±57.62	177±13.60	0.082	199.00±59.39	221.50±40.82	0.603
Triglyceride (mg/dl)	111.53±79.84	93.60±18.56	0.631	52.50±7.77	71.50±19.20	0.268
LDL(mg/dl)	111.53±79.84	93.60±18.56	0.169	119.00±46.10	143.20±36.17	0.512
HDL(mg/dl)	59.53±17.48	61.60±8.98	0.386	64.5±21.92	64.00±8.08	0.966

ND is not determined.

Discussion

Type 2 diabetes (T2D) is becoming increasingly prevalent throughout the world. The number of people living with diabetes is expected to increase from 387 million in 2014 to 592 million by 2035 according to the 6th Edition of the International Diabetes Federation's (IDF) Diabetes Atlas (1). In Thailand alone, there are more than over 4 million cases of diabetes in Thailand in 2015. Although lifestyle changes, an increasing prevalence of obesity, and an increasingly aging population are important drivers of this epidemic, genetic factors also play a major role in T2D susceptibility.

The recent development of high throughput methods of deoxyribonucleic acid (DNA) sequencing has made it possible to determine individual genome sequences and their specific variations. A region of particular interest is the protein-coding part of the genome, or exome, which is composed of gene exons. Exome sequencing has been used in the study of various diseases. Monogenic diseases with Mendelian inheritance are among these, but studies have also been carried out on genetic variations that represent risk factors for complex diseases. Diabetes is another intensive area for exome sequencing studies. Several examples were shown the use of exome sequencing in the diagnosis, prognosis, and treatment. Finally, remaining challenges and some practical and ethical considerations for the clinical application of exome sequencing were needed to further investigated.

Until now, it is clear that exome sequencing will often lead to the identification of the causative variant for Mendelian diseases as well as diabetes. In present study, we used exome sequencing approach to identify pathogenic variant a consanguineous pedigree with autosomal dominant T2D in which the exomes of two affected were sequenced. Approximately, 68,583 variants were identified by exome sequencing. Despite the enormous amount of exome sequencing data have been generated, potential filtering criterion were applied to reduce amount of such variants. By using filtering criterion, 140 variants heterozygous rare variants were chosen. The majority (99.79%) of variants were excluded. Application of these filters results in a significantly decreased large-scale exome sequencing dataset, remaining only 40 variants were selected Using various in silico programs, PolyPhen-2, SIFT, Mutation Taster, VarioWatch, Provean and SNP&GO, to help select proper putative candidate variants for further investigation.

A novel variant at c.C89T: p.P30L in the *PRMT8*, have been identified and partially co-segregated with fourteen affected and two unaffected individuals within family, absent in the other unaffected and two affected individuals (Figure 75). Two unaffected individuals, (III-10) and (III-1), carried heterozygous variant had age 38 and 16 year old, respectively. It may be possible that they are young adults thus they have opportunity to develop diabetes in the future.

The *PRMT8* (protein arginine methyltransferase 8) is a member of PRMTs family encoding the protein product that participated in Arginine methylation, a post-translational modification. Basis for functional characterization of the PRMTs mediated the methylation of a number of protein substrates of arginine residues and plays an important roles in many cellular processes, including signaling transduction, transcriptional modification, and mRNA splicing and transcriptional repression. Currently, PRMT family members are classified in two types. Type I enzyme, the most common type of PRMT, induces asymmetric dimethylation, adding two methyl groups to the terminal nitrogen atoms (***ω***-N^G,N^G-dimethylarginine). Six enzymes are categorized as type I PRMTs: PRMT1, PRMT2, PRMT3, PRMT4, PRMT6, and PRMT8. Another type, Type II enzymes add one methyl group to the terminal nitrogen atoms (***ω***-N^G,N^G-dimethylarginine) by catalyzing the symmetric dimethylation of arginine side chains, including PRMT5, PRMT7 and PRMT9 **(**Figure 82) (2).

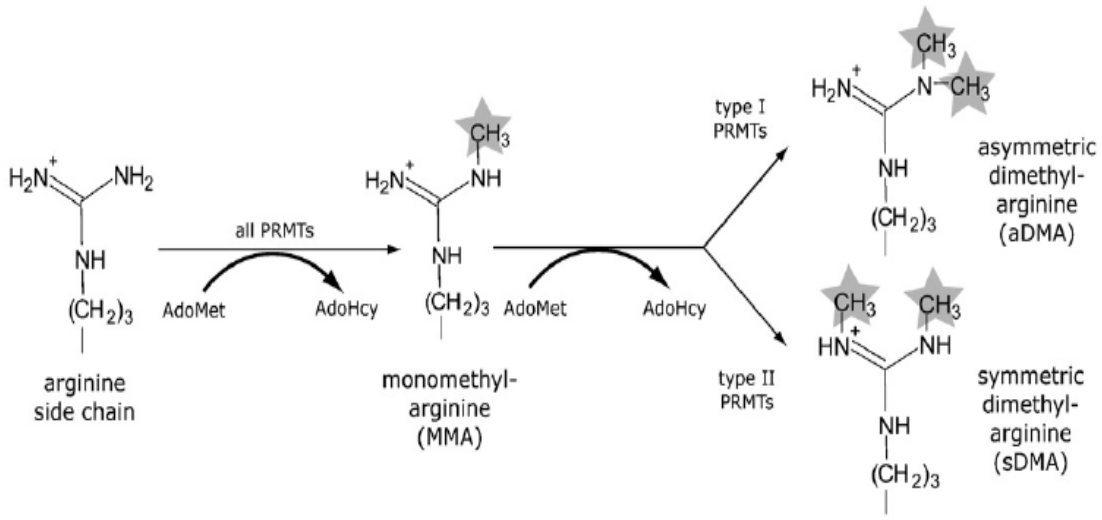

Figure 82 Methylation of the arginine side chain by PRMTs.

All of PRMTs catalyze the formation of monomethyl-arginine (MMA), at the same time the methyl donor S-adenosyl- L-methionine (AdoMet) is converted to S-adenosyl-L-homocysteine (AdoHcy). In a second step, the type I PRMTs transfer a second methyl group to the same guanidino nitrogen resulting in an asymmetric dimethylarginine (aDMA) whereas the type II PRMTs catalyze the formation of symmetric dimethylarginines (sDMA).

All PRMTs family there have common catalytic methyltransferase domain which consist of highly conserved region around 310 amino acid and subdomain that important to binding to methyl donor and substrate (2). In each individual PRMTs family member, they have different unique N-terminal region for example, SH3, zinc finger (ZnF), or myristoylation (Myr) motif from PRMT2, PRMT3 and PRMT8 respectively and PRMT4 has an unique C-terminal region, PRMT 7 and PRMT 10 have a second catalytic domain (Figure 83) (2).

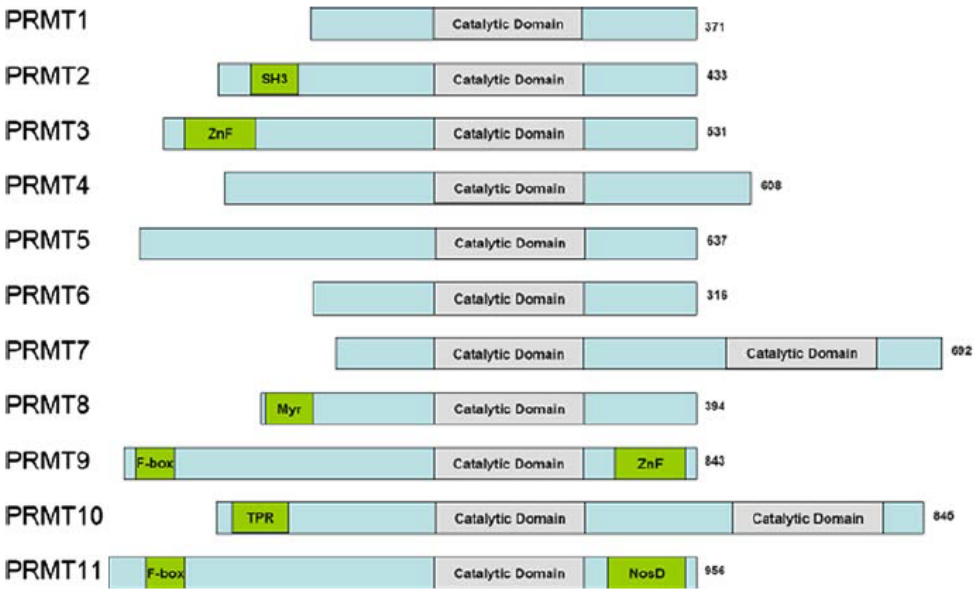

Figure 83 The human PRMT family. The different members are indicated as boxes and the protein length by the numbers of amino acids. Different additional domains are highlighted in (green)

In this study, we found that a novel missense *PRMT8* (c.C89T: p.P30L) contributed to familial from of T2D. PRMT8 gene (10 exons) spans ~10.27 kb of genomic DNA which is located on chromosome 12p13.3 and transcribed into mRNA (sizing 2,417bp). From SNPset: SS2642 of PRMT8 Gene that created on 5/25/2016, there are 780 SNPs reported along the gene including promoter, 5', 3' UTR and exons (Figure 84A-B). Only 8 variations found in coding regions. Alternative splicing produced two protein isoforms, containing 394 and 385 amino acid residues. These protein isoforms contained unique 76 amino acid N-terminal region (myristolation) that regulates PRMT8 activities (Figure 84C). (3). Expression analysis of PRMT8 by using northern analysis revealed a unique tissue specific expression was mostly found in human brain tissue. In addition, the PRMT8 activities are regulated by Myristolation (Mys) domain, a unique 76 amino acid N-terminal region. Previous reported has been shown that PRMT8 were localized in the plasma membrane of the human brain. In addition, the mutated PRMT8 on a unique N-terminal region contributed to ability of The PRMT8 localization in the plasma membrane (4). Surprisingly, in this report has been shown that mutation on Pro30Leu, exon 2 of the PRMT8 gene located in the part of myristolation (Mys) domain. It is possible that this mutation may cause defective ability to PRMT8 function (Figure 85).

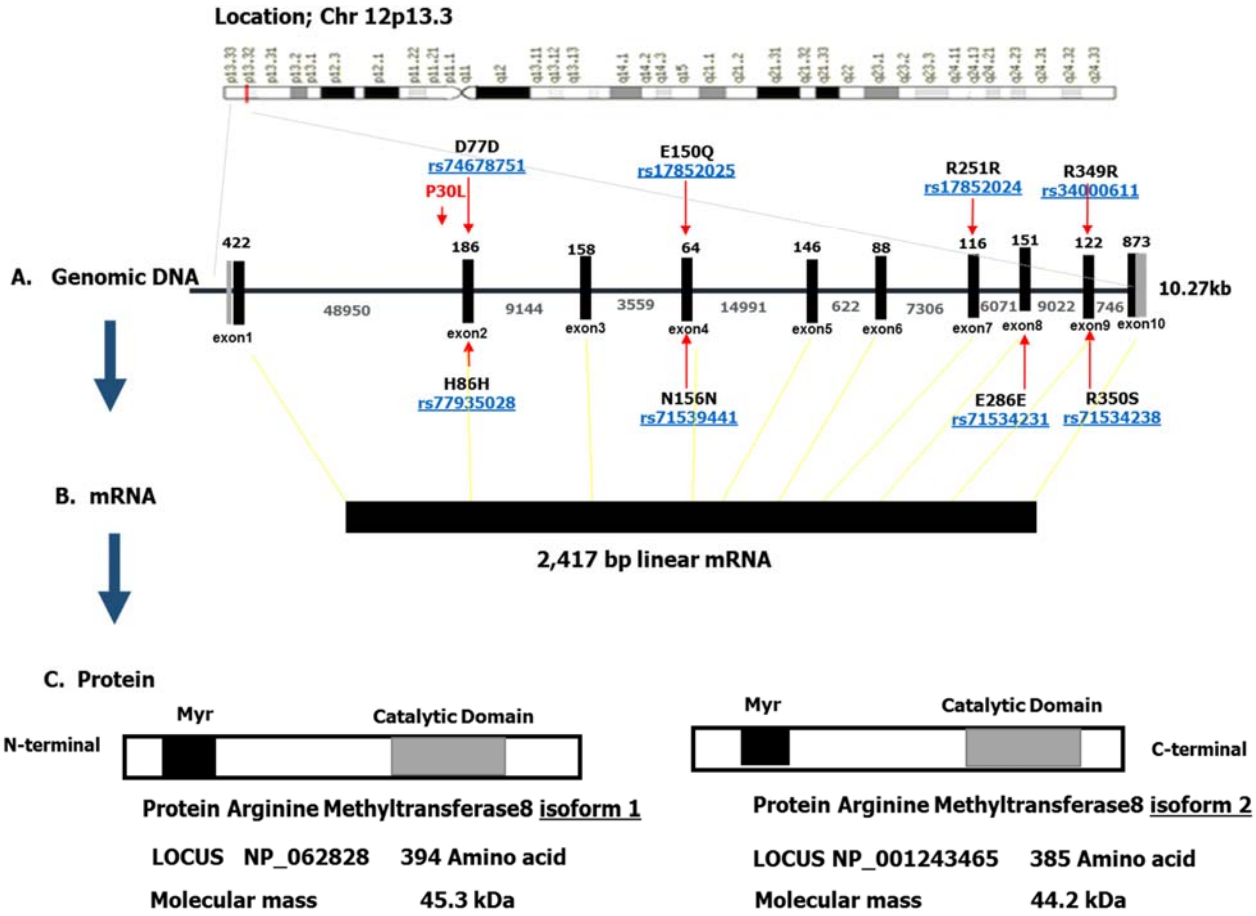

Figure 84 Illustrated the *PRMT8* gene showing the location on chromosome 12p13.3.

(A) Novel mutation on exon 2 and reported SNPs, (B) PRMT8 mRNA (C) and protein products of such gene.

.

PRMT8 is most closely related to PRMT1 that sheared more than 80 percent homology, although it has a distinctive N-terminal region (4). Moreover, type I enzyme of PRMTs; PRMT1, PRMT4, PRMT5 and PRMT6, potentially involve hepatic gluconeogenesis through specific regulation of FoxO1 and CREB-dependent transcriptional factor (5). However, the crucial role of PRMT8 in glucose homeostasis in pancreatic beta-cells has not been explored. Taken together, a novel mutation Pro30Leu of the *PRMT8* may contribute to development of familial T2D. The impact of this mutation on glucose homeostasis will be subjected for further investigation.

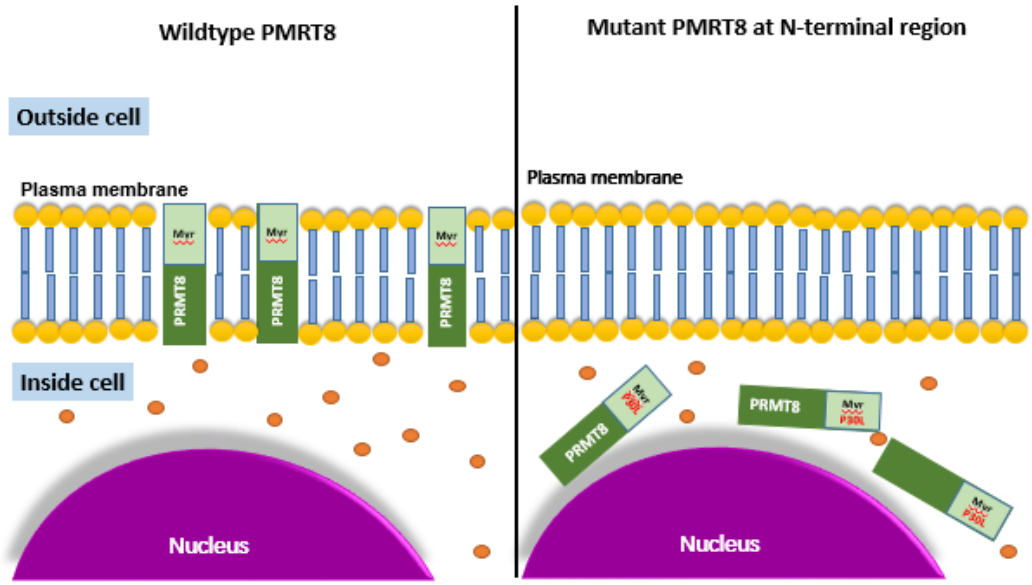

Figure 85 Illustrated the PRMT8s which were localized on plasma membrane. However, Single amino acid alteration (Pro30Leu) at N-terminal myristoylation region may contribute to ability of PRMT8 localization in plasma membrane.

Exome sequencing has led to the discovery of many causative variants in the genetic basis for a disease. However, the limitation of exome sequencing has been addressed to such as (i) genes in all exons. A few exons, such as those buried in stretches of repeats out towards the chromosome tips, which were not capture by capture kit .(ii). Structural variants, such as translocations and inversions, that move or flip DNA but don't alter the base sequence (iii) regulatory regions spread across the genome (transcription binding sites and enhancers).(6)

Conclusion

1. This study deal with genetic factors in type 2 diabetes family by using application of exome sequencing approach combined with a potentially filtration methodology. This work were successfully used to investigate the genetic cause of type 2 diabetes in Thai family.

2. A novel mutation in *PRMT8* gene (c.C89T: p.P30L) was possibly responsible for type 2 diabetes in a Thai family. This mutation was afflicting in three-generation of family

3. The conservation of amino acid at position 30 in various PRMT8 orthologs and bioinformatics prediction of the variant also supported the association between mutation and the disease.

4. .Molecular biology of *PRMT8* (c.C89T: p.P30L) variant should be further studied to promote a better understanding of mechanism underling the pathogenesis of diabetes.

References

1. Group IDFDA. Update of mortality attributable to diabetes for the IDF Diabetes Atlas: Estimates for the year 2013. Diabetes Res Clin Pract Sep;109(3):461-5.

2. Wolf SS. The protein arginine methyltransferase family: an update about function, new perspectives and the physiological role in humans. Cell Mol Life Sci2009 Jul;66(13):2109-21.

3. Xie Y, Ke S, Ouyang N, He J, Xie W, Bedford MT, et al. Epigenetic regulation of transcriptional activity of pregnane X receptor by protein arginine methyltransferase 1. J Biol Chem2009 Apr 3;284(14):9199-205.

4. Lee J, Sayegh J, Daniel J, Clarke S, Bedford MT. PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family. J Biol Chem2005;280(38):32890-6.

5. Han H-S, Choi D, Choi S, Koo S-H. Roles of Protein Arginine Methyltransferases in the Control of Glucose Metabolism. Endocrinology and Metabolism2014;29(4):435-40.

6. Li MH, Abrudan JL, Dulik MC, Sasson A, Brunton J, Jayaraman V, et al. Utility and limitations of exome sequencing as a genetic diagnostic tool for conditions associated with pediatric sudden cardiac arrest/sudden cardiac death. Hum Genomics;9:15.

Output

International Journal Publication

Molecular Genetics Study of Autosomal Dominant Type 2 Diabetes in Thai Families by Next Generation Sequencing (In preparation)

Connections with experts

1. Associate Professor Nattachet Plengvidhya, M.D.

   Head of Division of Endocrinology and Metabolism,
   Department of Medicine, Faculty of Medicine Siriraj Hospital,
   Mahidol University, Bangkok, Thailand

   - Mentor
   - Project consultant in clinical and genetics of diabetes aspects

2. Professor Pa-thai Yenchitsomanus, Ph.D.

   Head of Division of Molecular Medicine, Faculty of Medicine, Siriraj Hospital,
   Mahidol University, Bangkok, Thailand

   - Project consultant in molecular genetics of disease

3. Associate Professor Alessandro Doria, MD, Ph.D, MPH.

   Section on Genetics and Epidemiology, Joslin Diabetes Center,
   Harvard Medical School, Boston, Massachusetts, USA.

   - Project consultant in molecular genetics and epidemiology