รายงานโครงการวิจัยฉบับสมบูรณ์

โดรงการการค้นหายีนก่อโรคเบาหวานชนิดโมดีในครอบครัวผู้ป่วยไทย ด้วยวิธี Genome-wide linkage analysis โดยใช้ DNA microarray

โดย

รศ.นพ. ณัฐเชษฐ์ เปล่งวิทยา และคณะ

คณะแพทยศาสตร์ศิริราชพยาบาล

มหาวิทยาลัยมหิดล

เมษายน 2556

สัญญาเลขที่ BRG5280008

รายงานโครงการวิจัยฉบับสมบูรณ์

โดรงการการค้นหายีนก่อโรคเบาหวานชนิดโมดีในครอบครัวผู้ป่วยไทย ด้วยวิธี Genome-wide linkage analysis โดยใช้ DNA microarray

(Genome-wide linkage analysis in Thai families with maturity-onset diabetes of the young (MODY) by DNA microarray)

คณะผู้วิจัยและสังกัด

1.	รศ.นพ. ณัฐเชษฐ์ เปล่งวิทยา	สาขาวิชาต่อมไร้ท่อและเมตะบอลิสม ภาควิชาอายุรศาสตร์
2.	ศ.ดร.เพทาย เย็นจิตโสมนัส	หน่วยอณูเวชศาสตร์ สถานส่งเสริมการวิจัย
3.	รศ.พญ. สุวัฒณี คุคุปติวุฒิ	ภาควิชาสรีรวิทยา
4.	ดร. วทิพย์ ตั้งจิตติโภคิน	ภาควิชาวิทยาภูมิคุ้มกัน
5.	นางนลินี จงเจริญ	สถานส่งเสริมการวิจัย
6.	นางกาญจนา ชาญประเสริฐ	ภาควิชาวิทยาภูมิคุ้มกัน
7.	นางสาวนินารีมาน บินนิมะ	ภาควิชาวิทยาภูมิคุ้มกัน

คณะแพทยศาสตร์ศิริราชพยาบาลมหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

สัญญาเลขที่ BRG5280008

โครงการการ"การค้นหายีนก่อโรคเบาหวานชนิดโมดีในครอบครัวผู้ป่วยไทย ด้วยวิธี Genome-wide linkage analysis **โดยใช้** DNA microarray

รายงานโครงการวิจัยฉบับสมบูรณ์

ชื่อโครงการ (ไทย) การค้นหายีนก่อโรคเบาหวานชนิดโมดีในครอบครัวผู้ป่วยไทยด้วยวิธี Genome-wide linkage analysis โดยใช้ DNA microarray

(English) Genome-wide linkage analysis in Thai families with maturity-onset diabetes of the young (MODY) by DNA microarray

ระยะเวลาโครงการ 3 ปี

ชื่อหัวหน้าโครงการวิจัยผู้รับทุน รศ.นพ.ณัฐเชษฐ์ เปล่งวิทยา

รายงานในช่วงตั้งแต่วันที่ มิถุนายน 2552 – **เมษายน** 2556)

การดำเนินงาน

☑ ได้ดำเนินงานตามแผนที่วางไว้

☐ ได้ดำเนินงานล่าช้ากว่าแผนที่วางไว้

☐ ได้เปลี่ยนแผนงานที่วางไว้

Abstract

Type 2 diabetes (T2D) is a heterogeneous disorder resulting from interaction between genes and environment. The prevalence of diabetes is on the rising tide and the major part of this increase will occur in developing countries and T2D is a major contribution to this alarming figure. Maturity-onset Diabetes of the Young (MODY) is characterized by an autosomal dominant mode of inheritance, an early onset of hyperglycemia (usually less than 25 years). The well-defined mode of inheritance with high penetrance and the early age of onset of diabetes allow the collection of multigenerational pedigrees, making MODY an attractive model for genetic studies. The underlying pathophysiology of MODY is insulin secretion defect usually without insulin resistance. Up till now, defects in at least thirteen genes have been identified as the cause of MODY. These include *HNF-4A*, *HNF-1A*, *HNF-1B*, *GCK*, *IPF-1*, *NeuroD 1/B, KLF11, CEL, PAX4, INS, BLK*, *ABCC8* and *KCNJ11*. Nevertheless, MODY with unknown genetic etiology (MODY-X) in various ethnic groups including Thais is very common. To identify novel MODY genes in Thai patients, we performed genome-wide linkage analysis combined with whole-exome sequencing (WES) in selected MODY-X family. More than 1.8 million markers were genotyped in 27 family members of selected family by using Affymetrix® Genome-Wide Human SNP Array 6.0. Then, we performed WES in two affected and two unaffected relatives in selected family. For linkage analysis, we considered loci with logarithm of odds (LOD) scores > 2.5 as candidate regions for identification of rare variants. For WES, we focused on variants of interest (heterozygous, non-synonymous variants not reported in dbSNP 135 and MAF<0.5 from 1000 Genomes Project and NHLBI) present in the two affected relatives and not present in unaffected relatives. Six novel variations, which are located on candidate regions, were found. Only two novel variants p.M920L in *PTCH1* and p.P1060L in *WNK2* genes show an effect on protein function at least 3 programs by using 4 prediction programs including PolyPhen2, VarioWatch, Mutation Taster and SIFT. Both candidate genes are located on chromosome 9 and show expression in pancreatic beta-cells. They were not observed in 150 controls. Therefore, they may play a role in pancreatic beta-cells function. However, they need further investigation to study their possible role in diabetes. The combination of genome-wide linkage analysis and WES applied in this study provide a powerful tool to identify candidate genes causing MODY in Thais.

บทคัดย่อ

โรคเบาหวานชนิดที่ 2 (Type 2 Diabetes, T2D) เป็นโรคที่เกิดจากการมีปฏิสัมพันธ์ระหว่างยีนและสิ่งแวดล้อม ความชุกของโรคเบาหวานในปัจจุบันมีอัตราที่เพิ่มขึ้นโดยเฉพาะในประเทศกำลังพัฒนา นอกจากนี้ยังมีโรคเบาหวานชนิดที่พบในผู้ป่วยอายุน้อย (Maturity onset diabetes of the young, MODY) มีลักษณะที่สำคัญคือมีการถ่ายทอดการเป็นโรคแบบ autosomal dominant พบระดับน้ำตาลในเลือดสูงเมื่ออายุน้อย (โดยปกติจะน้อยกว่า 25 ปี) มีการถ่ายทอดการเป็นโรคหลายรุ่น ทำให้โรคเบาหวานชนิด MODY เป็นรูปแบบของโรคเบาหวานที่น่าสนใจสำหรับการศึกษาทางพันธุกรรม โดยพยาธิสรีรวิทยาพื้นฐานของ MODY เป็นโรคเบาหวานชนิดที่พบความผิดปกติของการหลั่งอินซูลิน มักจะไม่มี insulin resistance จนถึงขณะนี้ข้อบกพร่องอย่างน้อย 13 ยีนที่ถูกระบุว่าเป็นสาเหตุของ MODY เหล่านี้ คือ HNF-4A, HNF-1A, HNF-1A, GCK, IPF-1, NeuroD 1/B, KLF11, CEL, PAX4, INS, BLK, ABCC8 และ KCNJ11 อย่างไรก็ตามยังพบ MODY ที่ไม่ทราบสาเหตุทางพันธุกรรม (MODY-X) ในกลุ่มชาติพันธุ์ต่าง ๆ รวมทั้งคนไทย คณะผู้วิจัยจึงทำการวิเคราะห์หายีนก่อโรค โดยใช้วิธี linkage analysis ร่วมกับวิธี whole exome sequencing (WES) ในครอบครัว MODY X โดยใช้ marker เพื่อ genotyped มากกว่า 1.8 ล้านชนิดในสมาชิกในครอบครัวของครอบครัวที่เลือกจำนวน 27 คน โดยใช้ Affymetrix SNP array 6.0 โดยใช้ค่า LOD score> 2.5 เป็นในการหา candidate region และใช้วิธี WES ในคนที่เป็นโรค 2 คนและไม่เป็นโรค 2 คน และเลือกเฉพาะความแปรผันที่เป็น heterozygous, ไม่พบรายงานใน dbSNP135 และ MAF<0.5 จากโครงการ 1000 Genomes และ NHLBI ผลการวิเคราะห์พบความแปรผัน 6 ชนิด เมื่อนำมาพยากรณ์ด้วยโปรแกรม 4 โปรแกรมคือ PolyPhen2, VarioWatch, Mutation Taster และ SIFT พบว่าเพียง 2 mutations p.M920L ในยีน PTCH1 และ p.P1060L ในยีน WNK2 ที่น่าจะมีผลต่อการทำงานของโปรตีน โดยพบว่าทั้งสองยีนอยู่บนโครโมโซมที่ 9 และมีการแสดงออกในตับอ่อน Beta-cell นอกจากนี้ยังไม่พบในกลุ่มควบคุมที่ไม่เป็นโรคเบาหวานจำนวน 150 คนอีกด้วย ดังนั้นความแปรผันในยีนที่พบนี้อาจมีบทบาทสำคัญในการทำงานของเบต้าเซลล์ในตับอ่อน อย่างไรก็ตามยังมีจำเป็นที่จะต้องศึกษาต่อไปถึงบทบาทต่อการเป็นโรคเบาหวาน การใช้เทคนิค linkage analysis ร่วมกับการวิเคราะห์จีโนมด้วยวิธี WES น่าจะเป็นเครื่องมือที่มีประสิทธิภาพในการหายีนก่อโรคเบาหวานชนิด MODY ในคนไทยต่อไป

รายละเอียดผลการดำเนินงานของโครงการ

1. Executive summary

โรคเบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสำคัญด้านสาธารณสุขทั่วโลกรวมทั้งประเทศไทย เป็นสาเหตุที่สำคัญของการเสียชีวิตและทุพพลภาพจากโรคแทรกซ้อนปัจจุบันโรคเบาหวานแบ่งออกได้เป็นชนิดต่างๆตามสาเหตุของการเกิดโรค ได้แก่ โรคเบาหวานชนิดที่หนึ่ง (type 1 diabetes [T1D]), โรคเบาหวานชนิดที่สอง (type 2 diabetes [T2D]), โรคเบาหวานที่มีสาเหตุจำเพาะอื่นๆ และโรคเบาหวานขณะตั้งครรภ์ (gestational diabetes mellitus)

โรคเบาหวานชนิด Maturity-Onset Diabetes of the Young (MODY) เป็นโรคเบาหวานที่มีลักษณะทางคลีนิคคล้ายกับ T2D และมีการถ่ายทอดทางพันธุกรรม เป็นแบบโรคยีนเดี่ยวแบบ autosomal dominance แม้ว่าจะมีค้นพบยีนก่อโรคเบาหวานชนิด MODY แล้ว 13 ยีน ในประชากร Caucasians แต่พบว่ายีนดังกล่าวเป็นสาเหตุเพียงส่วนน้อยของโรคเบาหวานชนิด MODY ในประชากร Asians รวมทั้งไทยด้วย ดังนั้นคณะผู้วิจัยจึงได้ตรวจหายีนก่อโรคเบาหวานชนิด MODY ใหม่ด้วยวิธี genome-wide linkage analysis โดยใช้ Affymetrix GeneChip® Human Mapping 10K 2.0 ร่วมกับ Affymetrix Human SNP assay version 6.0 ในการทำ single nucleotide polymorphism (SNP) genotyping โดยทำการศึกษาในครอบครัว MODY ขนาดใหญ่ 1 ครอบครัวที่ตรวจยืนยันแล้วว่าโรคเบาหวานไม่ได้เกิดจากความแปรผันของยีน MODY ที่เคยมีรายงานมาก่อนหรือที่เรียกว่า MODY-X หลังจากวิเคราะห์ผลเบื้องต้นด้วยโปรแกรม Superlink แล้วพบว่า chromosome ที่ให้ค่า LOD score สูงสุดตกอยู่ในบริเวณ chromosome 9 คณะผู้วิจัยจึงได้ทำ sequence ในส่วนของ coding region ทั้งหมดโดยวิธี exome sequencing เพื่อหา candidate gene ที่น่าจะเป็นยีนก่อโรค โดยใช้วิธีเลือกความแปรผันที่น่าจะทำให้เกิดโรคดังนี้ (1) Heterozygous SNPs (2) พบความถี่ในฐานข้อมูล 1000 Genomes หรือ NIEHS SNPs น้อยกว่า 0.05% (3) ไม่พบข้อมูลใน dbSNPs135 รวมไปถึงการวิเคราะห์ผลกระทบของ SNPs ต่อโครงสร้างหรือการทำงานของโปรตีนด้วย 4 โปรแกรม (PolyPhen2, VarioWatch, Mutation Taster, SIFT) พบว่ายีนที่อาจจะเป็นยีนก่อโรคเบาหวานชนิดโมดี คือยีน *PTCH1* และ WNK2 ซึ่งคณะผู้วิจัยต้องทำการศึกษาวิเคราะห์เพื่อดูการถ่ายทอดความแปรผันกับโรคเบาหวานในครอบครัวนี้ต่อไป

2. ผลงานวิจัยที่ทำประกอบด้วย

(1) Objectives

1. To identify novel MODY genes by genome-wide linkage analysis in Thai MODY-X families
2. To demonstrate that the identified genes are MODY genes by genetic and functional studies

(2) Background/rationale of the problem for research and its significance

Diabetes mellitus (DM) is a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both. The chronic hyperglycemia of diabetes is associated with long-term damage, dysfunction, and failure of various organs, especially the eyes, kidneys, nerves, heart, and blood vessels. The cost of medical care especially for chronic complications is immense and has an enormous impact on global public health including Thailand. Moreover, the prevalence of DM is on a rising tide with an estimated prevalence in adult in 2025 is 300 million. The major part of this numerical increase will occur in developing countries and type2 diabetes (T2D) is a major contribution to this alarming figure. It was estimated in 2000 that 9.6% (2.4 millions) of Thai adults were affected with T2D and 5.4% (1.4 millions) had impaired fasting glucose.

A sensible way to address this issue is to understand disease pathophysiology. Since T2D is the most common form of diabetes, it is a foremost target for research. Generally, T2D is recognized as a complex disease resulting from interaction between several genes and environmental factors that trigger disease development. If such genes can be identified it will facilitate cure or more importantly disease prevention. Due to its polygenic nature, attempt to find the culprit genes has been unrewarding. However, 1%-5% of T2D are caused by defects in a single gene (monogenic), in which Maturity-onset Diabetes of the Young (MODY) is the most intensively investigated. MODY is characterized by an autosomal dominant mode of inheritance, an early onset of hyperglycemia (usually less than 25 years), The well-defined mode of inheritance with high penetrance and the early age of onset of diabetes allow the collection of multigenerational pedigrees, making MODY an attractive model for genetic studies. The underlying pathophysiology of MODY is insulin secretion defect usually without insulin resistance. MODY is not a single entity

but is a heterogeneous disease with regard to genetic, metabolic, and clinical features. and a primary defect in insulin secretion. Up till now, defects in at least six genes have been identified as the cause of MODY. These include genes encode transcription factors; Hepatocyte Nuclear Factor (HNF), HNF-4A, HNF-1A and HNF-1B, a gene encode glycolytic enzyme; Glucokinase (GCK) and two genes encode proteins involve in regulation of pancreatic Beta-cell function; Insulin Promotor Factor-1 (IPF-1) and Neurogenic Differentiation 1/B-cell E-box transactivator 2 (NeuroD 1/B 2). By studying MODY, there is breakthrough in understanding of Beta-cell metabolism, neogenesis and turnover. Novel pathways regulating insulin secretion have been identified. New drug for treating diabetes such as glucokinase activator is currently in development.

Sequence variations of known MODY genes are responsible for 85% of MODY in Caucasians. This is in contrast with reported in Asian in which genetic variability of know MODY genes are accounted for 15-20% of early-onset diabetes. Thus, the remaining families, referred to as MODY-X could harbor novel MODY genes. Study at Siriraj Hospital by Siriraj Diabetes Research Group (SiDRG) has revealed that more than 85% of early-onset diabetic patients are MODY-X. These families are precious resources for identifying new MODY genes.

Two tactics have been used to discover new genes. A candidate gene approach, in this case, by looking at genes involve in pancreatic beta-cell function. It can be speculated that one or more of them should play important roles in regulating insulin secretion. Recently, SiDRG have studied Pax4 which encodes Pax4 transcription factor that plays important roles in pancreatic beta-cell development as a candidate gene for MODY. We found possible pathogenic mutations in three

MODY-X probands and showed that one of these mutations decreases transcriptional activity of mutant protein on insulin and glucagon promoters. We conclude that Pax4 mutation can cause diabetes in few Thai MODY-X families. This candidate genes approach is straightforward. However, the outcome may not as fruitful as expect because there is no guarantee of accomplishment even if those certain genes may be very important in regulation of normal β-cell function. Moreover, preexisting knowledge of molecular mechanisms underlying the disease phenotypes is limited and biological molecules or processes contributing to pathogenesis are largely remained unknown. The other method to use is to do the linkage analysis in the remaining MODY-X families to find the promising region on certain chromosomes that may harbor the culprit genes. Genome-wide linkage analysis can break the restriction of bottleneck information, allowing

the discovery of novel molecules which might responsible for new pathogenic pathways. Moreover, this method is commonly used for identification of causative genes in monogenic disorder.

We therefore aim to identify gene causing MODY-X in Thais by using genome-wide linkage analysis. We initially selected the most informative MODY-X family for identification of novel MODY gene by this approach. More than 10,000 single nucleotide polymorphisms (SNPs) were genotyped in 29 family members of the selected family by using Affymetrix 10K SNP microarray. The preliminary result yielded significant log of odd (LOD) scores on certain regions of chromosomes 15 and 16. We are optimistic that gene(s) responsible for our MODY-X should be located on one of these two chromosomal regions. The positional candidate genes in this critical regions are further explored by dense-SNP genotyping and sequencing. In addition if we can recruit more members of other MODY-X families and made them as extended MODY-X families, this approach will also be applied for these families. We will also recruit more MODY families for screening the six known MODY genes and expect that we will obtain a few extended MODY-X families for the genome-wide linkage analysis and candidate gene identification. Once a possible pathogenic mutation in the candidate gene is identified, it will be tested for segregation with diabetes and its impact on protein function will be assessed to confirm the pathogenic effect. This study will primarily generate the wealthy information of genes responsible for MODY specifically to Thai population. Further research in molecular genetics and molecular biology area will promote a better understanding of mechanisms underlying the pathogenesis of the MODY as well as other forms of DM which is not only fulfill an image of complex biological networks maintaining glucose homeostasis but also lead to development of novel methods for therapeutic management of DM.

(3) การดำเนินงานที่ผ่านมา

1. Subject recruitment

Families selected for genome-wide linkage analysis were carefully chosen from 44 MODY families in which probands carried none of possible pathogenic mutation of six known MODY genes, according to the study of MODY subtypes performed by Siriraj Diabetes Research Group (SiDRG). The criteria for selection of informative families including: (i) an extended-mutigeneration (3 generations or more) pedigree, (ii) composed of enough number of affected as well as unaffected members to be studied and (iii) diabetes passed through the family from one parent (uniparental diabetes). Additional members of the recruited MODY-X families and also new MODY families were recruited during this study. A family was selected as a representative family and during the first year of the project we had recruited 3 new MODY families. The study protocol and informed-consent procedures have been approved by Ethic Committee Faculty of Medicine Siriraj Hospital Mahidol University. All subjects were informed for the purpose and extent of the study and asked to sign a consent form before enrollment into this project.

2. Sample collection, laboratory assays, and DNA preparation

Venous blood (~10 ml) and urine samples were collected from affected and unaffected family members. Plasma glucose was determined by glucose oxidase method. Glycosylated hemoglobin (HbA1C), total cholesterol, triglyceride and HDL-cholesterol will be assayed by standard methods. LDL-cholesterol was directly measured or calculated by Friedewald formula (LDL-cholesterol = total cholesterol-HDL-cholesterol-triglyceride/5). Urine microalbumin and creatinine were measured by DCA 2000 Microalbumin/Creatinine Analyzer (Bayer Corporation Elkhart, IN USA) using immunoturbidity method. Ten to 15 ml of peripheral venous blood was collected from each subject into a sterile tube containing 20 µl of 20% EDTA for DNA preparation. The red blood cell (RBC) lysis buffer was used to destroy RBC. Leukocytes were separated and stored in a 15-ml screw-capped tube at –70°C. Standard phenol/chloroform extraction method was used to prepare genomic DNA from the leukocytes.

3. Excluding mutations of six known MODY genes in the selected families

To completely rule out mutation of six known MODY genes as cause of diabetes in selected families, these genes wereexamined by re-sequencing. DNA samples extracted from peripheral

blood leukocytes of probands from a family and newly collected families were subjected to amplifications by polymerase chain reaction (PCR) of each exon with flanking introns, 3'UTR and 5' flanking regions of the six known MODY genes. The sets of primers were shown. The purified PCR products were screened for mutations by direct sequencing. If sequence variations identified are segregated with diabetes, the families which harbor them were excluded from genome-wide linkage analysis.

ตารางที่ 1 แสดง Primersและสภาวะที่เหมาะสมในการทำ PCR ของยีน *HNF4A*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Promoter	HNF4A pro 1-F HNF4A pro 3-R	CCAAGGTCCCAGTTAGTAGACGGTAG CCTCTCTGCCTTCCTTTCAA	26 20	63	779
Ex1a-Ex1c	HNF4A Ex1A-F HNF4A Ex1C-R	GCCAAGACTCCCAGCAGATC CGTGCTGGGCTGAATCGCTG	20 20	63	622
Ex2	HNF4A Ex2-F HNF4A Ex2-R	TCTGTTCTTCCTGAAGCCTCACTC CAAGTGTGCCCATTTCCCAG	24 20	65	260
Ex3	HNF4A Ex3-F HNF4A Ex3-R	CCTAGTTCTGTCCTAAGAGG GTCATAAAGTGTGGCTACAG	20 20	60	253
Ex4-5	HNF4A Ex4-F HNF4A Ex5-R	CAGACACCCCCACCCCCTAC CACTGCCCACTACTGCCCAC	20 21	68	1072
Ex6	HNF4A Ex6-F HNF4A Ex6-R	GCGTCACTGAGTTGGCTACGG GCTAGGCATACCCTCCCTGGAG	22 22	55	210
Ex7	HNF4A Ex7-F HNF4A Ex7-R	CCCACAGGCACCAGCTATCTTG AGCGTTCTGGAGAGAGAGTCAGG	23 23	65	306
Ex8	HNF4A Ex8-F HNF4A Ex8-R	TTGCCCACCCTCTTCCATTG TCCCCACTCCAACCCCGCCC	20 20	65	300
Ex9	HNF4A Ex9-F HNF4A Ex9-R	TCTGCATCCCAGACTCTCCATC AGCCCCATCCTCACCCTTTG	22 20	58	246
Ex10	HNF4A Ex10-F HNF4A Ex10-R	CATTTACTCCCACAAAGGCT GACCACGTGATCACCAGGTG	20 20	55	277

ตารางที่ 2 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *GCK*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Promoter	GCK pro 1-F GCK pro 3-R	TCTGGTTCAGATTTCAAGGAGA CTCCTGGTTGTGTTGAGCTGTG	22 22	64	747
Pro-Ex1A	GCK pro 4-F GCK Ex 1-R	GGACACTAAGCCCCACAGCTCA TCAGATTCTGAGGCTCAAAC	22 20	66	358
Ex1B	GCK Ex1B-F GCK Ex1B-R	GCCCTTTCAGTGCAGAAGCC GCTCTCCCAGTGCAAAGTCCC	20 21	68	260
Ex1C	GCK Ex1C-F GCK Ex1C-R	AAGGACTGTCTCTGTACTGATGGCTC TTCTGAAGGGAGGTGGGAGG	26 20	65	236
Ex2	GCK Ex2-F GCK Ex2-R	TGCAGATGCCTGGTGACAGC CACAGCTGCTTCTGGATGAG	20 20	68	290
Ex3	GCK Ex3-F GCK Ex3-R	TAATATCCGGGCTCAGTCACC CTGAGATCCTGCATGGCCTTG	20 20	65	298
Ex4	GCK Ex4-F GCK Ex4-R	TAGCTTGGCTTGAGGCCGTG TGAAGGCAGAGTTCCTCTGG	20 20	65	272
Ex5-6	GCK Ex5-F GCK Ex6-R	TGCCTCCAGTATATGTTAGCAGCC CAGGCTCTGCTCTGACATCACC	23 22	70	476
Ex7	GCK Ex7-F GCK Ex7-R	AGTGCAGCTCTCGCTGACAG CATCTGCCGCTGCACCAGAG	20 20	70	286
Ex8	GCK Ex8-F GCK Ex8-R	TGCCTGCTGATGTAATGGAC TGAGACCAAGTCTGCAGTGC	20 20	63	262
Ex9-10	GCK Ex9-F GCK Ex10-R	CACTCAGCGACCGCCCTACC TGTGGCATCCTCCCTGCGCT	20 20	65	664

ตารางที่ 3 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *HNF1A*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Promoter	HNF1A pro 2-F HNF1A pro 1-R	TCCCATCGCAGGCCATAGCTC CCGTCTGCAGCTGGCTCAGTT	20 21	65	615
Ex1	HNF1A Ex1-F HNF1A Ex1-R	GGCAGGCAAACGCAACCCACG GAAGGGGGGCTCGTTAGGAGC	21 21	63	483
Ex2	HNF1A Ex2-F HNF1A Ex2-R	CATGCACAGTCCCCACCCTCA CTTCCAGCCCCCACCTATGAG	21 21	63	390
Ex3	HNF1A Ex3-F HNF1A Ex3-R	GGGCAAGGTCAGGGGAATGGA CAGCCCAGACCAAACCAGCAC	21 21	70	304
Ex4	HNF1A Ex4-F HNF1A Ex4-R	CAGAACCCTCCCCTTCATGCC GGTGACTGCTGTCACTGGGAC	21 21	63	397
Ex5-6	HNF1A Ex5-F HNF1A Ex6-R	GGCAGACAGGCAGCTGGCCTA GTTGCCCCATGAGCCTCCCAC	21 21	70	668
Ex7	HNF1A Ex7-F HNF1A Ex7-R	GGTCTTGGGCAGGGGTGGGAT CTGCAATGCCTGCCAGGCACC	21 21	70	347
Ex8	HNF1A Ex8-F HNF1A Ex9-R	GAGGCCTGGGACTAGGGCTGT CGGACAGCAACAGAAGGGGTG	21 21	70	523
Ex10	HNF1A Ex10-F HNF1A Ex10-R	GTACCCCTAGGGACAGGCAGG ACCCCCCAAGCAGGCAGTACA	21 21	70	248

ตารางที่ 4 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *IPF1*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Enhancer	IPF enhancer-F IPF enhancer-R	GCCGCAGACAATGGACTC AGATGCCCTTGCTGTCACC	18 19	55	182
Pro-Ex1-2	IPF promoter-F IPF Ex1-2-R	GCCTAGCCTCTTAGTGCG CGCTTGGAGGTAAGGCGG	18 18	63	745
Ex2-1-2.2	IPF Ex2-1-F IPF Ex2-2-R	TGGGGGCTGTGCGGGGCTC CCGAGTGGTTGAAGCCCCTCAG	19 21	63	659

ตารางที่ 5 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *HNF1B*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Pro-Ex1-2	HNF1B pro 1-F HNF1B Ex1-2-R	CTGCAAGGCACTGGCTTAAC GAGTGTGGTCGGGCGCAGTG	20 20	69	883
Ex2	HNF1B Ex2-F HNF1B Ex2-R	CTCCCACTAGTACCCTAACC GAGAGGGCAAAGGTCACTTCAG	20 22	56	291
Ex3	HNF1B Ex3-F HNF1B Ex3-R	AGTGAAGGCTACAGACCCTATC TTCCTGGGTCTGTGTACTTGC	22 21	56	365
Ex4	HNF1B Ex4-F HNF1B Ex4-R	CCCTCACTCACCATCTCCCCTCCA CCGAGGCAGTGAGGCCCAAC	24 20	60	301
Ex5	HNF1B Ex5-F HNF1B Ex5-R	TGCCGAGTCATTGTTCCAGG CCTCTTATCTTATCAGCTCCAG	20 22	58	276
Ex6	HNF1B Ex6-F HNF1B Ex6-R	CTGCTCTTTGTGGTCCAAGTCC GAGTTTGAAGGAGACCTACAG	22 21	56	288
Ex7	HNF1B Ex7-F HNF1B Ex7-R	ATCCACCTCTCCTTATCCCAG ACTTCCGAGAAAGTTCAGACC	21 21	56	341
Ex8	HNF1B Ex8-F HNF1B Ex8-R	TTTGCCTGTGTATGCACCTTG GCCGAGTCCATGCTTGCCAC	21 20	60	257
Ex9	HNF1B Ex9-F HNF1B Ex9-R	CTTTGCTGGTTGAGTTGGGC TTCCATGACAGCTGCCCAGAG	20 21	58	208

ตารางที่ 6 แสดง Primersและสภาวะที่เหมาะสมในการทำ PCR ของยีน *NEUROD1*

Fragment	Primer	Nucleotide sequence (5'→3')	No. of nucleotides	Ta (°C)	Product size (bp)
Promoter	NeuroD pro 1-F NeuroD pro 3-R	GCTTTTCCCTTCCTTCCCTC ACACACTCTCGCAAACGCAC	20 20	64	666
Ex2-1-2.3	NeuroD Ex2-1-F NeuroD Ex2-3-R	CAAGCATTTGTACAGGTTTAG CTGTAAGCACAGTGGGTTCG	21 20	56	825

ตารางที่ 7 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *PAX4*

Fragment	Primer name	Nucleotide sequence 5'>3'	No. of nucleotides	Ta (°C)	Product size (bp)
PAX4-Pro1	Pro PAX4 F1-F	GACTTCCCTACAGACACG	18	58	700
	Pro PAX4 F1-R	CTACTTAGCCATGCTTGG	18		
PAX4-Pro2	Pro PAX4 F2-F	GGTCAGGACATAAGCCTT G	19	65	622
	Pro PAX4 F2-R	GGATGCTGAAAGTGAAACA	19		
PAX4-Pro3	Pro PAX4 F3-F	GGT TTCACAGACTCTGAGG	19	63	610
	Pro PAX4 F3-R	CGTGTGTGTTGCTATAAAAG	20		
PAX4-Pro4	Pro PAX4 F4-F	GGTTAGGACACTCTCTGG	18	63	644
	Pro PAX4 F4-R	GTGATGATCTGGTCTTGG	18		
PAX4-Ex1-3	Pax4 Ex1-F	AGGTGGTGTGTGGATACCTC	20	70	1128
	Pax4 Ex3-R	GAGATTTGGCTGTGATTAGCCC	22		
PAX4-Ex4-5	Pax4 Ex4-F	CTGACCAGAGGAATCACCATC	21	70	539
	Pax4 Ex5-R	GGCCCAGACTCTTCCTCCTTG	21		
PAX4-Ex6	Pax4 Ex6-F	GATCAGCAGGTGACAGGCAGC	21	69	174
	Pax4 Ex6-R	AGATGACTGAGCGGGCAGATG	21		
PAX4-Ex7-9	Pax4 Ex7-F	AGTGGCTGACTTTCCTAGAAC	21	70	1112
	Pax4 Ex9-R	GTAAGGACAATGGGCAGGATG	21		

ตารางที่ 8 แสดง Primers และสภาวะที่เหมาะสมในการทำ PCR ของยีน *INS*

Fragment	Primer name	Nucleotide sequence 5'>3'	No. of nucleotides	Ta (°C)	Product size (bp)
INS-Promoter	INS promoter-F	CCCGCCCTGCAGCCTCCAGC	20	70	437
	INS-promoter-R	TGGAATCCTGAGCCCACCTGACGC	24		
INS-Exon1-2	INS-exn1-2-F	CGGCAGGGGTTGAGAGGTAGGGGA	24	68	557
	INS-exon1-2-R	CATGCTGGGTGGGAGCGCCA	20		
INS-Exon3	INS-exon3-F	GGTGACCCTCCCTCTAACCT	20	63	529
	INS-exon3-R	CTCCCATACTGGACCCTGAG	20		

4. SNP genotyping by Affymetrix GeneChip Human Mapping 10K

Affymetrix GeneChip Human Mapping 10K (The GeneChip® Human maping 10K Array Xba 142 2.0) is used for genome-wide linkage analysis, according to the methods described by the manufacturer. Briefly, 250 ng of genomic DNA is digested with XbaI restriction enzymes and ligated to adaptors that recognize the overhangs. The ligated DNA is amplified under recommended conditions, using primers complementary to the universal adapter. PCR product then be fragmented, labeled with biotinylated ddATP using terminal deoxynucleotidyl transferase and hybridized to the chip. After hybridization, washing, and signal development in the Affymetrix Fluidic System, the chips are then scanned by a laser scanner to acquire data for examination of a quality control (QC) call rate and genotype calls.

5. SNP genotyping by Affymetrix GeneChip Human Mapping version 6.0

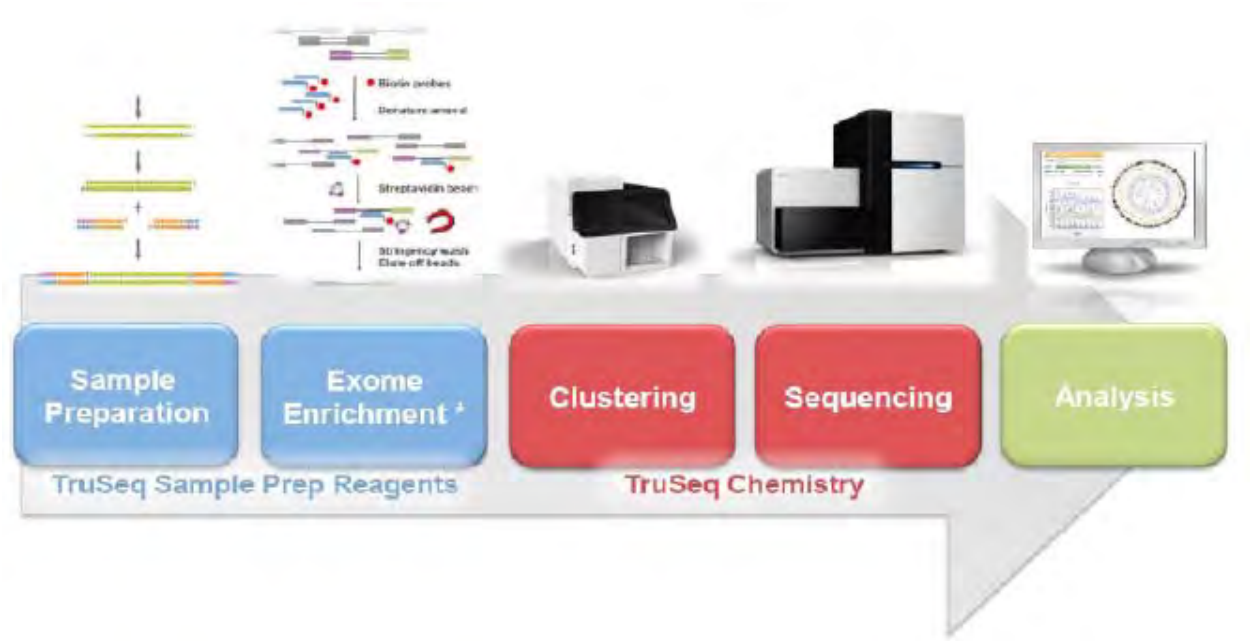
Twenty-seven individuals from selected MODY-X family will be genotype with the Affymetrix® Genome-Wide Human SNP Array 6.0 according to the method described by the manufacturer. Briefly, 250 ng of genomic DNA is digested with the restriction enzyme Sty I and Nps I and ligated to a common adaptor with T4 DNA ligase. Following ligation, the template undergoes PCR amplification using TITANIUM™ Taq DNA polymerase. Then, the Sty and Nsp PCR products are pooled to a single deep well pooling plate and undergo purification. The purified PCR products are then fragmented with Fragmentation Reagent (DNAseI) and end-labeled using terminal deoxynucleotidyl transferase and hybridized to the chips. After hybridization, the chips are washed and signals are developed in the Affymetrix Fluidic System. The chips are then scanned by laser scanner to acquire data for a quality control (QC) call rate and genotype calls.

6. Candidate region identification from linkage analysis

The genotype data generated from Affymetrix® Genome-Wide Human SNP Array 6.0 together with pedigree and marker information were used to identify candidate regions that show significant linkage with diabetes in the families by using the easyLINKAGE Analysis Program (easyLINKAGE v5.08 beta).

7. Variants identification by whole-exome sequencing (WES)

Samples were prepared as an Illumina sequencing library, and in the second step, the sequencing libraries were enriched for the desired target using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencer



รูปที่ 1 Experiment overview of next generation sequencing

8. Candidate variants selection

Approximately 15,000 to 20,000 coding variants were identified from WES. Thus, filtering criteria must be applied to select suitable candidate variants. Since MODY is an autosomal dominant disease, heterozygous-nonsynonymous variants (introduce an amino acid change) presented only in diabetic patients were explored first. Moreover, variants which show minor allele frequency (MAF) more than 5% in 1000 Genome and National Institute of Environmental Health Sciences (NIEHS) Exome Project were also excluded because MODY is a rare Mendelian disorder. Under the assumption that variant responsible for Mendelian disorders should not be presented in publicly available databases of human genetic variations, we then, excluded variants in dbSNP135 database. To select putative variants, remaining variants located in candidate regions resulted from linkage analysis were selected. Then, the putative variants were studied further for effects on

protein expression and function by *in silico* programs (PolyPhen V2.0.23, SIFT, Mutation Taster and VarioWatch) to help select proper candidate variants.

9. Segregation analysis by genotyping in selected MODY-X family

The selected candidate genes were tested for segregation with diabetes in the family by Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) method. Novel variant(s), which showed segregation with diabetes in the family, was validated by Sanger sequencing.

10. Additional genotyping of identified novel variation(s)

Novel variation(s) were analyzed further in 200 non-diabetic controls and in other MODY-X probands by PCR-RFLP method to confirm an association of these novel variants with diabetes.

(4) Scope of the research work:

An extended, informative multi-generation MODY-X family has been studied by genome-wide linkage analysis using Affymetrix 10K Microarray. A preliminary result reveals high LOD score in two different chromosomal regions. Extensive search for candidate genes in these critical regions are being performed by dense-SNP genotyping and direct sequencing. This approach will also be applied to other one or two MODY-X families, if extended; informative multi-generation families can be recruited. The identified candidate gene harboring deleterious mutation will be tested for segregation and linkage with diabetes within the affected family. The effect of mutation of the functional change of the encoded protein will be analyzed by cDNA cloning, site-directed mutagenesis, protein expression, and appropriate functional assay.

(5) Equipment

Basic equipment for molecular genetics, molecular and cellular biology techniques are available and allowable to access at the Division of Medical Molecular Biology, Department of Research and Development, Department of Immunology, and Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University. Affymetrix GeneChip Microarray System including hybridization oven, automated fluidics station, laser scanner, PC provided with system, and analysis

and control software were lent by RI Technologies, Bangkok. The system is also available at Ramathibodi Hospitatl, Mahidol University, Bangkok.

(6) Expected benefits

This study will generate the information of novel genes responsible for MODY in Thai population, which may suggest new molecular pathophysiology pathway of diabetes. The forthcoming paper will be the very first report regarding new MODY subtype in Thailand and South East Asia. It will promote better understanding of mechanisms underlying molecular pathogenesis of MODY and perhaps, of other types of DM which is very useful for clinical implication, treatment, and prevention of the disease.

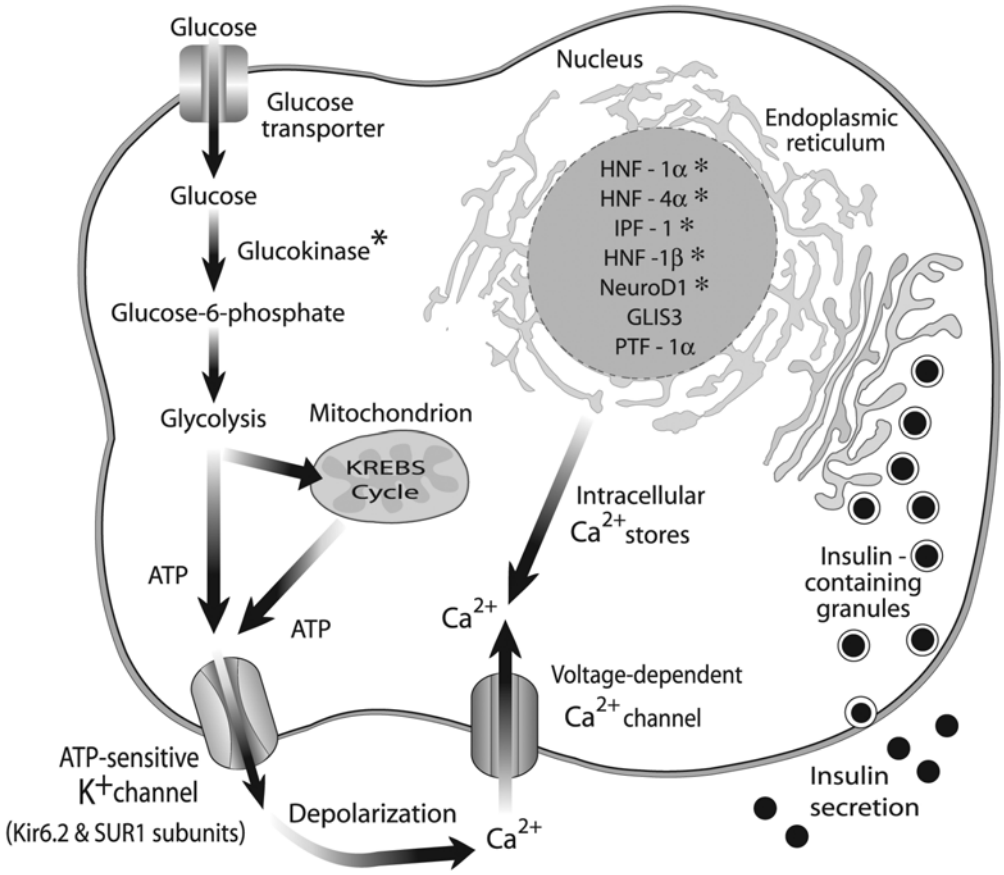
(7) Literature review and references:

Diabetes mellitus (DM) is a serious threat to global public health problem. The prevalence of DM has increased sharply in recent decades. The number of affected individual worldwide is currently estimated at 150 million and will reach 220 million by 2010 and 300 million by 2025 (1, 3). In Thailand, the most recent survey was conducted in 2000 and showed that 9.6% (2.4 millions) of Thai adults were affected with DM and 5.4% (1.4 millions) had impaired fasting glucose (2). DM is characterized by chronic hyperglycemia resulting from defects in insulin secretion, insulin action or a combination of both. DM is classified into four etiologic types. Type 1 diabetes (T1D) and type 2 diabetes (T2D) contribute to the majority of cases. Both are complex diseases in which several genetic abnormalities together with environmental triggering are required for the development of disease. While T1D is mainly caused by autoimmune-mediated destruction of ?-cells and usually occurs in children, T2D is caused by an inability of -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. However, these types of diabetes have been study extensively, in particular Maturity-Onset Diabetes of the Young (MODY). This special form of diabetes has genetic defects of beta-cell function (4, 5). The diagnosis of MODY is based upon the following criteria: (i) early onset of

diabetes (usually before 25 years), (ii) autosomal dominant inheritance, (iii) rarely obese and nonketotic diabetes, and (iv) diabetes resulting from beta -cell dysfunction.

Since MODY is a monogenic disease, dissection of its genetic determinant is straightforward. To date, mutations in six different genes are known to cause MODY. These genes encode the glucokinase enzyme (associated with MODY2) (6) and the transcription factors expressed in pancreatic beta-cells, including hepatocyte nuclear factor (HNF)-4A (MODY1), (7), HNF-1A (MODY3) (8), insulin promotor factor (IPF)-1 (MODY4) (9), HNF-1B (MODY5), and neurogenic differentiation 1/ B-cell E-box transactivator 2 (NeuroD1/BETA2, MODY6) (10). Investigation in knockout mice and humans indicated that these transcription factors play coordinately roles in embryonic development of pancreas and final differentiation to beta-cells. In addition, they are involved in normal beta-cell functions by regulating gene expression in fully differentiated beta-cells. Once glucose is transported into the beta-cell via a specific glucose-transporter protein (GLUT2) on beta-cell cell membrane, it is catalyzed into glucose-6-phosphate by glucokinase, a rate-limiting enzyme in glycolysis pathway before passing through the sequential steps of energy production. In turn, increasing of ATP and ADP ratio inhibits and closes the ATP-sensitive potassium channels, leading to depolarization of plasma membrane. As a result, membrane depolarization opens the voltage-dependent calcium channels. Increased intracellular calcium elicits movement of insulin-containing secretory vesicles to the plasma membrane and insulin is then secreted into the circulation (Figure 1).

Hyperglycaemia in MODY2 patients appears to be resulted from a reduction in the activity of glucokinase which leads to decreased beta-cell sensitivity to glucose. Since HNF-4A (MODY1) regulates genes involved in glucose transport and glycolysis (11) the underlying pathophysiology of MODY1 is described as an impairment of glucose-stimulated insulin secretion. Because HNF-1A expression is regulated by HNF-4A, the underlying pathophysiology associated with MODY3 (HNF-1A mutations) is occurred in the same manner. HNF proteins also play role in liver and kidney in tissue-specific gene expression.



รูปที่ 2 The glucose stimulated-insulin secretion in pancreatic beta cell. Proteins associated with MODY are marked with asterisk (*) [modified from Fajans et al].

Therefore, mutations of HNF-1A, HNF-4A and HNF-1B are associated with abnormalities in liver and kidney functions. Study of MODY4 (IPF-1 mutation) is based on information from a single family whose proband was a homozygotes of the mutant. This infant had neonatal diabetes and pancreatic exocrine insufficiency resulting from pancreatic agenesis (9). Due to its seldom occurrence, the molecular mechanism associated with MODY6 (NeuroD1/BETA2 mutation) is rarely examined. By studying MODY, there is breakthrough in understanding of beta-cell metabolism, neogenesis and turnover. Novel pathways regulating insulin secretion have been identified. New drug for treating diabetes such as glucokinase activator is currently under development.

The prevalence of each MODY subtype varies among various ethnic groups. MODY2 are the most common cause of MODY in France, accounting for more than 60% of studied families (6), whereas the prevalence of MODY2 in United Kingdom and Germany were 11% (12) and 8% (13), respectively. In general, MODY3 is the most common cause of MODY in Caucasians and the prevalence varies from 21% to 64%. The other four types of MODY are rare. Additionally, MODY with defects of unknown genes, MODY-X does exist and represents 16-45% of MODY cases in Caucasians. The prevalence of MODY-X in Asians is even more frequent. It was estimated that as many as 60-80% of Chinese (14) and Japanese (15) MODY patients are MODY-X.

In Thais, Siriraj Diabetes Research Group (SiDRG) has recruited MODY and early-onset T2D patients to investigate whether genetic variations of the six known MODY genes are associated with diabetes in the families (manuscript in press). Seven mutations (five are novel) in four known MODY genes (HNF-4A, GCK, HNF-1A, and NeuroD) are identified in 7 of 51 patients being studied. This result showed that variants of six known MODY genes account for a small proportion of both classic MODY (19%) and early-onset type 2 diabetes patients (10%). It is suggested that the majority of MODY in Thais are caused by unknown genes, waiting to be identified.

Two techniques, candidate gene approach and genome-wide scan, are commonly used for identification of novel genes associated with monogenic disorders. In the first approach, candidate genes are selected from an existing knowledge of their biological functions that are presumed to be related with disease phenotype. GCK, HNF-1A, IPF-1, and NeuroD1 are the examples of candidate genes rising from its biological functions. GCK encodes glucokinase, a glycolytic enzyme that acts as a glucose sensor in pancreatic beta-cells and plays an important role in the regulation of insulin

secretion. Heterozygous mutations in GCK result in MODY2 phenotypes (6) while homozygous mutations result in permanent neonatal diabetes mellitus (PND) (16). Transgenic mice with IPF-1 disruption expressed pancreatic agenesis which is associated with MODY4 phenotype (17, 18). HNF-1A encodes transcription factor that forms a heterodimer with HNF-1A, protein that associated with MODY3. It was considered as an excellent candidate gene, and mutations in HNF-1A gene were subsequently described as MODY5 (19). NeuroD1 encodes transcription factor that functions as a regulatory switch for endocrine pancreatic development. Mutations in NeuroD1 were formerly found to be associated with T2D and subsequently be defined as a MODY6 gene (10, 20). Recently, SiDRG has investigated whether Pax4, a gene encoding Pax4 transcription factor that plays an important role in beta-cell development is linked to MODY-X in Thais (21). Forty-four Thai MODY-X probands were screened for mutations in Pax4. Three possible pathogenic mutations were discovered and one of them, R164W, segregated with diabetes in the family. The mutant protein showed less transcriptional repressor activity compared to the wild type. We therefore concluded that mutations in Pax4 are linked to diabetes in few Thai MODY-X

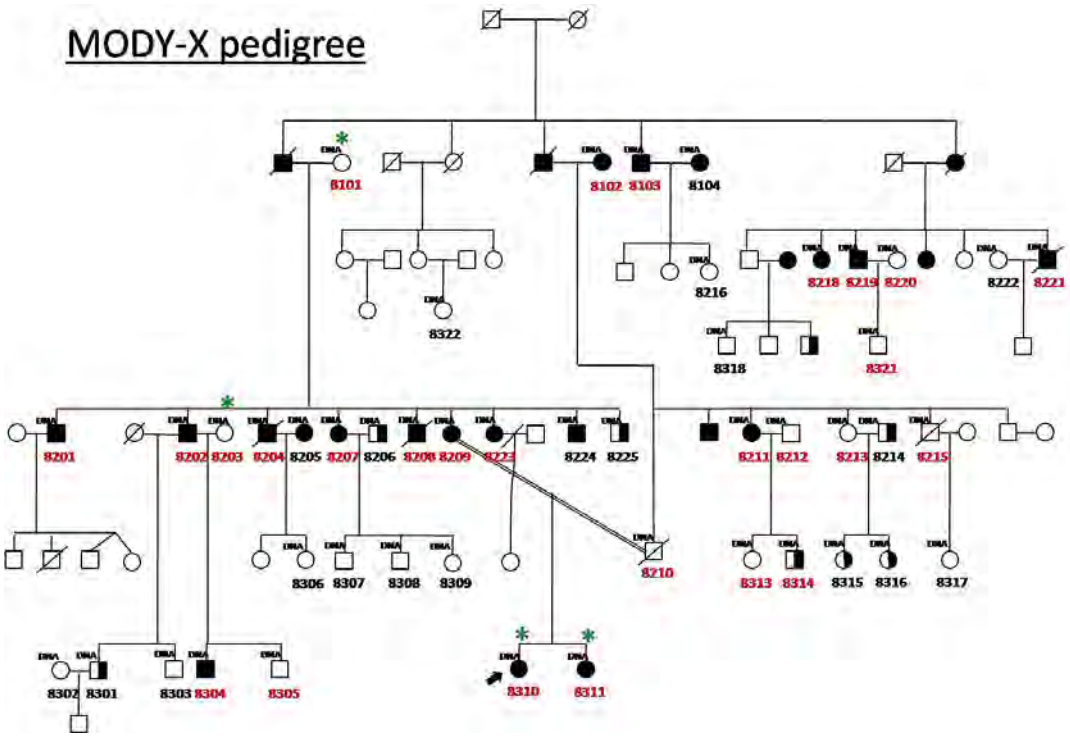
Unlike candidate gene approach, genome-wide scan is an independent way to identify new genes without any background information. By using this approach, novel molecules underlining disease pathways would be identified. Genome-wide scan can be performed by association study or linkage analysis. Association study is a population-based study which is appropriate for studying complex diseases using available case-control cohort. For monogenic disorder like MODY in which the disease has Mendelian pattern of inheritance, genome-wide linkage analysis is a more suitable way. Linkage analysis is a family-based study. The concept is that the region of genome where genetic abnormality is located should be co-inherited from common ancestor to the affected members more frequency than would be expected by chance. Practically, the transmission of genomic regions from parents to children can be tracked with genetic markers or genetic variations that can be specifically located in genome and are polymorphic in studied populations. The linkage between trait and marker locus at frequency more than would be expected by chance is determined by LOD (log10 of ODD ratio) score >3. This score represents the probability that there is a linkage between marker and disease locus over the probability of no linkage between the two loci. Therefore, LOD score more than 3 means that the probability of linkage between 2 loci is 1000 (103) times more than probability of random co-segregation (22). The genomic region that is linked

to the disease loci may cover more than hundreds or thousands kilobases. Fine mapping with another set of genetic markers is required for narrowing the regions of interest. Genes in this region, or positional candidate genes, are identified by positional cloning. HNF-4A and HNF-1A which are responsible for MODY1 and MODY3 (23, 24) are the examples of disease related genes identified by positional cloning. Using the advantage of a complete Human Genome Project, the positional candidate genes in the regions of interest can be now directly identified from the genome sequence databases.

Even though there are several types of genetic markers, single nucleotide polymorphisms (SNPs) that are distributed throughout the genome are the most commonly used genetic markers. They are considered as powerful genetic markers for genome-wide scan because of their abundance, stability and relative ease of scoring (25). Technology advancement has made possible to conduct genome-wide mapping by genotyping of 500,000 SNPs or more in a single array using microarray. Up till now, several commercial platforms of SNPs-microarray are available. Affymetrix GeneChip Human Mapping 10K is one of the most popular platforms for mapping disease genes by linkage analysis. More than 10,000 SNPs with means of intermarker and genetic gap distances of 210 kb and 0.32 cM, respectively, can be genotyped (26). Identification of mutation of VPS33B as a cause of arthrogryposis-renal-dysfunction-cholestatis (ARC) syndrome (27), TSPYL nonsense mutation in sudden infant death with dysgenesis of testes syndrome (SIDDT) (28), and LRP6 mutation in a family with early coronary disease (29) are examples of mapping disease genes with this technique. Even though the culprit genes have not been discovered in some studies, several novel disease-linked regions were localized. These include a novel neonatal diabetic susceptible region on chromosome 10p12.1-p13 (30). Other important findings that were facilitated by using Affymetrix GeneChip Human Mapping 10K are shown in Appendix.

Since the majority of MODY in Thais are MODY-X and using candidate genes approach to find new disease-related genes is straightforward. The outcome may not as fruitful as expected because there is no guarantee of accomplishment even if those certain genes may be very important in regulation of normal beta-cell function. Moreover, pre-existing knowledge of molecular mechanisms underlying the disease phenotypes is limited and biological molecules or processes' contributing to pathogenesis is largely remained unknown. We therefore aim to identify gene causing MODY-X in Thais by genome-wide linkage analysis. We carefully selected an informative

MODY-X family for performing this task. The most informative family chosen is characterized by (i) an extended-mutigeneration (3 generations) pedigree, (ii) composed of enough affected as well as unaffected individuals (11 are affected with DM, 2 are classified as impaired fasting glucose, and 16 are unaffected) and (iii) no bi-parental diabetes. The pedigree is shown in Figure 2. The power to detect linkage in this family was estimated by simulation that an allele of markers co-segregates with the disease in pedigree and calculating an electronic LOD score (ELOD) by SLINK, a general simulation program for linkage analysis. The selected pedigree yielded ELOD of 4.73 which indicated that this family is promising to be used in linkage analysis.



รูปที่ -3 Pedigree of MODY-X family being in this studied. The filled and half-filled boxes (and circles) represent individuals with diabetes and impaired fasting glucose, respectively. The unfilled boxes and circles represent males and females without diabetes. There are 26 affected by diabetes and 21 unaffected who DNA were collected. Twenty seven samples, which were genotype with Affymetrix® Genome-Wide Human SNP Array 6.0, are indicated by red color of ID number. Two affected and two unaffected, who were performed by WES, are indicated with a star. Proband is indicated with black arrow.

Once a gene is located and a possible pathogenic mutation is identified, it will be examined for segregation with diabetes within the families. Impact on protein function will be investigated in appropriate system. This study would primarily generate the information of novel genes responsible for MODY in Thai population. Further studies in genetic and molecular biology will promote the better understanding of mechanisms underlying the pathogenesis of the MODY and perhaps of other types of DM which is very useful for clinical implication, treatment, and prevention of the disease.

REFERENCES

1. Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. Nature. 2001 Dec 13;414(6865):782-7.
2. Aekplakorn W, Stolk RP, Neal B, Suriyawongpaisal P, Chongsuvivatwong V, Cheepudomwit S, et al. The prevalence and management of diabetes in Thai adults: the international collaborative study of cardiovascular disease in Asia. Diabetes Care. 2003 Oct;26(10):2758-63.
3. Wolford JK, Vozarova de Courten B. Genetic basis of type 2 diabetes mellitus: implications for therapy. Treat Endocrinol. 2004;3(4):257-67.
4. Fajans SS, Bell GI, Polonsky KS. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. N Engl J Med. 2001 Sep 27;345(13):971-80.
5. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care. 2003 Jan;26 Suppl 1:S5-20.
6. Pearson ER, Velho G, Clark P, Stride A, Shepherd M, Frayling TM, et al. beta-cell genes and diabetes: quantitative and qualitative differences in the pathophysiology of hepatic nuclear factor-1alpha and glucokinase mutations. Diabetes. 2001 Feb;50 Suppl 1:S101-7.
7. Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature. 1996 Dec 5;384(6608):458-60.
8. Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). Nature. 1996 Dec 5;384(6608):455-8.

9. Wright NM, Metzger DL, Borowitz SM, Clarke WL. Permanent neonatal diabetes mellitus and pancreatic exocrine insufficiency resulting from congenital pancreatic agenesis.Am J Dis Child. 1993 Jun;147(6):607-9.
10. Kristinsson SY, Thorolfsdottir ET, Talseth B, Steingrimsson E, Thorsson AV, Helgason T, et al. MODY in Iceland is associated with mutations in HNF-1alpha and a novel mutation in NeuroD1. Diabetologia. 2001 Nov;44(11):2098-103.
11. Stoffel M, Duncan SA. The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism. Proc Natl Acad Sci U S A. 1997 Nov 25;94(24):13209-14.
12. Beards F, Frayling T, Bulman M, Horikawa Y, Allen L, Appleton M, et al. Mutations in hepatocyte nuclear factor 1beta are not a common cause of maturity-onset diabetes of the young in the U.K. Diabetes. 1998 Jul;47(7):1152-4.
13. Lindner TH, Cockburn BN, Bell GI. Molecular genetics of MODY in Germany.Diabetologia. 1999 Jan;42(1):121-3.
14. Ng MC, Cockburn BN, Lindner TH, Yeung VT, Chow CC, So WY, et al. Molecular genetics of diabetes mellitus in Chinese subjects: identification of mutations in glucokinase and hepatocyte nuclear factor-1alpha genes in patients with early-onset type 2 diabetes mellitus/MODY. Diabet Med. 1999 Nov;16(11):956-63.
15. Nishigori H, Yamada S, Kohama T, Utsugi T, Shimizu H, Takeuchi T, et al. Mutations in the hepatocyte nuclear factor-1 alpha gene (MODY3) are not a major cause of early-onset non-insulin-dependent (type 2) diabetes mellitus in Japanese. J Hum Genet. 1998;43(2):107-10.
16. Glaser B, Kesavan P, Heyman M, Davis E, Cuesta A, Buchs A, et al. Familial hyperinsulinism caused by an activating glucokinase mutation. N Engl J Med. 1998 Jan 22;338(4):226-30.
17. Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet. 1997 Oct;17(2):138-9.
18. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. Nat Genet. 1997 Jan;15(1):106-10.

19. Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet. 1997 Dec;17(4):384-5.
20. Malecki MT, Jhala US, Antonellis A, Fields L, Doria A, Orban T, et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. Nat Genet. 1999 Nov;23(3):323-8.
21. Plengvidhya N, Kooptiwut S, Songtawee N, Doi A, Furuta H, Nishi M, et al. PAX4 mutations in Thais with maturity onset diabetes of the young. J Clin Endocrinol Metab. 2007 Jul;92(7):2821-6.
22. Sudberly P. Human Molecular Genetics. 1 ed. Essex: Addison Wesley Longman Limited; 1998.
23. Njolstad PR, Sovik O, Cuesta-Munoz A, Bjorkhaug L, Massa O, Barbetti F, et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. N Engl J Med. 2001 May 24;344(21):1588-92.
24. Vaxillaire M, Boccio V, Philippi A, Vigouroux C, Terwilliger J, Passa P, et al. A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. Nat Genet. 1995 Apr;9(4):418-23.
25. Hwang JS, Shin CH, Yang SW, Jung SY, Huh N. Genetic and clinical characteristics of Korean maturity-onset diabetes of the young (MODY) patients. Diabetes Res Clin Pract. 2006 Oct;74(1):75-81.
26. Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res. 2004 Mar;14(3):414-25.
27. Gissen P, Johnson CA, Morgan NV, Stapelbroek JM, Forshew T, Cooper WN, et al. Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome. Nat Genet. 2004 Apr;36(4):400-4.
28. Puffenberger EG, Hu-Lince D, Parod JM, Craig DW, Dobrin SE, Conway AR, et al. Mapping of sudden infant death with dysgenesis of the testes syndrome (SIDDT) by a SNP genome scan

and identification of TSPYL loss of function. Proc Natl Acad Sci U S A. 2004 Aug 10;101(32):11689-94.

29. Mani A, Radhakrishnan J, Wang H, Mani A, Mani MA, Nelson-Williams C, et al. LRP6 mutation in a family with early coronary disease and metabolic risk factors. Science. 2007 Mar 2;315(5816):1278-82.

30. Sellick GS, Garrett C, Houlston RS. A novel gene for neonatal diabetes maps to chromosome 10p12.1-p13. Diabetes. 2003 Oct;52(10):2636-8.

(8) Result

1. Subject recruitment

Families selected for genome-wide linkage analysis were carefully chosen from 7 MODY families in which probands carried none of possible pathogenic mutation of six known MODY genes, according to the study of MODY subtypes performed by Siriraj Diabetes Research Group (SiDRG). The criteria for selection of informative families including: (i) an extended-mutigeneration (3 generations or more) pedigree, (ii) composed of enough number of affected as well as unaffected members to be studied and (iii) diabetes passed through the family from one parent (uniparental diabetes). Additional members of the recruited MODY-X families and also new MODY families were recruited during this study. A family was selected as a representative family and during the first year of the project we had recruited 3 new MODY families. The study protocol and informed-consent procedures have been approved by Ethic Committee Faculty of Medicine Siriraj Hospital Mahidol University. All subjects were informed for the purpose and extent of the study and asked to sign a consent form before enrollment into this project.

2. Sample collection, laboratory assays, and DNA preparation

Venous blood (~10 ml) and urine samples were collected from affected and unaffected family members. Plasma glucose was determined by glucose oxidase method. Glycosylated hemoglobin (HbA1C), total cholesterol, triglyceride and HDL-cholesterol will be assayed by standard methods. LDL-cholesterol was directly measured or calculated by Friedewald formula (LDL-cholesterol = total cholesterol-HDL-cholesterol-triglyceride/5). Urine microalbumin and creatinine were measured by DCA 2000 Microalbumin/Creatinine Analyzer (Bayer Corporation Elkhart, IN USA) using immunoturbidity method. Ten to 15 ml of peripheral venous blood was collected from each subject into a sterile tube containing 20 µl of 20% EDTA for DNA preparation. The red blood cell (RBC) lysis buffer was used to destroy RBC. Leukocytes were separated and stored in a 15-ml screw-capped tube at –70^{o}C. Standard phenol/chloroform extraction method was used to prepare genomic DNA from the leukocytes.

3. Excluding mutations of six known MODY genes in the selected families

To completely rule out mutation of six known MODY genes as cause of diabetes in selected families, these genes were examined by re-sequencing. DNA samples extracted from peripheral blood leukocytes of probands from a family and newly collected families were subjected to amplifications by polymerase chain reaction (PCR) of each exon with flanking introns, 3'UTR and 5' flanking regions of the six known MODY genes. The purified PCR products were screened for mutations by direct sequencing. If sequence variations identified are segregated with diabetes, the families which harbor them were excluded from genome-wide linkage analysis.

4. SNP genotyping by Affymetrix GeneChip Human Mapping 10K

Affymetrix GeneChip Human Mapping 10K (The GeneChip® Human maping 10K Array Xba 142 2.0) is used for genome-wide linkage analysis, according to the methods described by the manufacturer. Briefly, 250 ng of genomic DNA is digested with XbaI restriction enzymes and ligated to adaptors that recognize the overhangs. The ligated DNA is amplified under recommended conditions, using primers complementary to the universal adapter. PCR product then be fragmented, labeled with biotinylated ddATP using terminal deoxynucleotidyl transferase and hybridized to the chip. After hybridization, washing, and signal development in the Affymetrix Fluidic System, the chips are then scanned by a laser scanner to acquire data for examination of a quality control (QC) call rate and genotype calls.

5. SNP genotyping by Affymetrix GeneChip Human Mapping version 6.0

Twenty-seven individuals from selected MODY-X family were genotype with the Affymetrix® Genome-Wide Human SNP Array 6.0 according to the method described by the manufacturer. Briefly, 250 ng of genomic DNA was digested with the restriction enzyme Sty I and Nps I and ligated to a common adaptor with T4 DNA ligase. Following ligation, the template undergoes PCR amplification using TITANIUM™ Taq DNA polymerase. Then, the Sty and Nsp PCR products were pooled to a single deep well pooling plate and undergo purification. The purified PCR products were then fragmented with Fragmentation Reagent (DNase I) and end-labeled using terminal

deoxynucleotidyl transferase and hybridized to the chips. After hybridization, the chips were washed and signals were developed in the Affymetrix Fluidic System. The chips were then scanned by laser scanner to acquire data for a quality control (QC) call rate and genotype calls.

6. Candidate region identification from linkage analysis

The genotype data generated from Affymetrix GeneChip Human Mapping 10K and Affymetrix® Genome-Wide Human SNP Array 6.0 together with pedigree and marker information were used to identify candidate regions that show significant linkage with diabetes in the families by using the easyLINKAGE Analysis Program (easyLINKAGE v5.08 beta). This study, two-point parametric analysis by SuperLink v1.6 and multipoint non-parametric linkage analysis by Simwalk 2.91 were used to identify linkage regions. The regions that show significant linkage with diabetes were selected to investigate candidate variants together with identified variants from WES.

When we used genotyping data from affymetric 10K to identify high LOD scores, significant LOD score (LOD>3) was not identified. However, highest LOD score on chromosome 4 which present LOD score 2.4220 at 78.04 cM was investigated (Fig 4, Fig 8). Because number of SNP markers on affymetric 10K array less than SNP array 6.0, we also used SNP array 6.0 for increasing the chance to identify significant LOD score. The significant LOD score was not identified also when we used SNP array 6.0. The highest LOD score was investigated on chromosome 5 which present LOD score 2.8981 at 101.08 cM (Fig 36, Fig 37). LOD score of each chromosome were shown in Table1 and Fig 4 to Fig 57.

Because significant LOD score was not identified, chromosomes presenting high LOD score $\geq$ 2.5, when used SNP array 6.0, were selected. There are 11 chromosomes showed LOD score $\geq$ 2.5 including chromosome 1 (LOD = 2.6879 at 160.85 cM), chromosome 4 (LOD = 2.6309 at 74.29 cM), chromosome 5 (LOD = 2.8981 at 101.08 cM), chromosome 6 (LOD = 2.7871 at 114.87 cM), chromosome 8 (LOD = 2.8121 at 118.92 cM), chromosome 9 (LOD = 2.5447 at 111.35 cM), chromosome 10 (LOD = 2.5781 at 105.91 cM), chromosome 11 (LOD = 2.5427 at 48.79 cM), chromosome 13 (LOD = 2.8102 at 8.76 cM), chromosome 16 (LOD = 2.6713 at 27.89 cM) and chromosome 22 (LOD = 2.6281 at 28.20 cM). Then, high LOD regions and regions where closely

with high LOD were selected to investigate pathogenic gene causing MODY in this family together with data received from WES (Table 10).

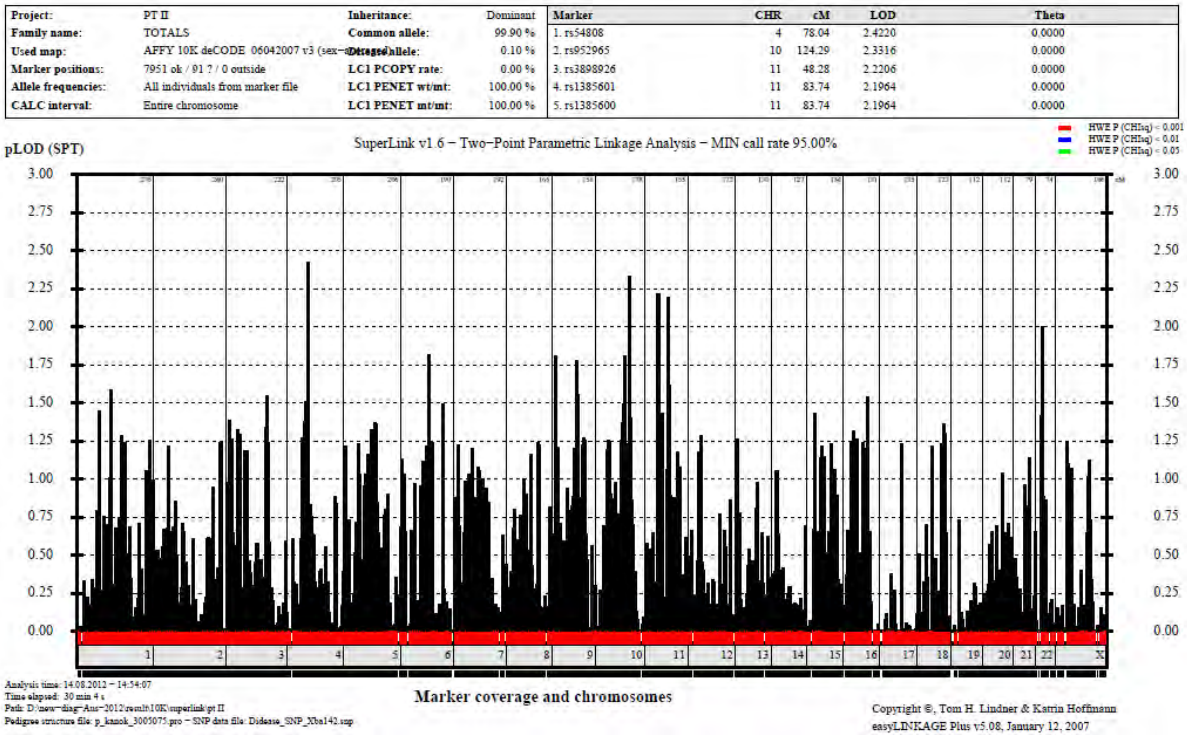
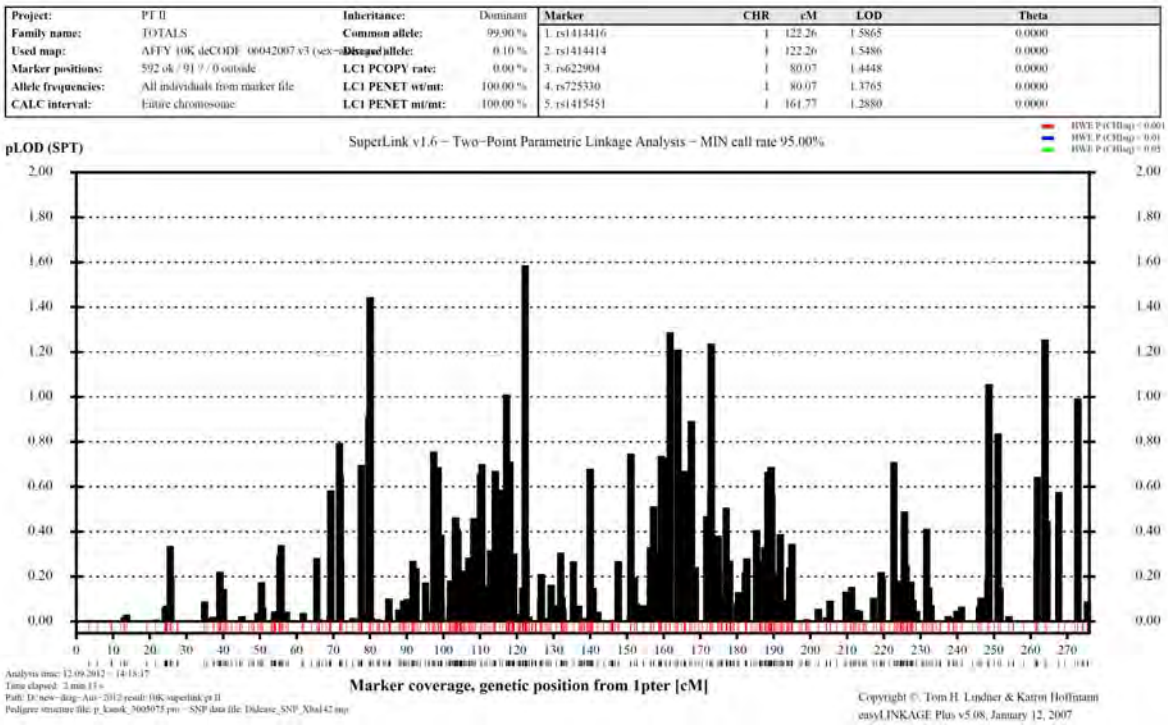


Figure 4 Two-point parametric LOD score was observed on all chromosomes by using data from genotyping affymetrix 10K array



Figur

e 5 Two-point parametric LOD score was observed on chromosome 1 by using data from genotyping affymetrix 10K array

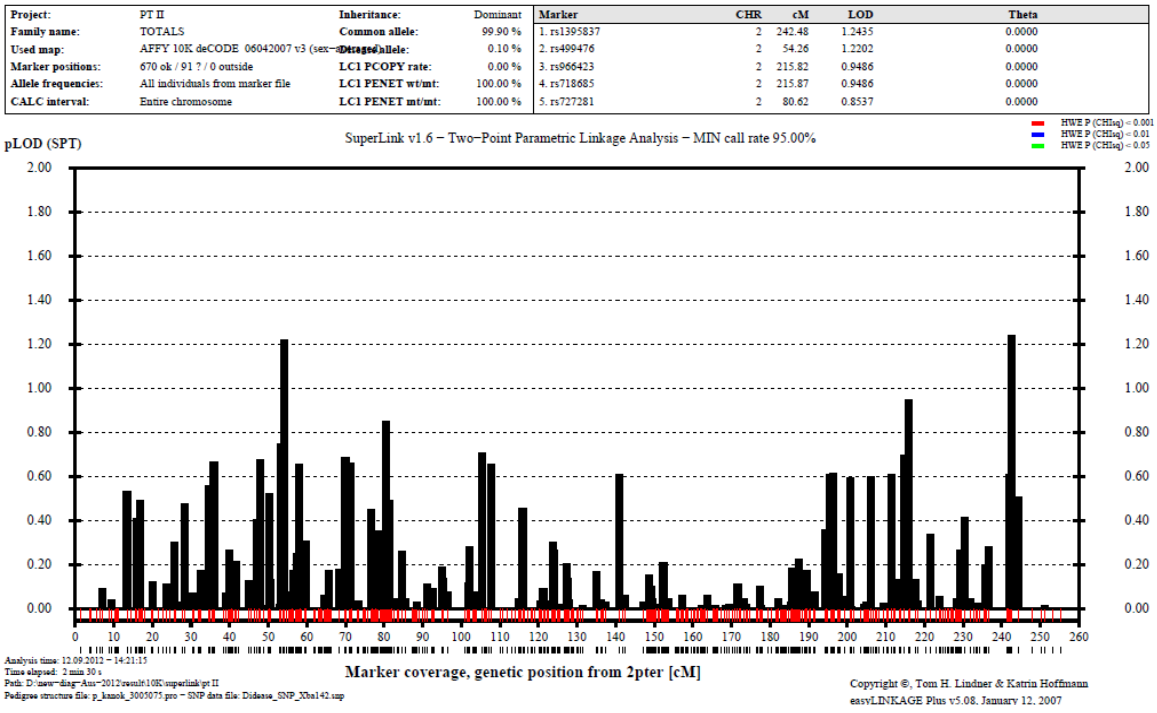


Figure 6 Two-point parametric LOD score was observed on chromosome 2 by using data from genotyping affymetrix 10K array

Project:	PT II	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 10K deCODE 06042007 v3 (sex-a	Disease allele:	0.10 %
Marker positions:	574 ok / 91 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs953239	3	149.20	1.5501	0.0000
2. rs1074650	3	12.64	1.3857	0.0500
3. rs724016	3	147.41	1.3386	0.0500
4. rs1344870	3	42.34	1.3245	0.0000
5. rs1350867	3	51.90	1.2959	0.0000

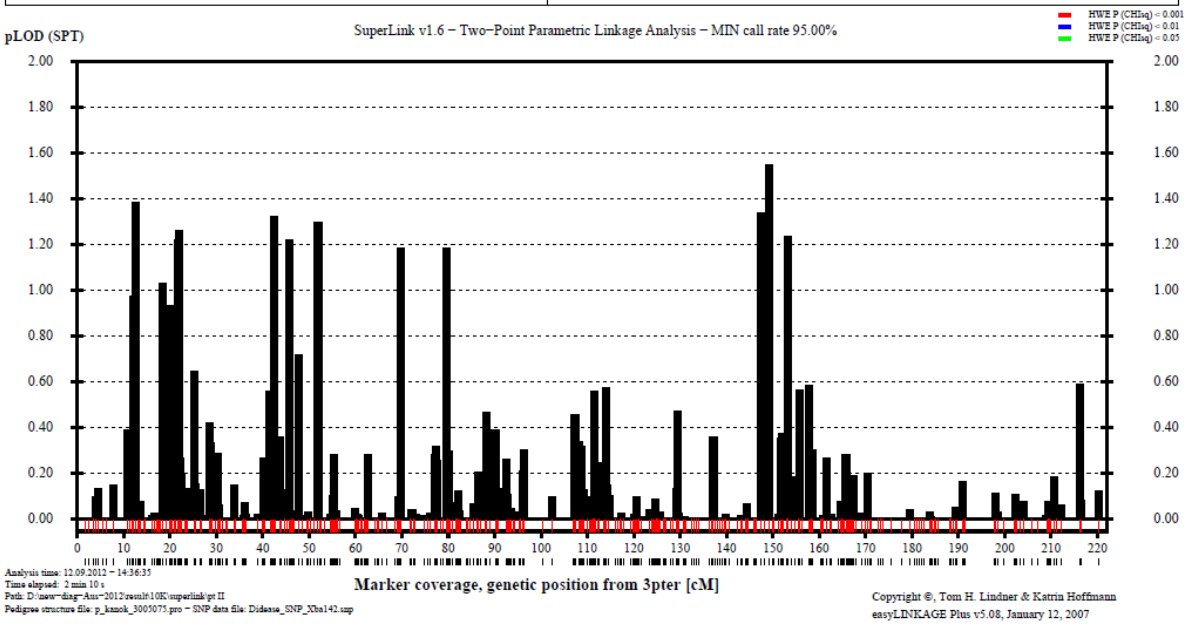


Figure 7 Two-point parametric LOD score was observed on chromosome 3 by using data from genotyping affymetrix 10K array

Project:	PT II	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 10K deCODE 06042007 v3 (sex-a	Disease allele:	0.10 %
Marker positions:	546 ok / 91 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs54808	4	78.04	2.4220	0.0000
2. rs1384696	4	66.95	1.5078	0.0000
3. rs2350135	4	65.37	1.3819	0.0000
4. rs1497478	4	69.12	1.2874	0.0000
5. rs958927	4	55.92	1.2738	0.0000

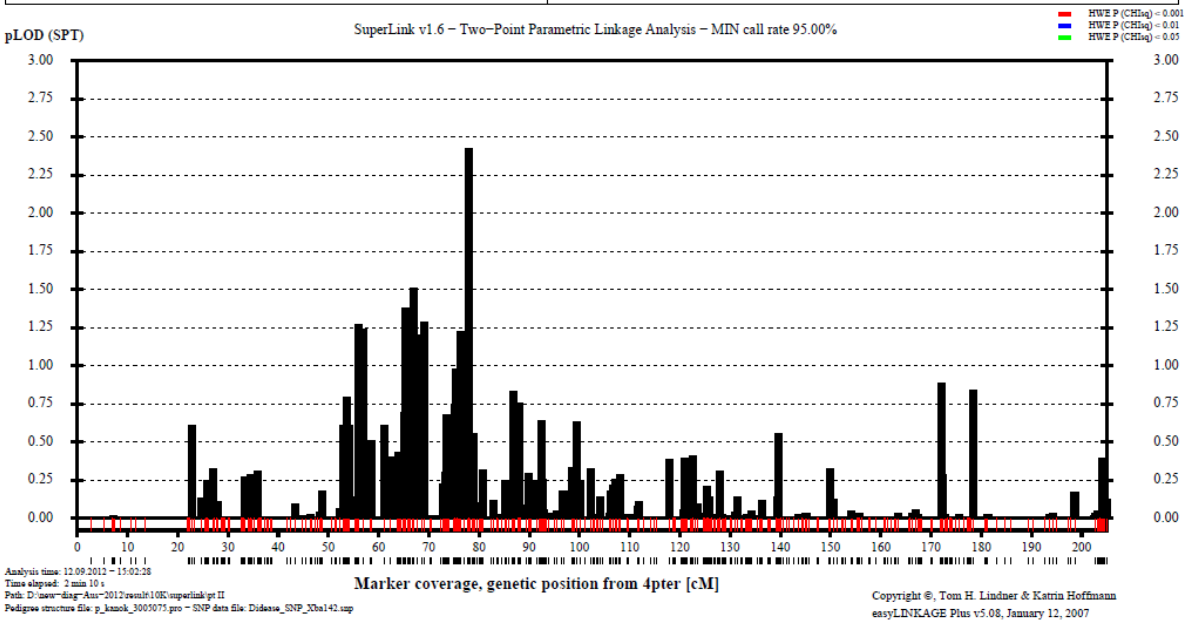


Figure 8 Two-point parametric LOD score was observed on chromosome 4 by using data from genotyping affymetrix 10K array

Project:	PT II	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 10K deCODE 06042007 v3 (sex-a	Disease allele:	0.10 %
Marker positions:	551 ok / 91 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1037796	5	111.60	1.3714	0.0000
2. rs36950	5	119.37	1.3603	0.0000
3. rs1445754	5	101.21	1.3234	0.0000
4. rs3915320	5	102.63	1.3234	0.0000
5. rs1395015	5	105.32	1.2467	0.0000

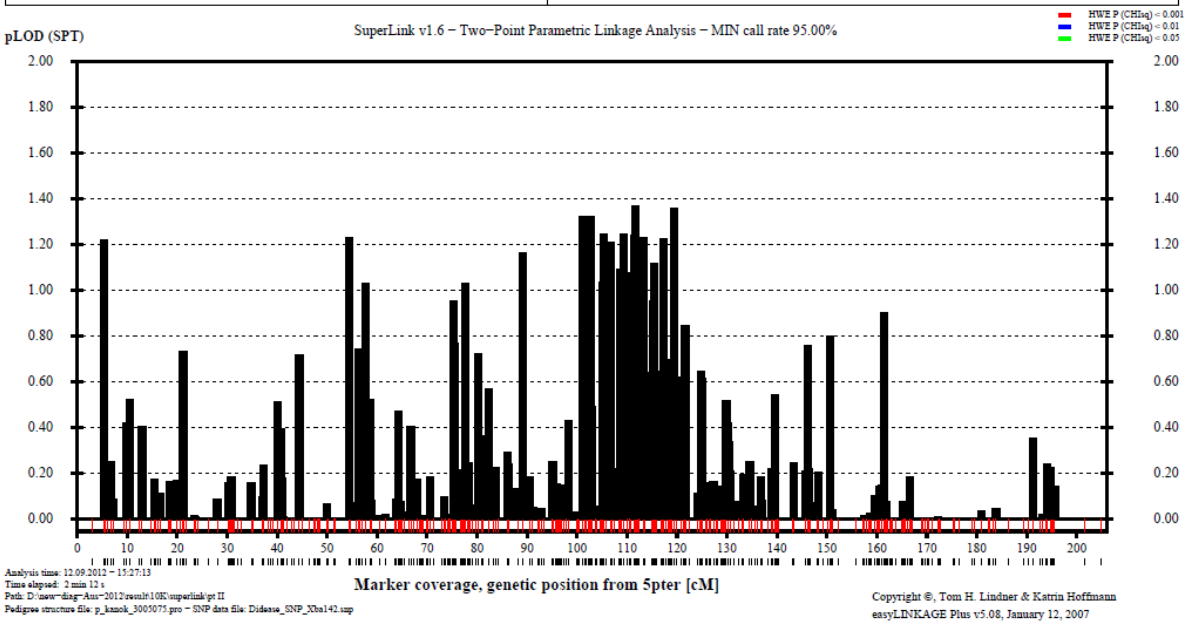


Figure 9 Two-point parametric LOD score was observed on chromosome 5 by using data from genotyping affymetrix 10K array

Project:	PT II	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 10K deCODE 06042007 v3 (sex-a	Disease allele:	0.10 %
Marker positions:	550 ok / 91 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs2387156	6	101.75	1.8204	0.0500
2. rs1408504	6	153.33	1.4910	0.0000
3. rs423246	6	100.78	1.2967	0.0000
4. rs1409836	6	115.99	1.2426	0.0000
5. rs1321754	6	93.27	1.2166	0.0000

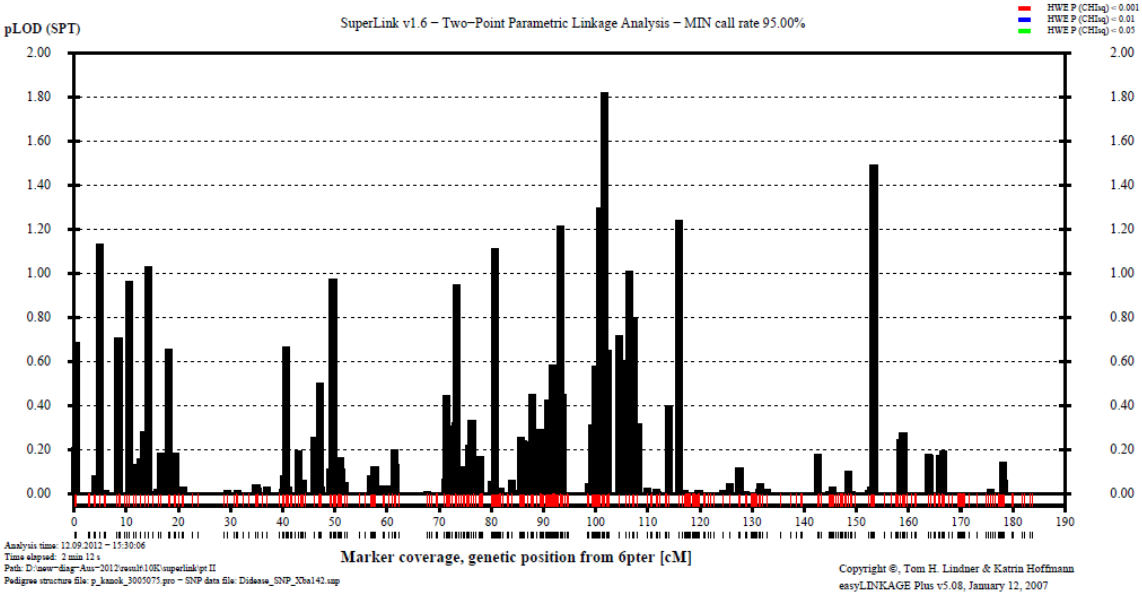


Figure 10 Two-point parametric LOD score was observed on chromosome 6 by using data from genotyping affymetrix 10K array

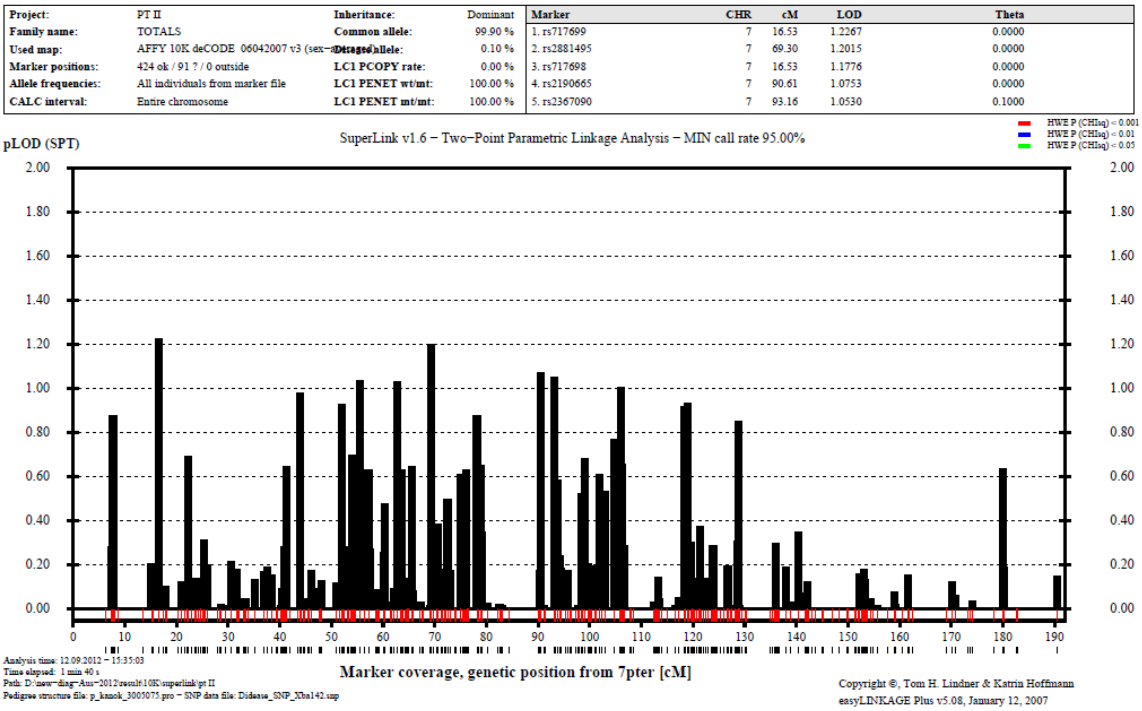


Figure 11 Two-point parametric LOD score was observed on chromosome 7 by using data from genotyping affymetrix 10K array

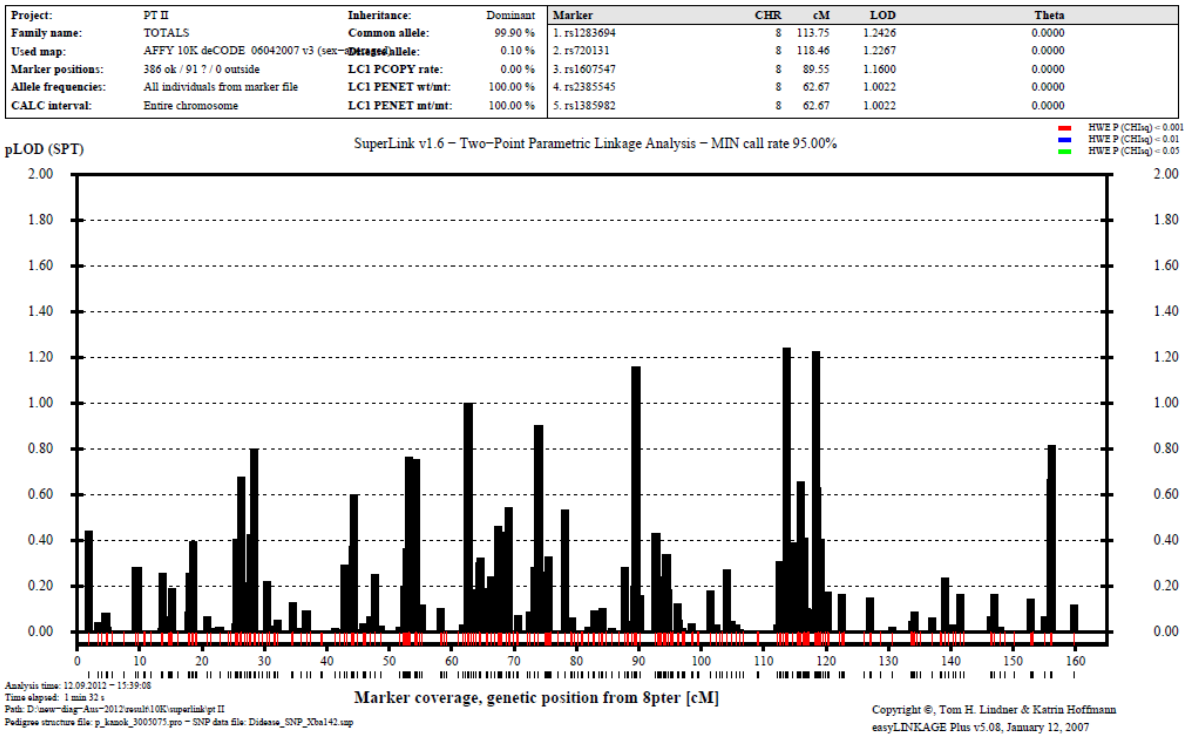


Figure 12 Two-point parametric LOD score was observed on chromosome 8 by using data from genotyping affymetrix 10K array

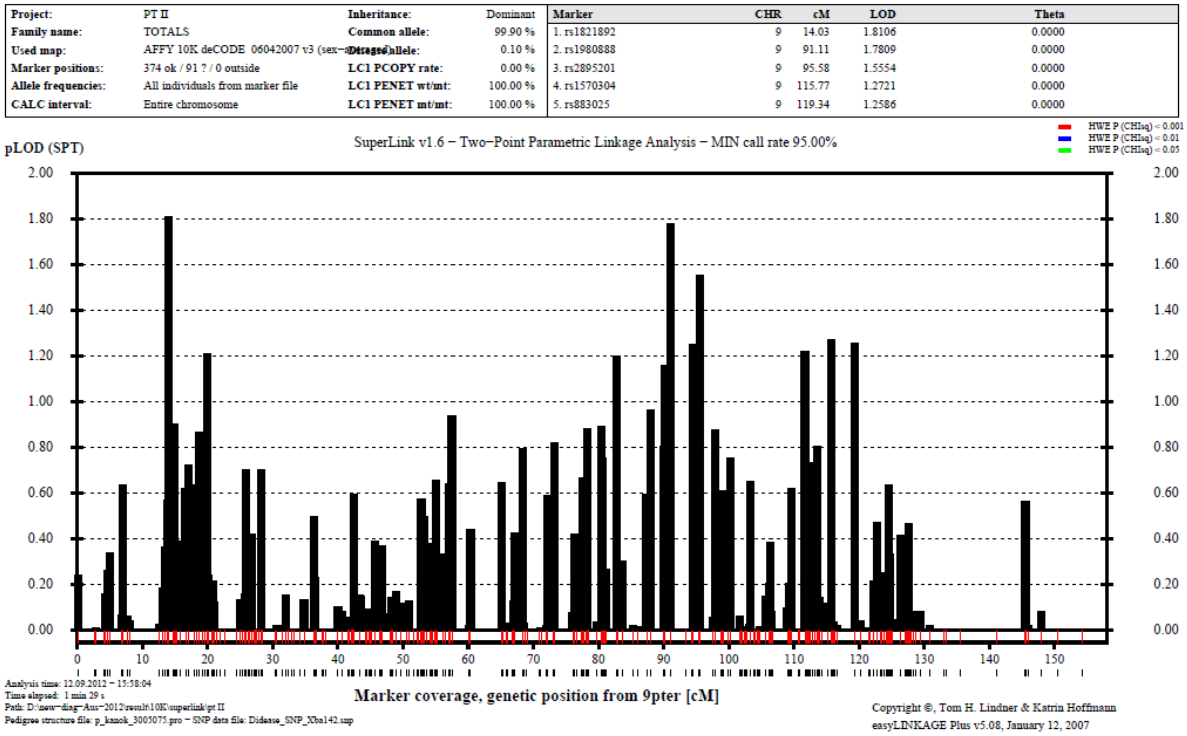


Figure 13 Two-point parametric LOD score was observed on chromosome 9 by using data from genotyping affymetrix 10K array

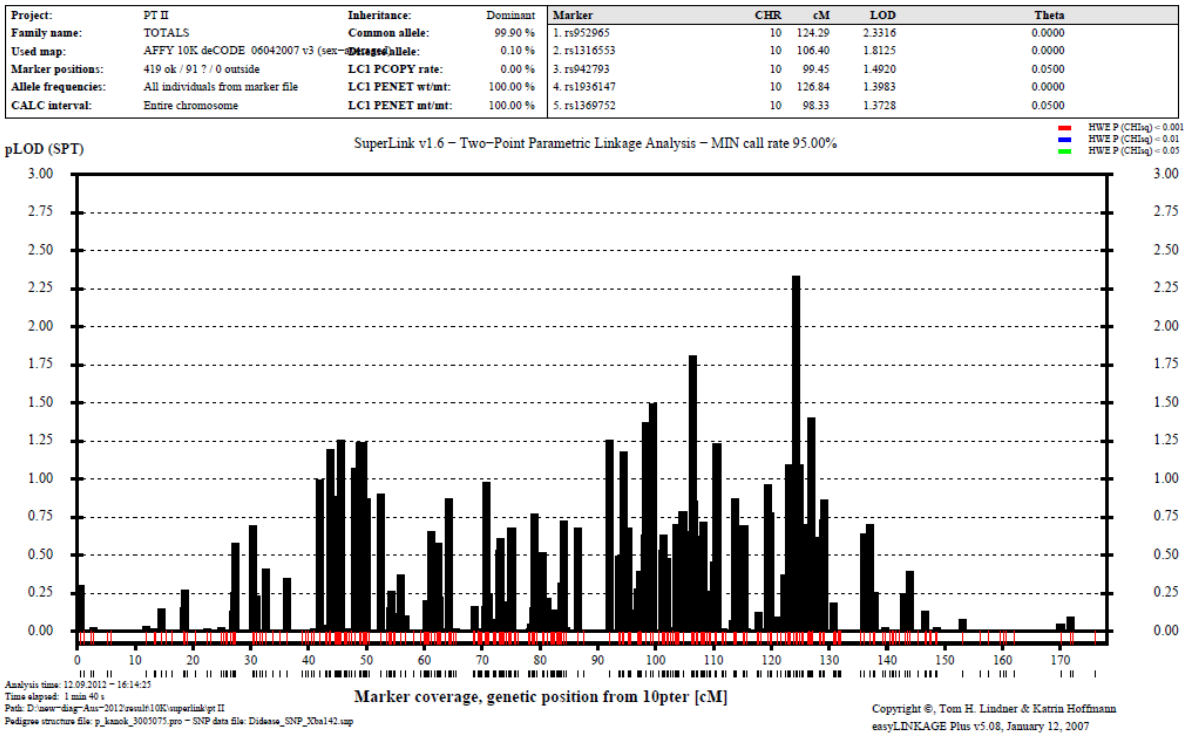


Figure 14 Two-point parametric LOD score was observed on chromosome 10 by using data from genotyping affymetrix 10K array

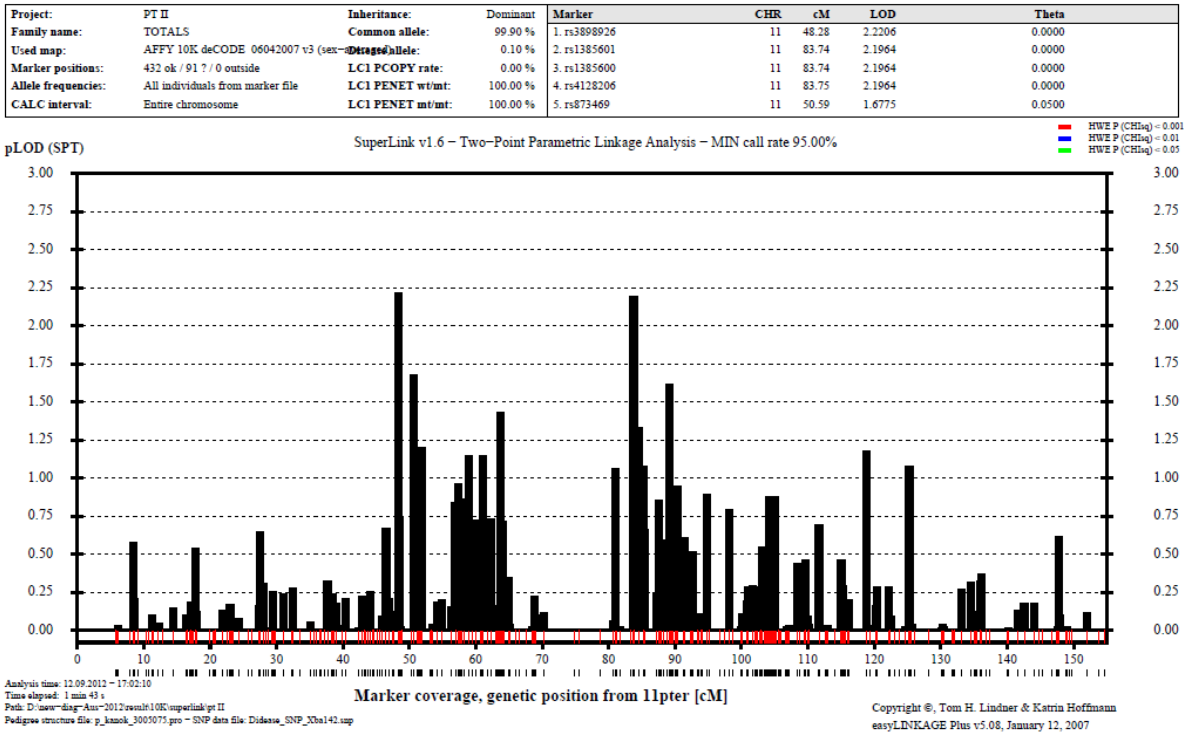


Figure 15 Two-point parametric LOD score was observed on chromosome 11 by using data from genotyping affymetrix 10K array

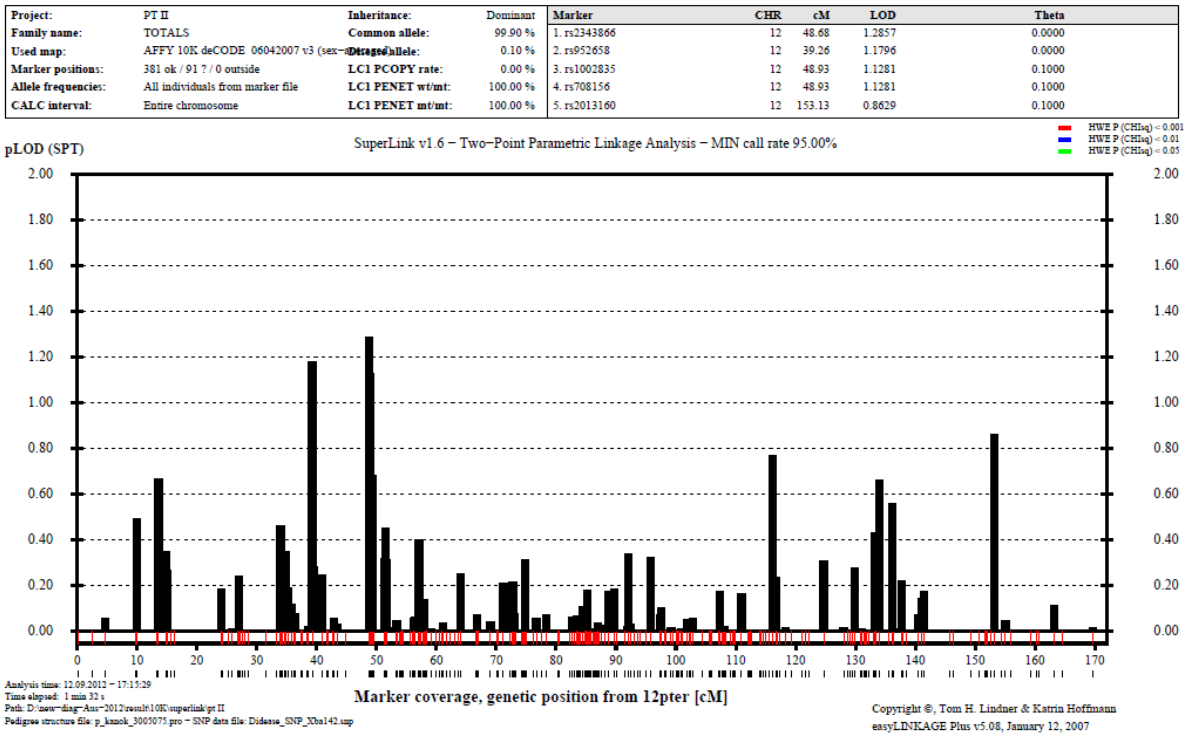


Figure 16 Two-point parametric LOD score was observed on chromosome 12 by using data from genotyping affymetrix 10K array

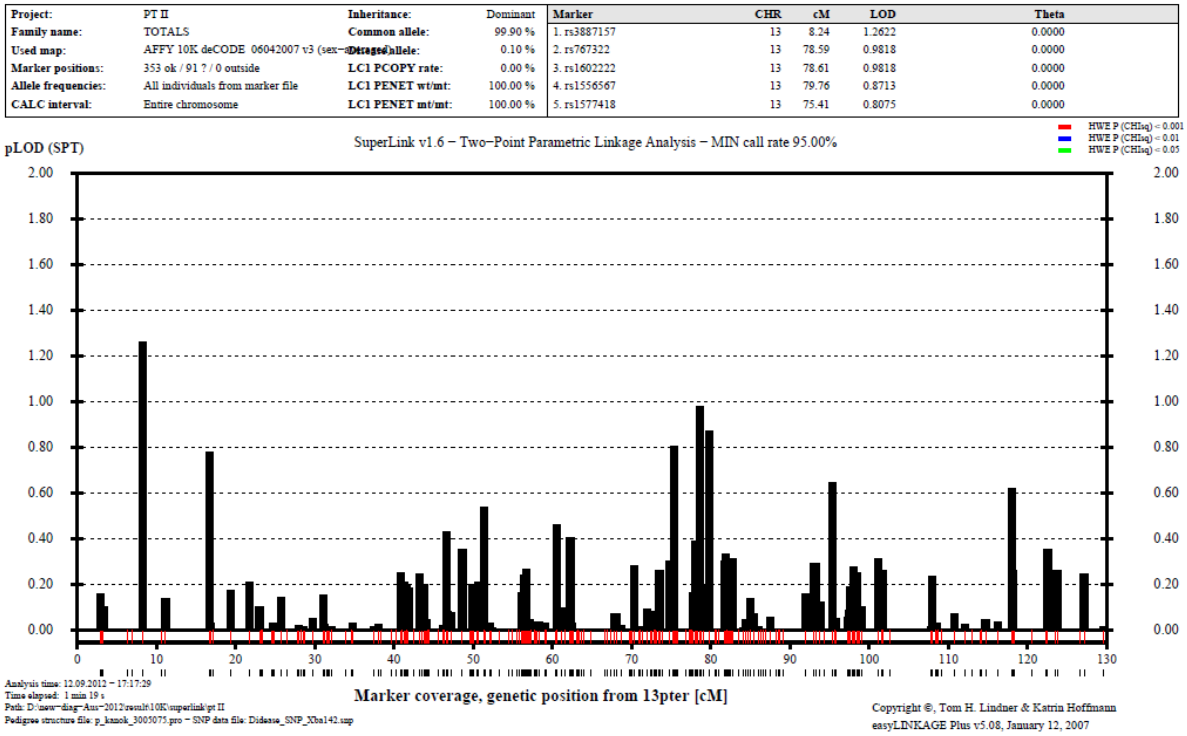


Figure 17 Two-point parametric LOD score was observed on chromosome 13 by using data from genotyping affymetrix 10K array

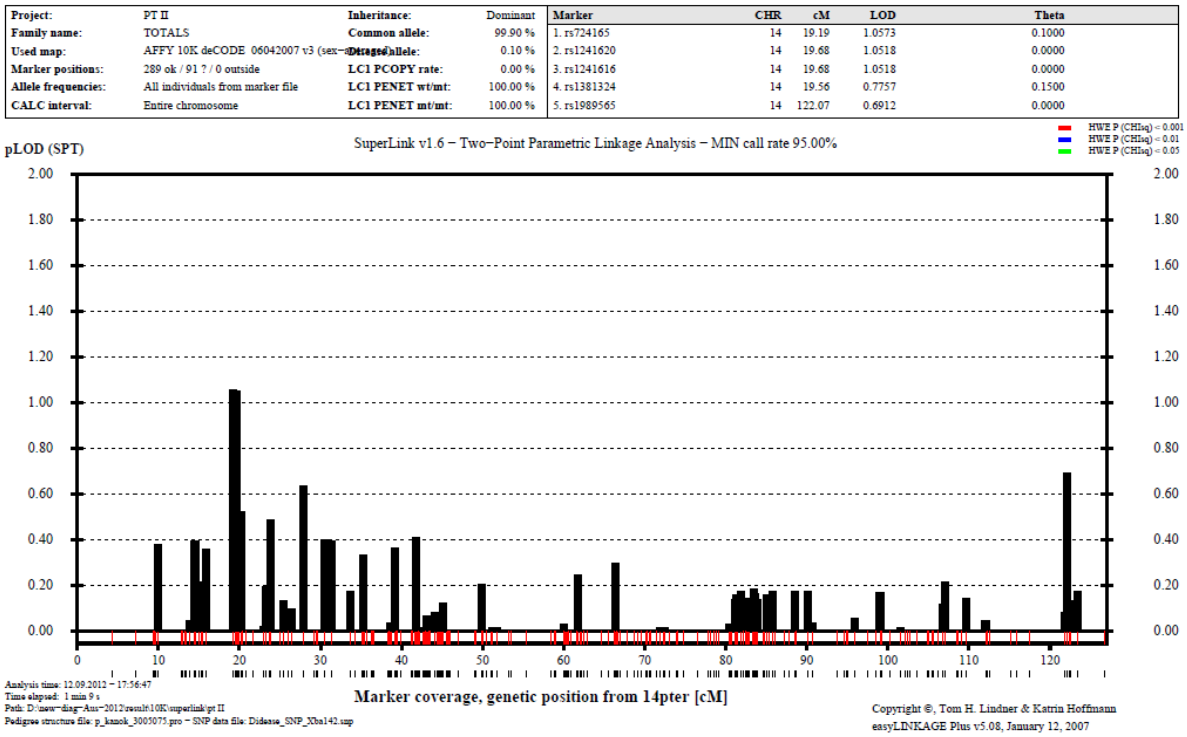


Figure 18 Two-point parametric LOD score was observed on chromosome 14 by using data from genotyping affymetrix 10K array

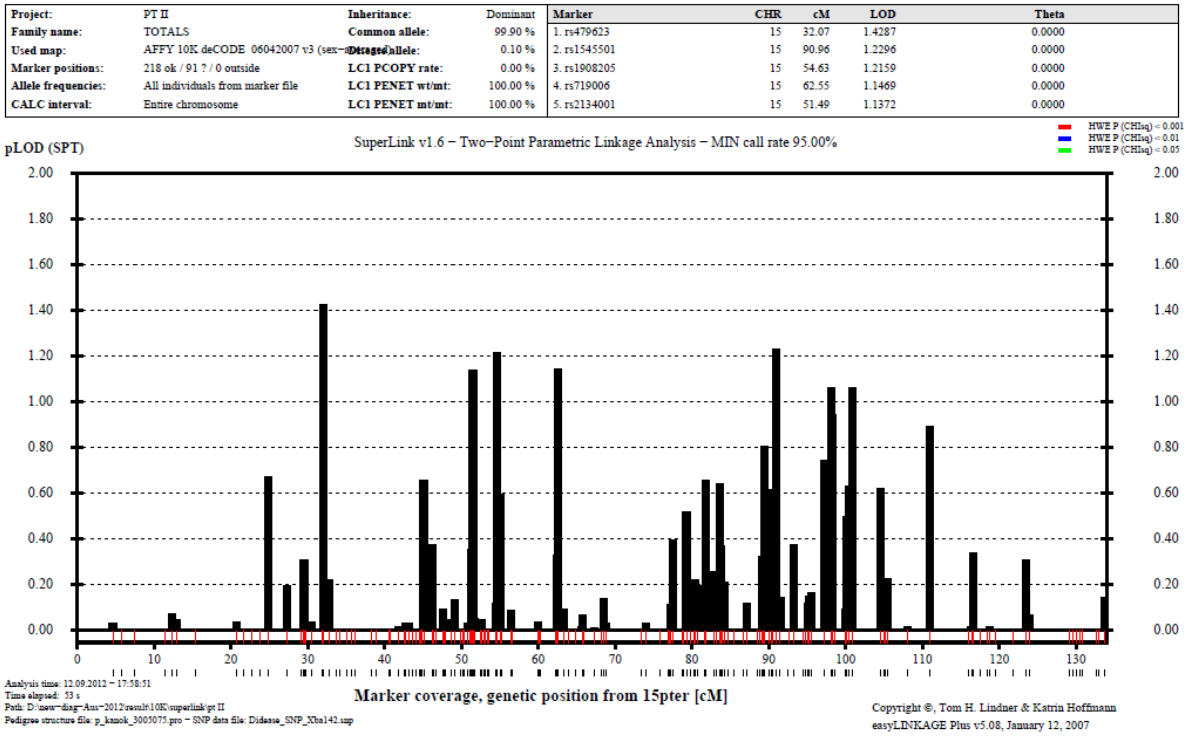


Figure 19 Two-point parametric LOD score was observed on chromosome 15 by using data from genotyping affymetrix 10K array

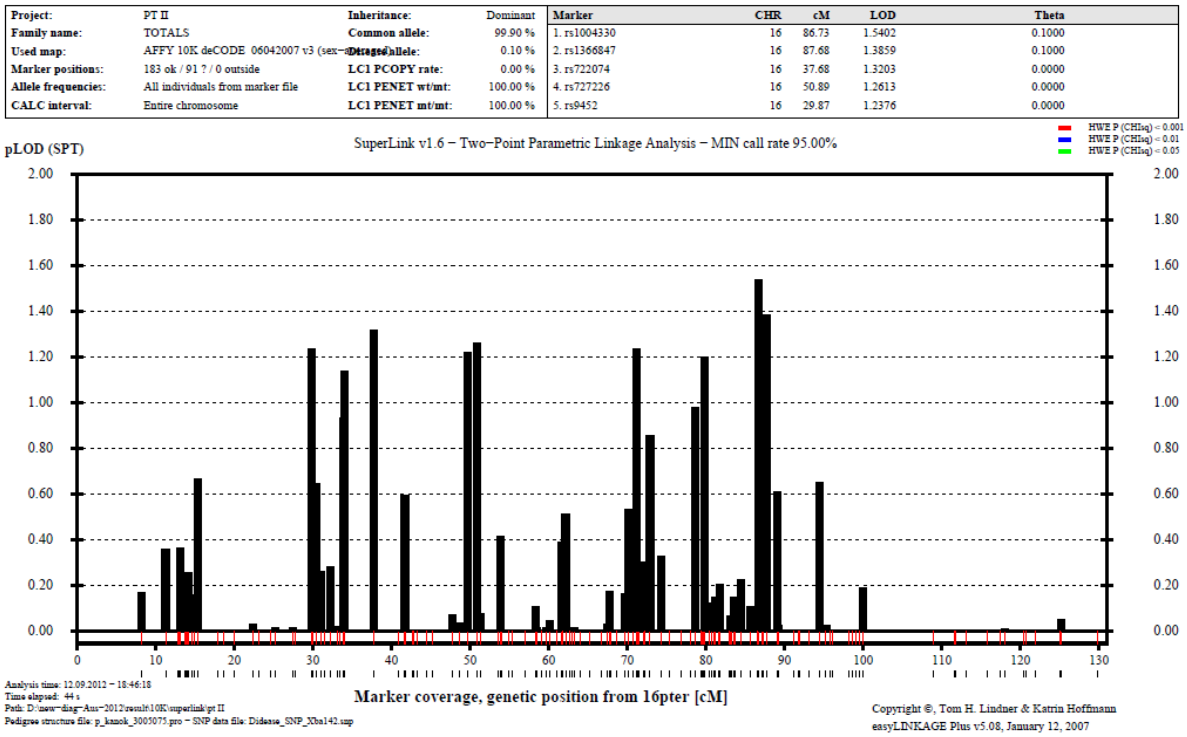


Figure 20 Two-point parametric LOD score was observed on chromosome 16 by using data from genotyping affymetrix 10K array

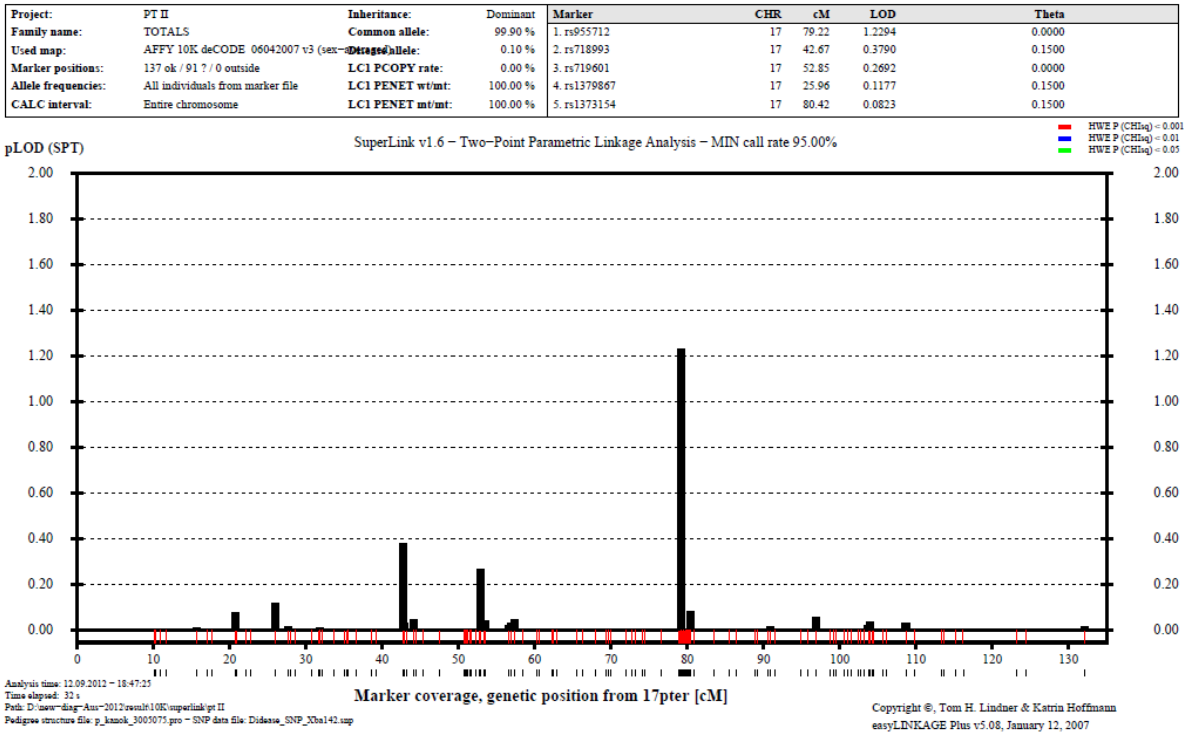


Figure 21 Two-point parametric LOD score was observed on chromosome 17 by using data from genotyping affymetrix 10K array

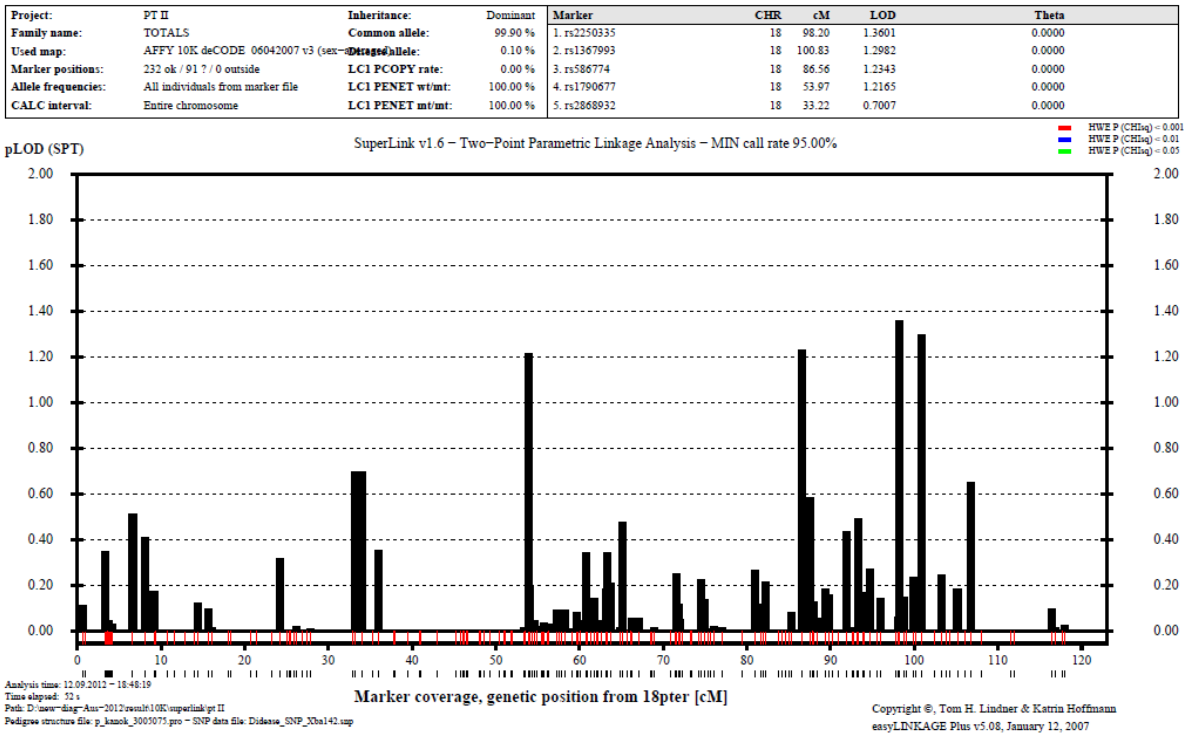


Figure 22 Two-point parametric LOD score was observed on chromosome 18 by using data from genotyping affymetrix 10K array

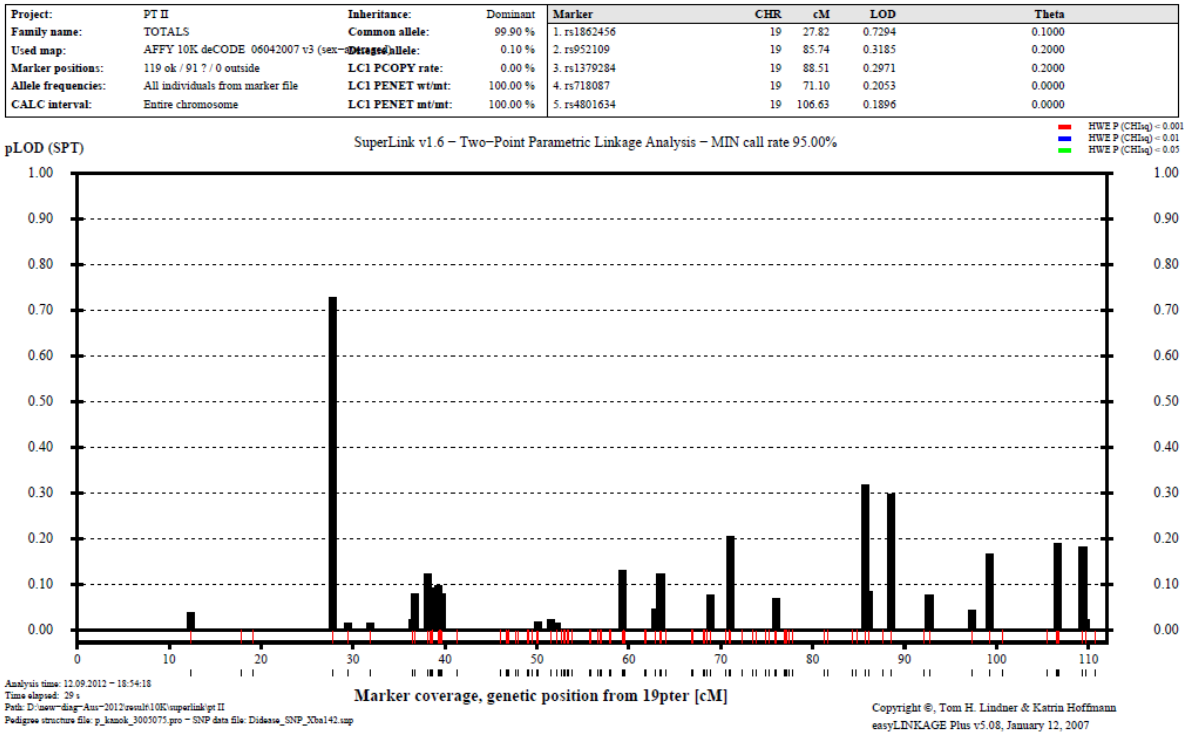


Figure 23 Two-point parametric LOD score was observed on chromosome 19 by using data from genotyping affymetrix 10K array

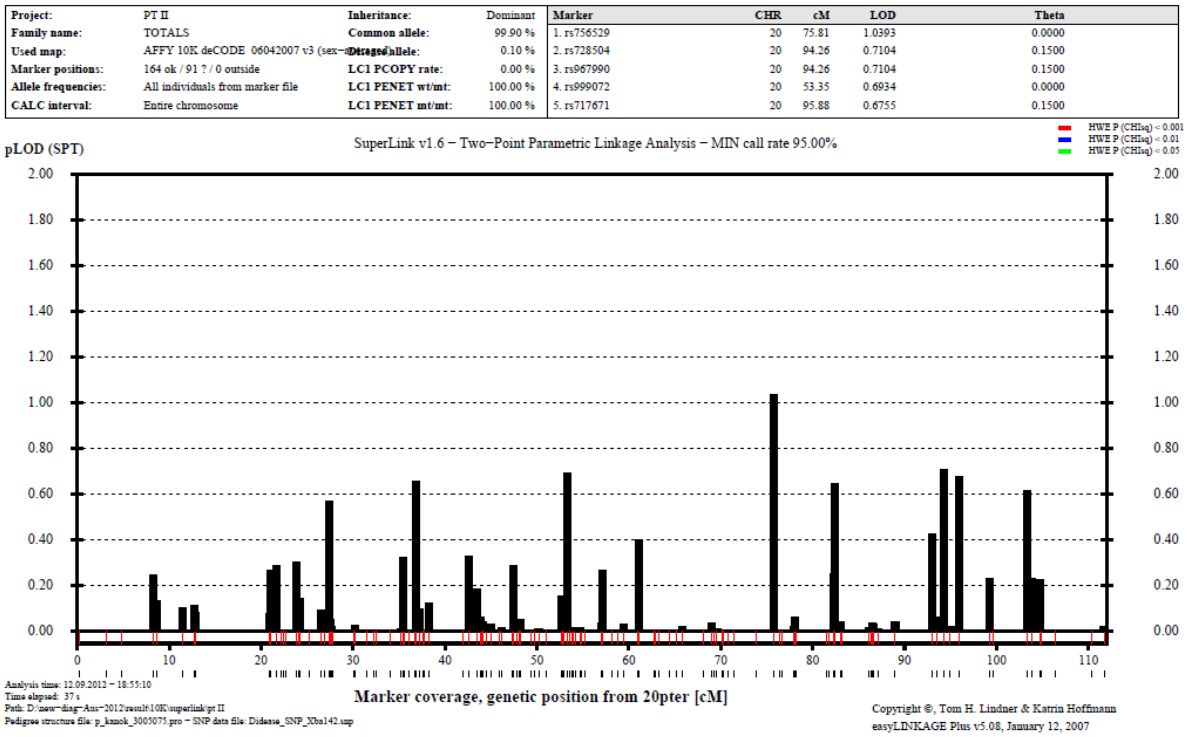


Figure 24 Two-point parametric LOD score was observed on chromosome 20 by using data from genotyping affymetrix 10K array

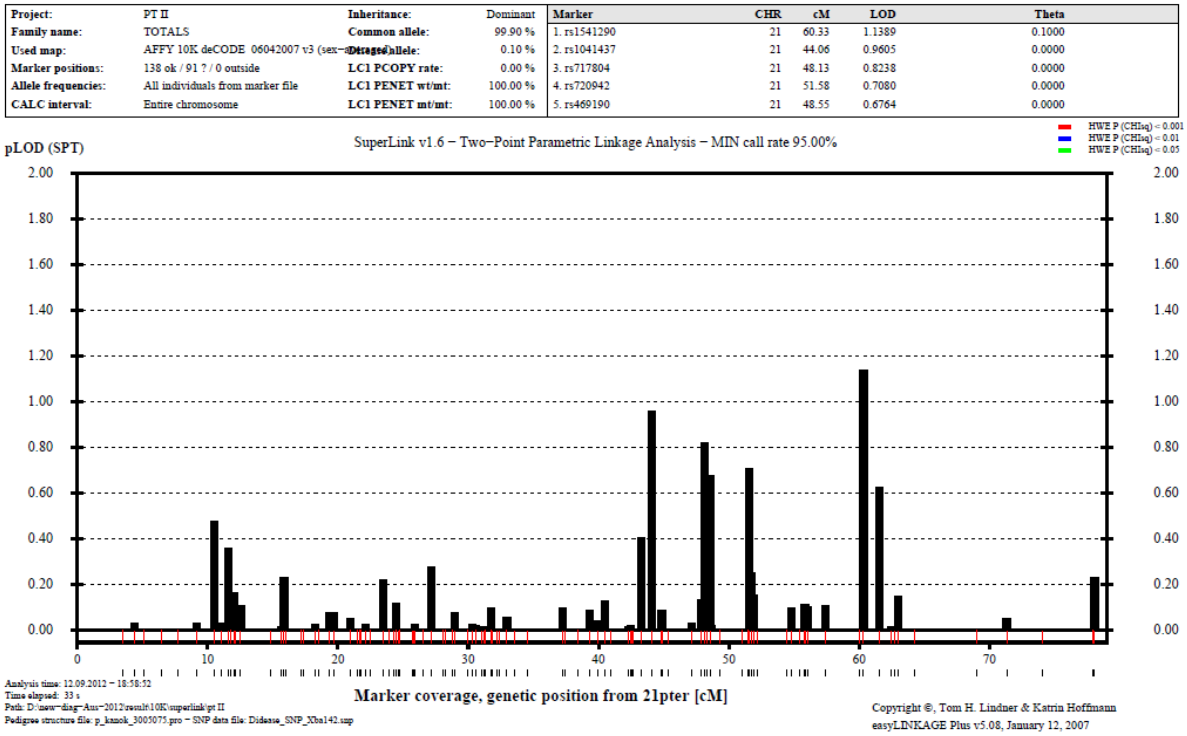


Figure 25 Two-point parametric LOD score was observed on chromosome 21 by using data from genotyping affymetrix 10K array

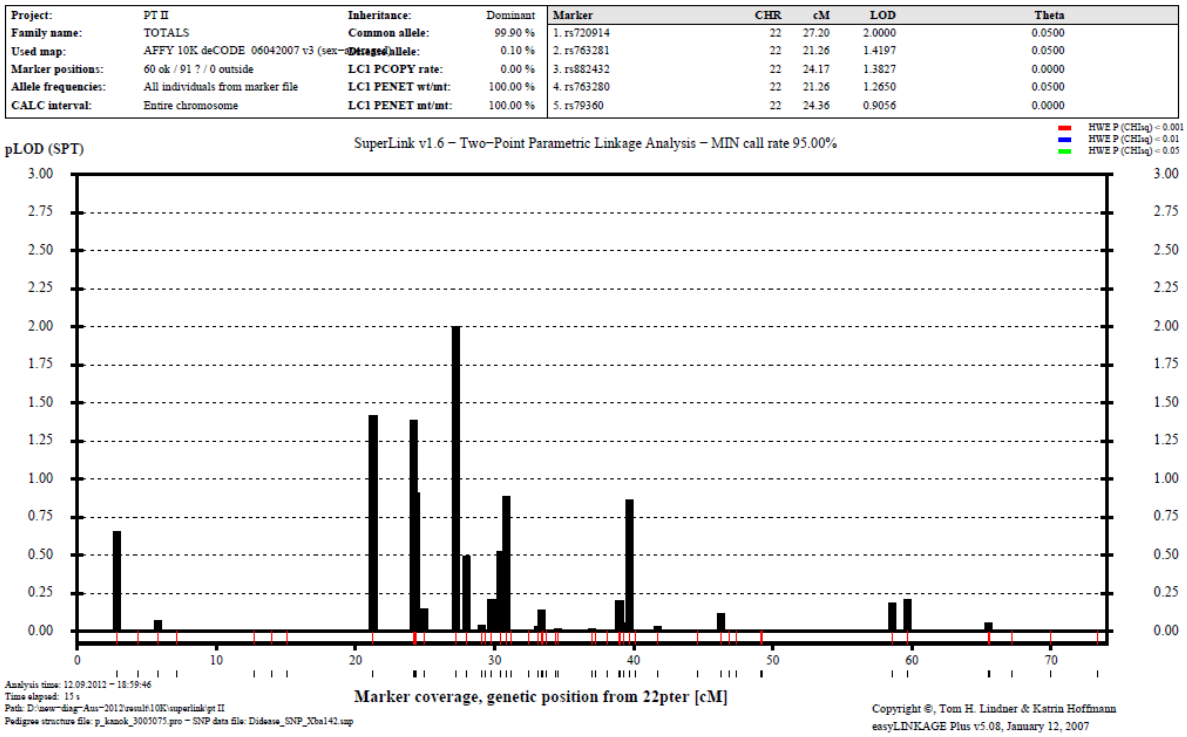


Figure 26 Two-point parametric LOD score was observed on chromosome 22 by using data from genotyping affymetrix 10K

array

				Marker	CHR	cM	LOD	Theta
Project:	PT II	**Inheritance:**	Dominant					
Family name:	TOTALS	**Common allele:**	99.90 %	1. rs724087	X	43.67	1.2422	0.1000
Used map:	AFFY 10K deCODE 06042007 v3 (sex-averaged)	**Disease allele:**	0.10 %	2. rs1151719	X	123.42	1.1224	0.0000
Marker positions:	159 ok / 91 ? / 0 outside	**LC1 PCOPY rate:**	0.00 %	3. rs1074193	X	45.95	1.1007	0.0000
Allele frequencies:	All individuals from marker file	**LC1 PENET wt/wt:**	100.00 %	4. rs2316179	X	59.37	1.0724	0.0000
CALC interval:	Entire chromosome	**LC1 PENET mt/mt:**	100.00 %	5. rs4132117	X	116.90	1.0183	0.0000

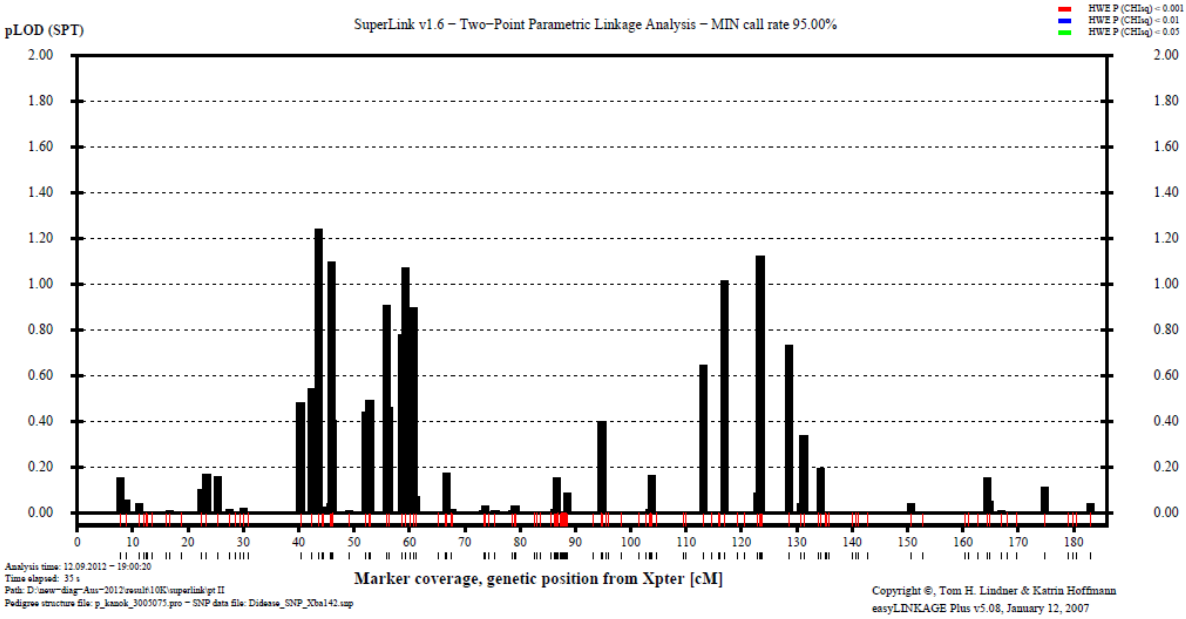


Figure 27 Two-point parametric LOD score was observed on chromosome X by using data from genotyping affymetrix 10K array

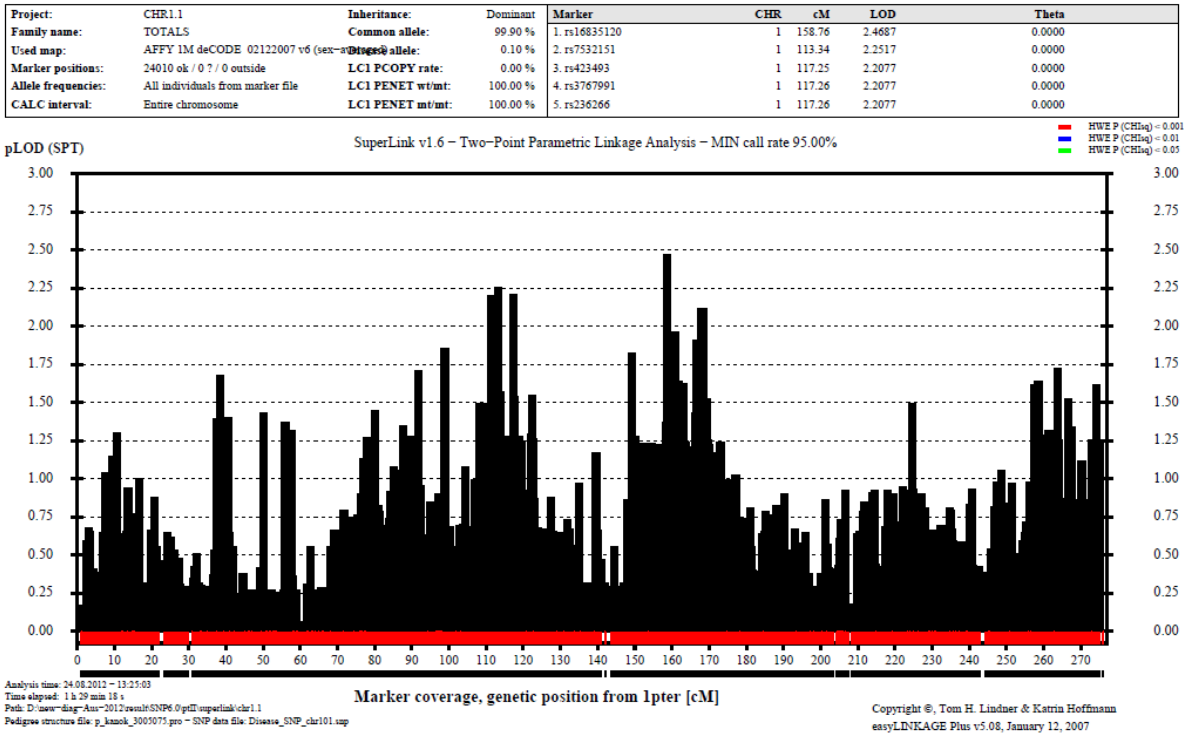


Figure 28 Two-point parametric LOD score was observed on chromosome 1 (file1) by using data from genotyping SNP array 6.0

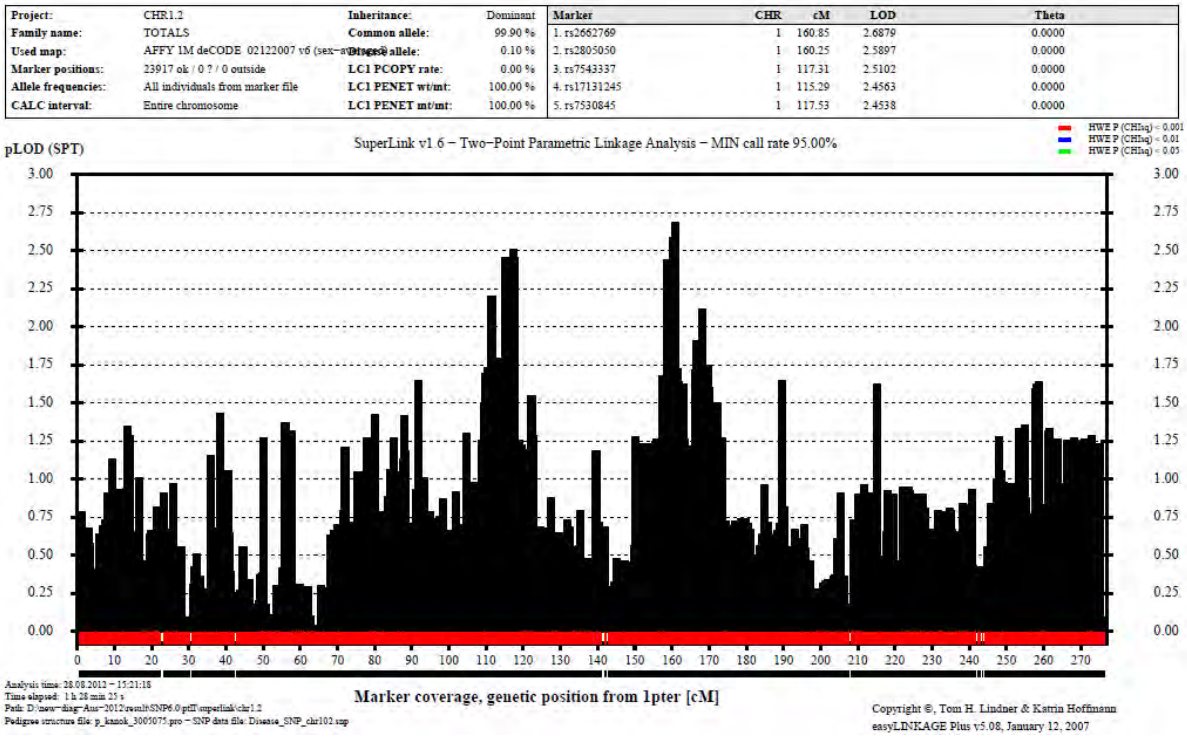


Figure 29 Two-point parametric LOD score was observed on chromosome 1 (file2) by using data from genotyping SNP array 6.0

Project:	CHR2.1	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	24490 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs13398373	2	44.98	2.0366	0.0000
2. rs1017476	2	66.43	1.9807	0.0000
3. rs16824919	2	233.02	1.8987	0.0000
4. rs16824867	2	232.98	1.8689	0.0000
5. rs7608256	2	234.01	1.8267	0.0000

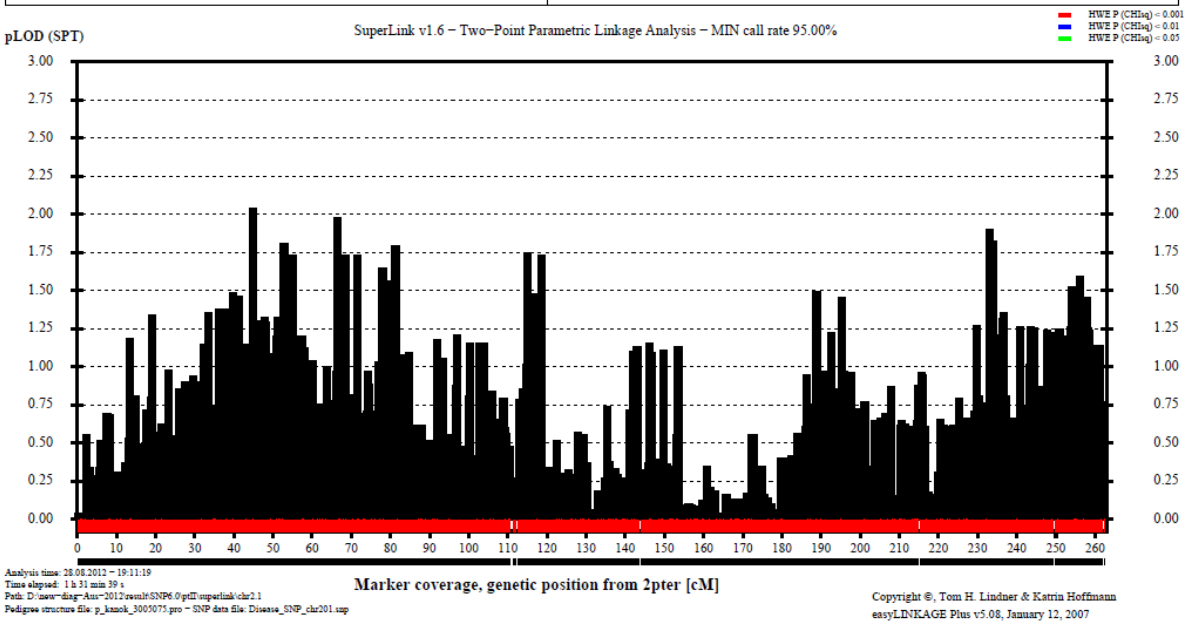


Figure 30 Two-point parametric LOD score was observed on chromosome 2 (file1) by using data from genotyping SNP array 6.0

Project:	CHR2.2	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	25362 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs4131238	2	67.25	2.2206	0.0000
2. rs7558605	2	67.23	2.0007	0.0000
3. rs12478290	2	231.95	1.9564	0.0000
4. rs2270388	2	233.97	1.9162	0.0000
5. rs7559213	2	233.98	1.9162	0.0000

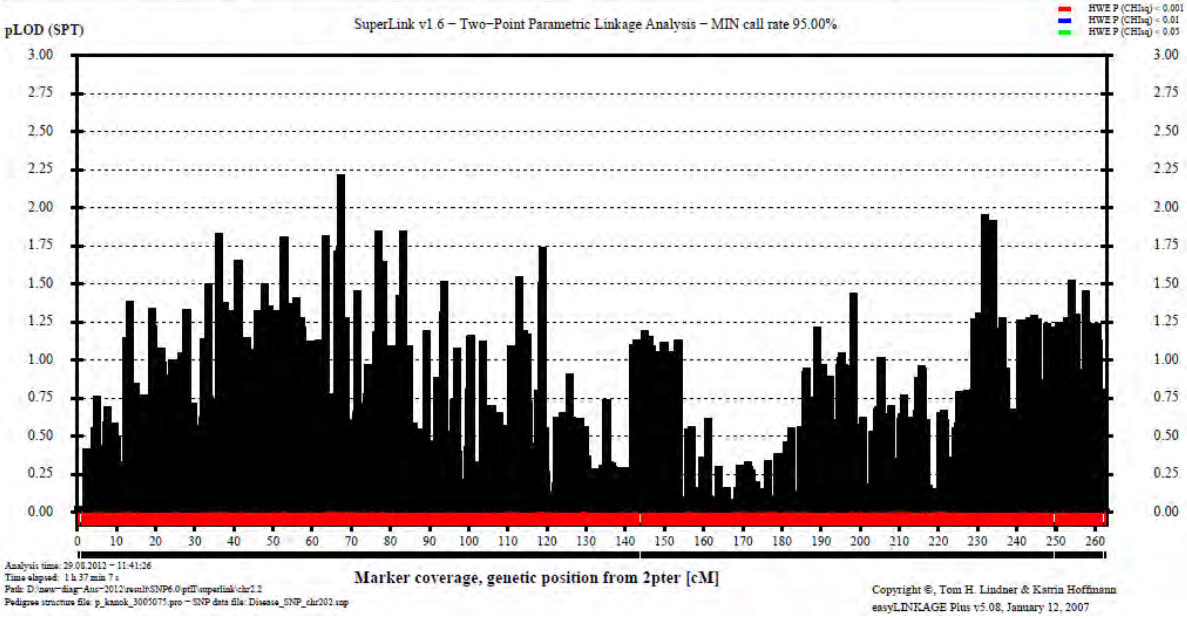


Figure 31 Two-point parametric LOD score was observed on chromosome 2 (file2) by using data from genotyping SNP array 6.0

Project:	CHR3.1	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	21105 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1517380	3	18.23	2.0180	0.0000
2. rs13076437	3	18.23	2.0180	0.0000
3. rs155403	3	2.74	1.8399	0.0000
4. rs2729309	3	4.91	1.8259	0.0000
5. rs17016781	3	48.34	1.6911	0.0000

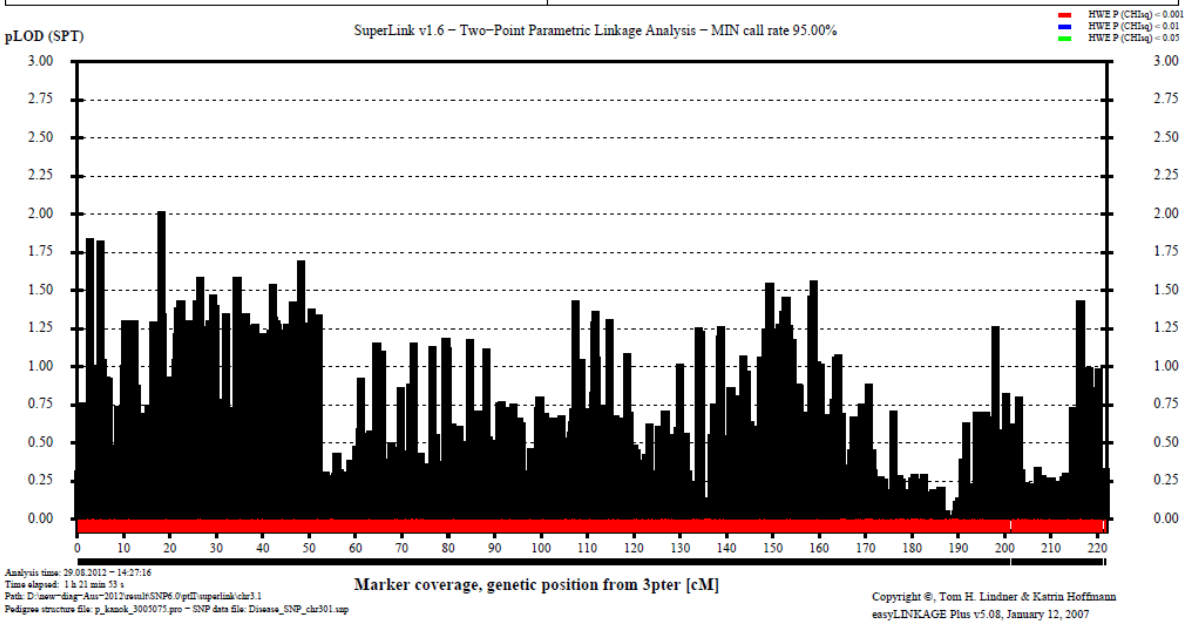
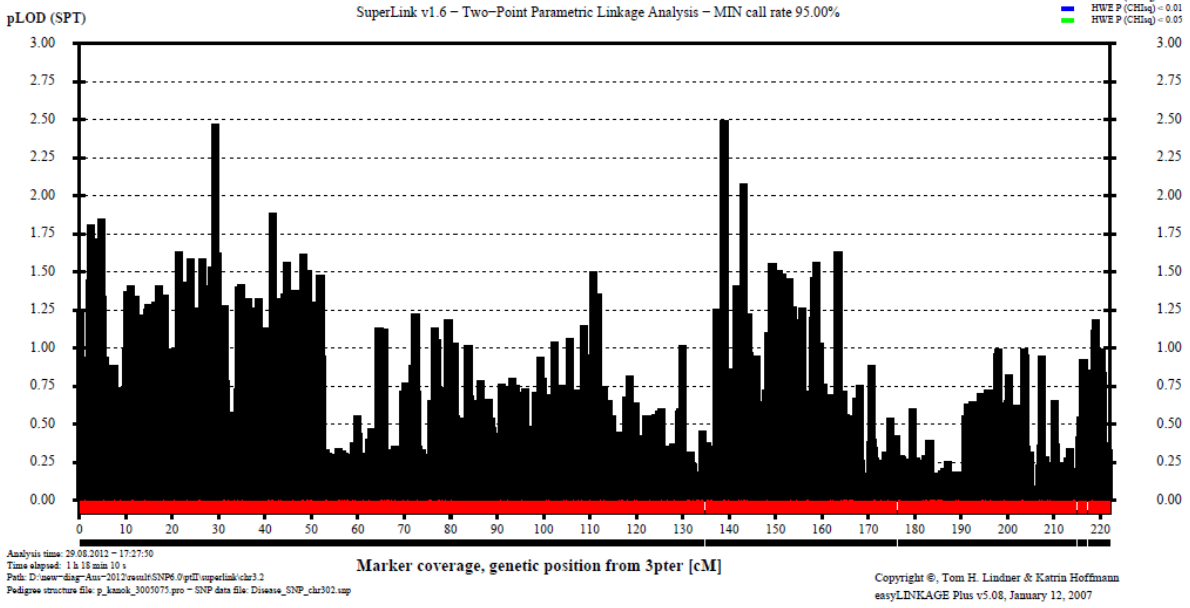


Figure 32 Two-point parametric LOD score was observed on chromosome 3 (file1) by using data from genotyping SNP array

Project:	CHR3.2	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	20337 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/mt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1356802	3	138.95	2.4918	0.0000
2. rs238615	3	29.18	2.4737	0.0000
3. rs11917704	3	143.19	2.0787	0.0000
4. rs6796194	3	41.74	1.8886	0.1000
5. rs1488757	3	4.79	1.8472	0.0000



6.0

Figure 33 Two-point parametric LOD score was observed on chromosome 3 (file2) by using data from genotyping SNP array 6.0

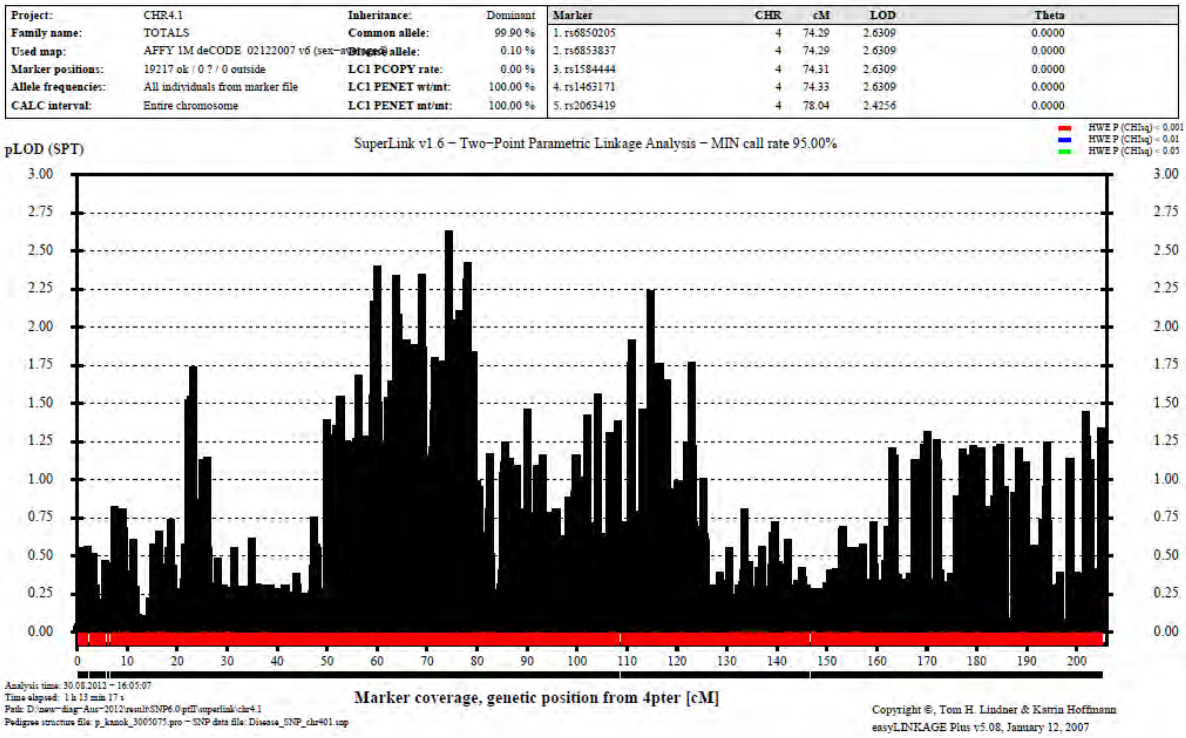


Figure 34 Two-point parametric LOD score was observed on chromosome 4 (file1) by using data from genotyping SNP array 6.0

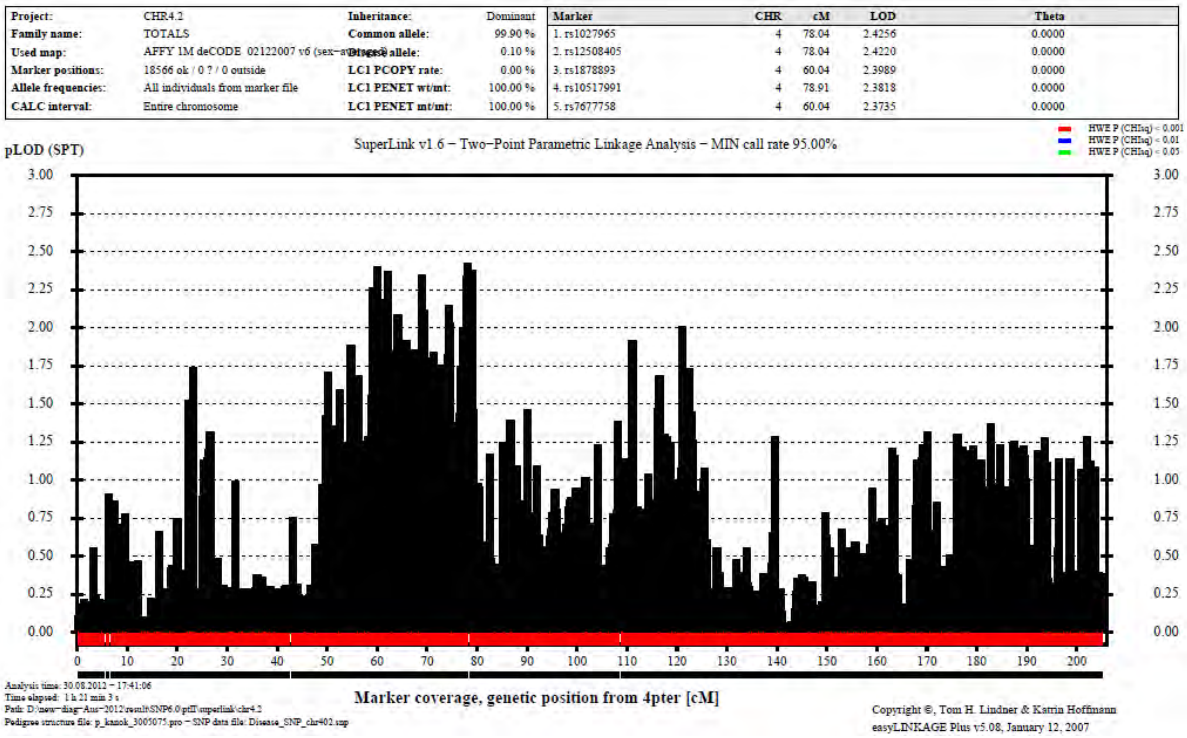


Figure 35 Two-point parametric LOD score was observed on chromosome 4 (file2) by using data from genotyping SNP array 6.0

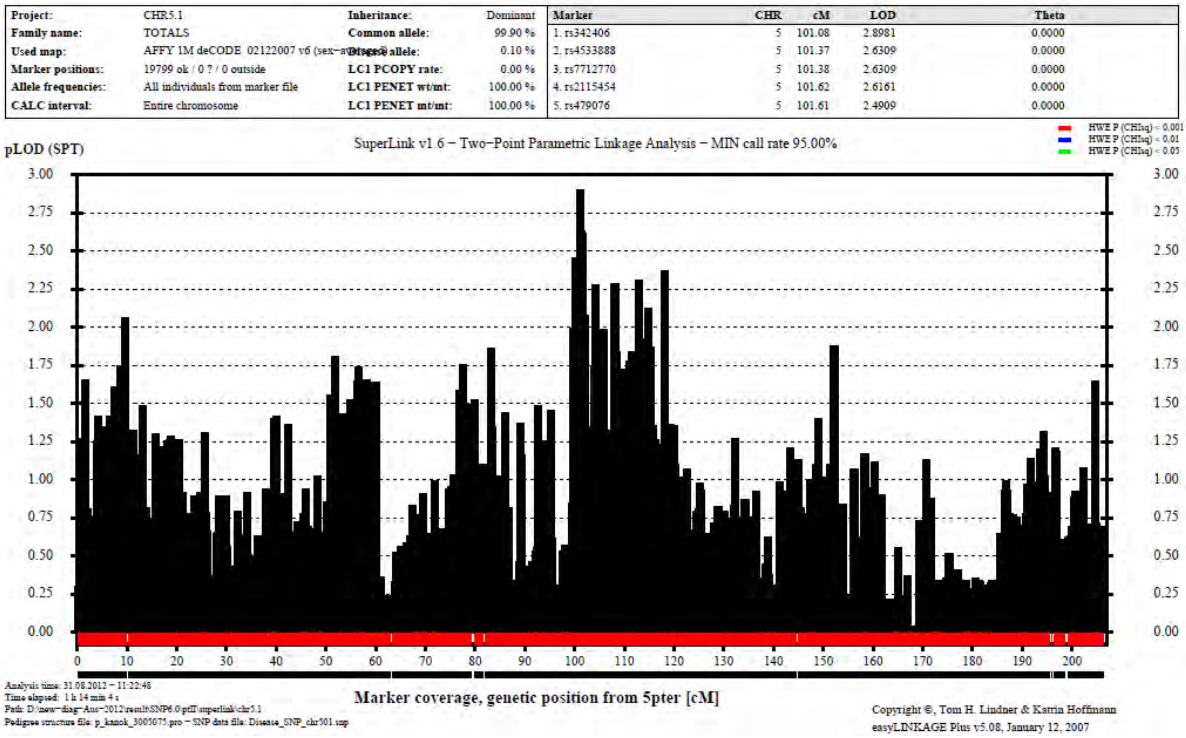


Figure 36 Two-point parametric LOD score was observed on chromosome 5 (file1) by using data from genotyping SNP array 6.0

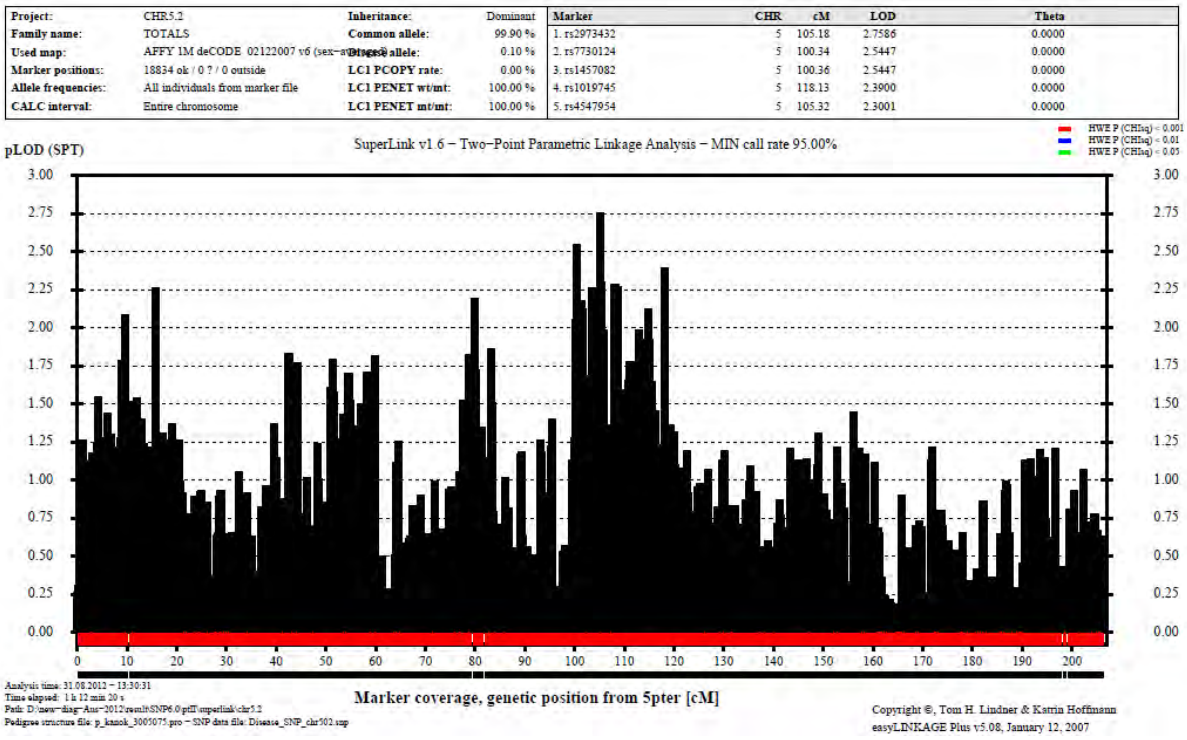


Figure 37 Two-point parametric LOD score was observed on chromosome 5 (file2) by using data from genotyping SNP array 6.0

Project:	CHR6.1	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	20431 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs2503627	6	99.78	2.3185	0.0000
2. rs2452987	6	99.78	2.3185	0.0000
3. rs2387699	6	102.15	2.0037	0.1000
4. rs9400365	6	102.19	2.0037	0.1000
5. rs7744614	6	102.21	2.0037	0.1000

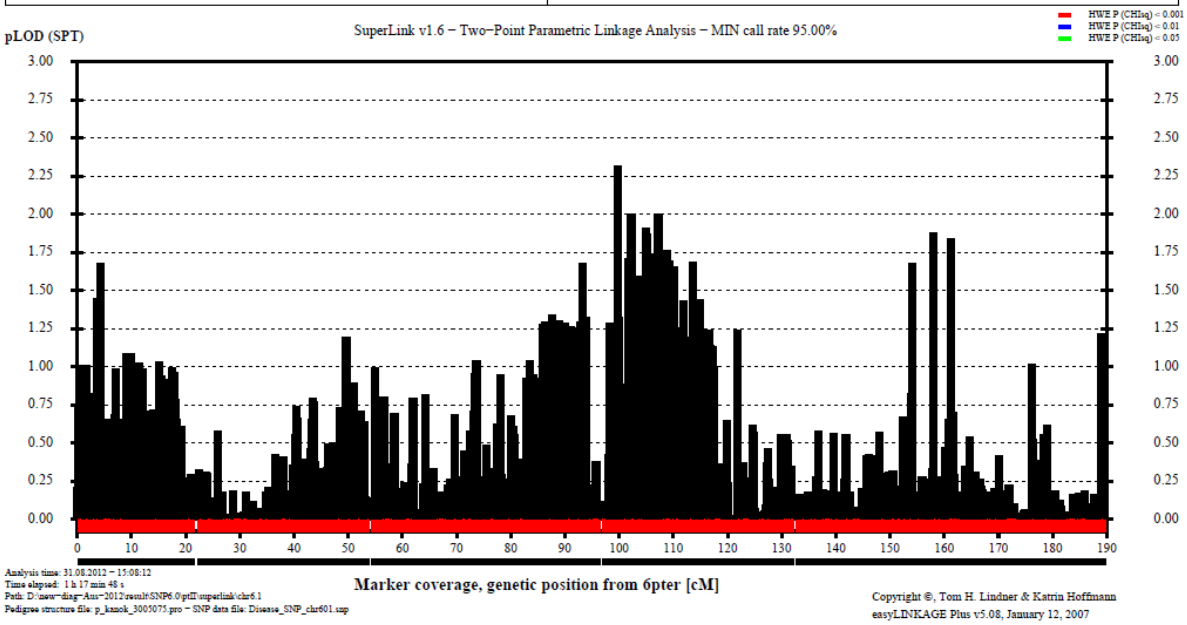


Figure 38 Two-point parametric LOD score was observed on chromosome 6 (file1) by using data from genotyping SNP array 6.0

Project:	CHR6.2	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	18655 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs2428190	6	114.87	2.7871	0.0000
2. rs434993	6	102.05	2.5196	0.0500
3. rs401858	6	102.05	2.5196	0.0500
4. rs17056117	6	101.70	2.3341	0.0500
5. rs11752153	6	101.72	2.3341	0.0500

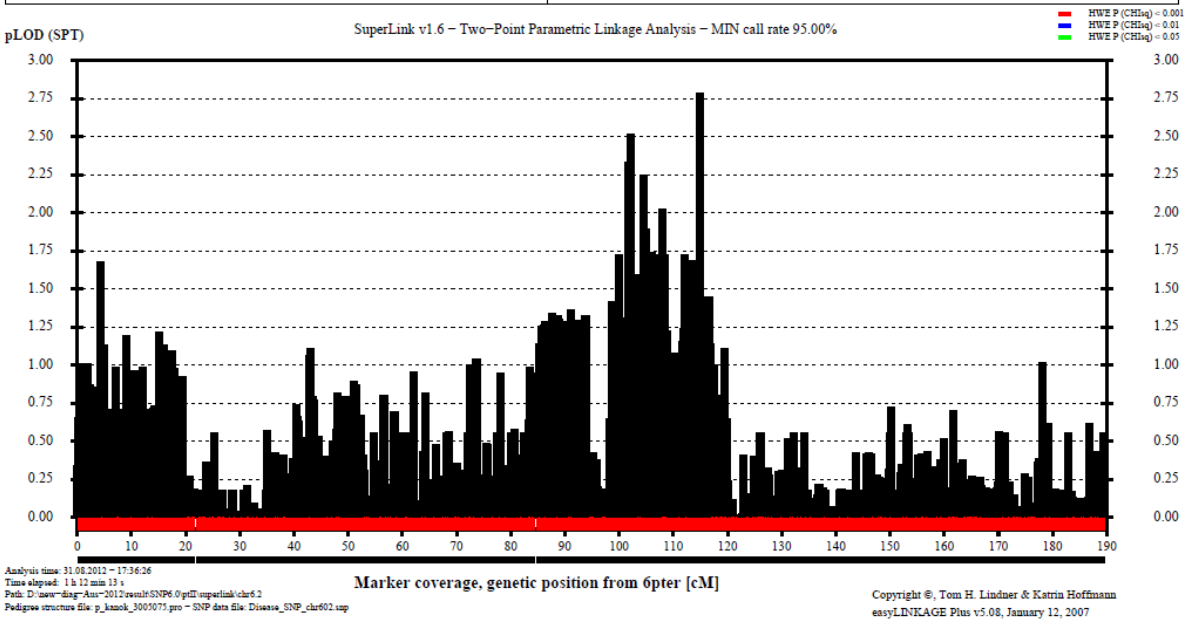


Figure 39 Two-point parametric LOD score was observed on chromosome 6 (file2) by using data from genotyping SNP array 6.0

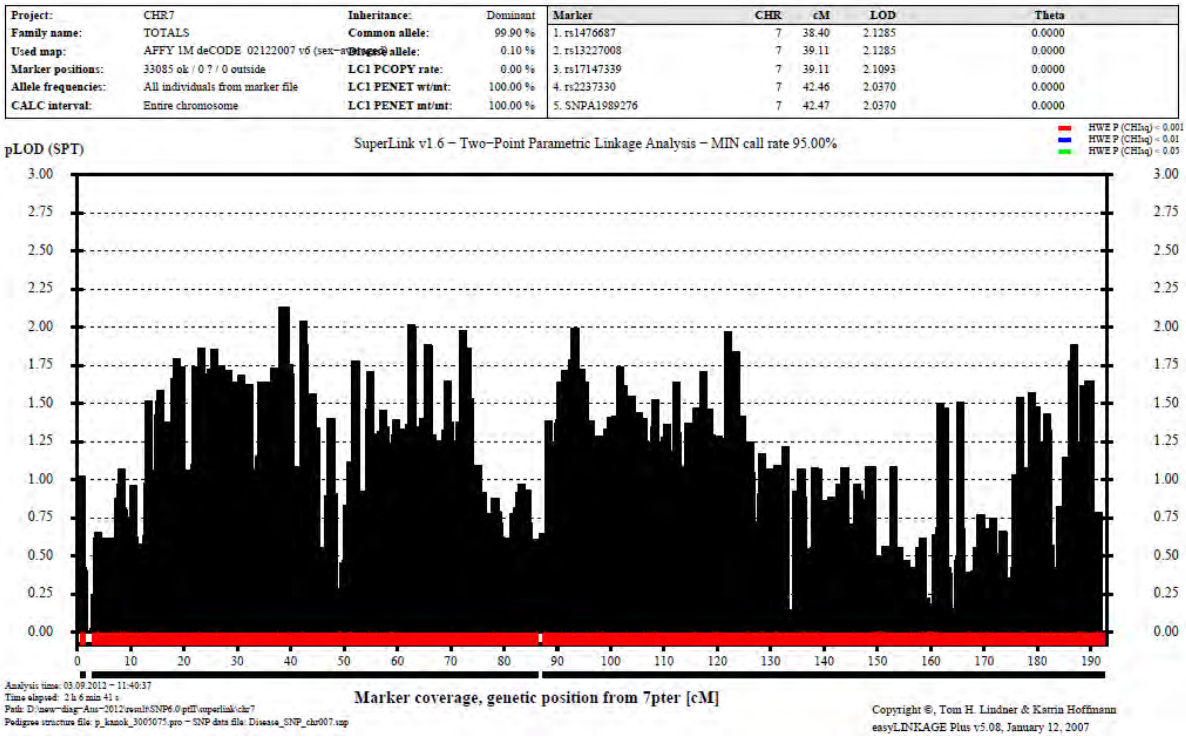


Figure 40 Two-point parametric LOD score was observed on chromosome 7 by using data from genotyping SNP array 6.0

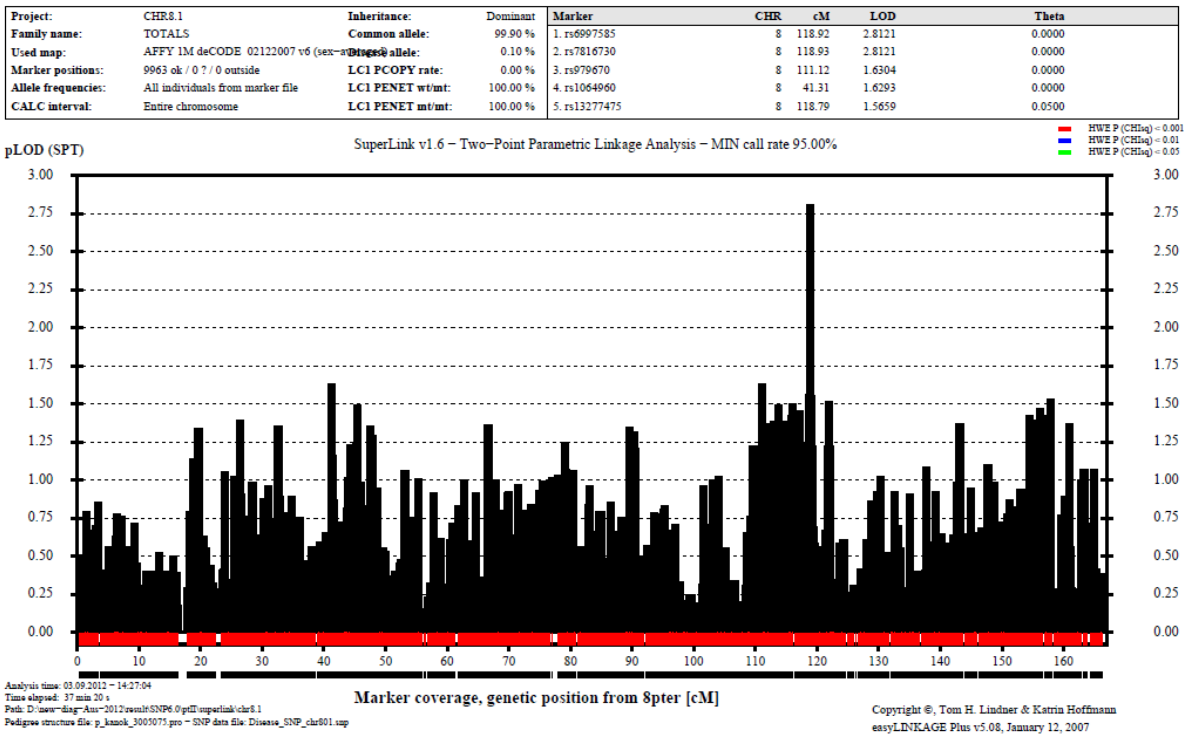


Figure 41 Two-point parametric LOD score was observed on chromosome 8 by using data from genotyping SNP array 6.0

Project:	CHR8.2	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	9975 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs12548886	8	113.24	1.8140	0.0000
2. rs7818918	8	113.29	1.7234	0.0000
3. rs1560308	8	113.07	1.6773	0.0500
4. rs13264061	8	113.28	1.6265	0.0000
5. rs13249969	8	118.98	1.5703	0.0000

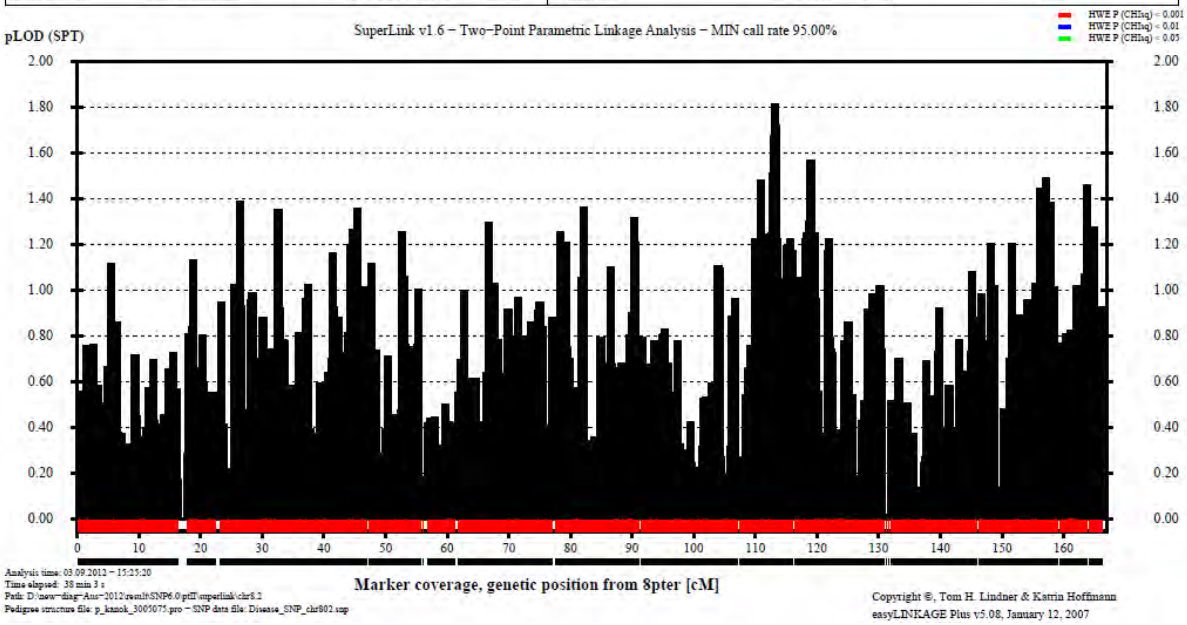


Figure 42 Two-point parametric LOD score was observed on chromosome 8 by using data from genotyping SNP array 6.0

Project:	CHR9	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	28999 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs657659	9	111.35	2.5447	0.0000
2. rs1556939	9	111.37	2.5447	0.0000
3. rs11792080	9	111.29	2.5149	0.0000
4. rs158689	9	94.54	2.5073	0.0000
5. rs10283612	9	3.51	2.4606	0.0000

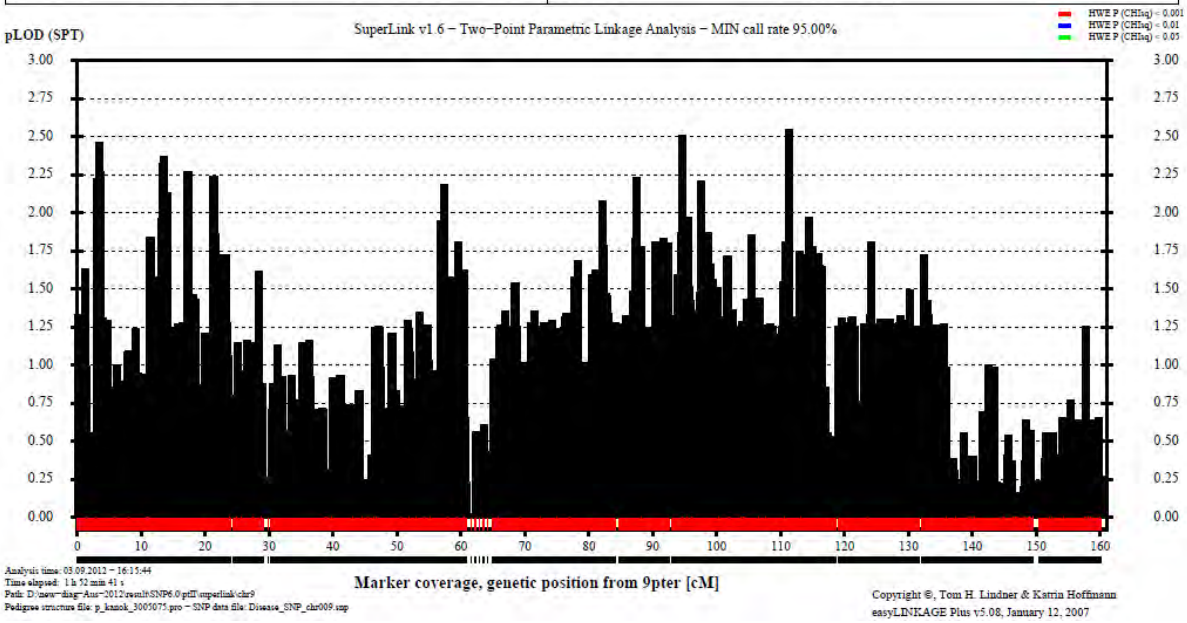


Figure 43 Two-point parametric LOD score was observed on chromosome 9 by using data from genotyping SNP array 6.0

Project:	CHR10	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	33056 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1113145	10	105.19	2.5781	0.0000
2. rs7097606	10	108.12	2.5731	0.0000
3. rs662166	10	105.09	2.5294	0.0000
4. rs2664395	10	106.78	2.5294	0.0000
5. rs4601665	10	48.67	2.5257	0.0000

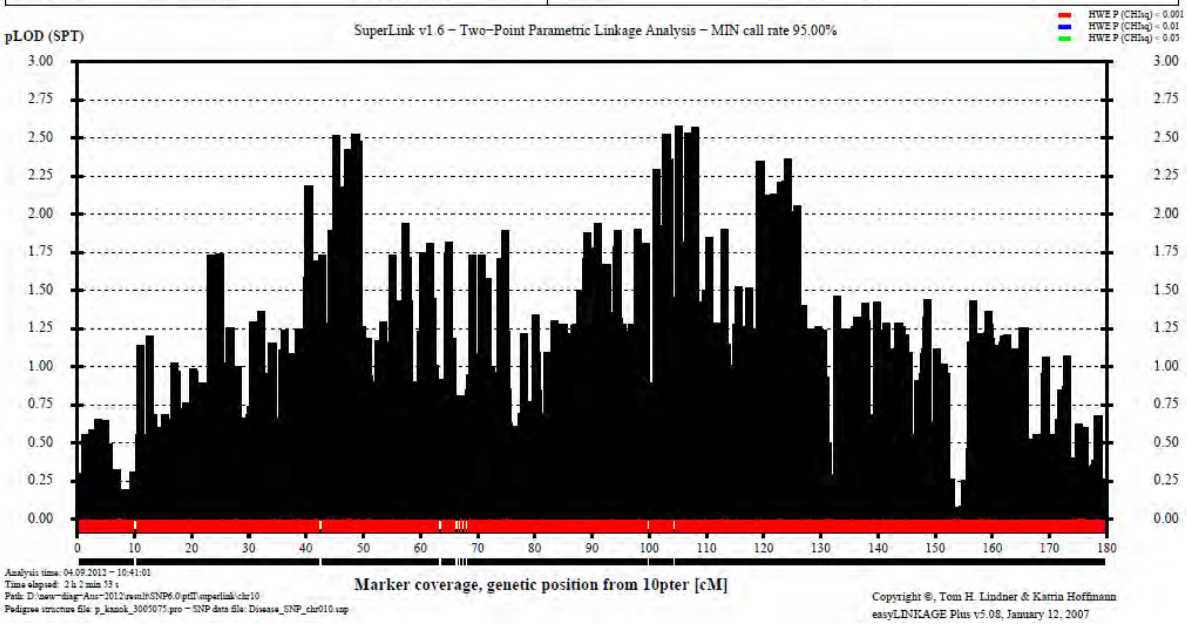


Figure 44 Two-point parametric LOD score was observed on chromosome 10 by using data from genotyping SNP array 6.0

Project:	CHR11	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	30223 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs10742304	11	48.79	2.5427	0.0000
2. rs4755367	11	50.57	2.5265	0.0000
3. rs10160235	11	56.99	2.4423	0.0000
4. rs6485536	11	61.00	2.3650	0.0000
5. rs495033	11	117.02	2.3160	0.0500

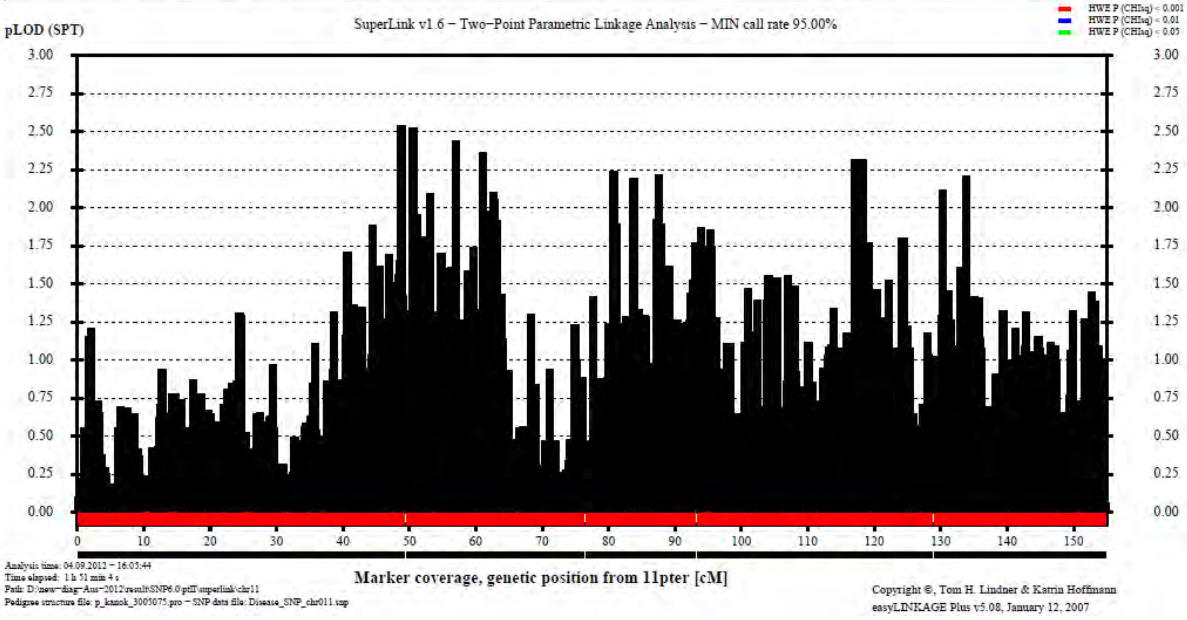


Figure 45 Two-point parametric LOD score was observed on chromosome 11 by using data from genotyping SNP array 6.0

Project:	CHR12	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	29120 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs936954	12	67.08	2.0867	0.0000
2. rs936956	12	67.09	2.0867	0.0000
3. rs2354997	12	73.02	2.0507	0.0000
4. rs10877300	12	73.50	2.0507	0.0000
5. rs10876317	12	67.07	2.0154	0.0000

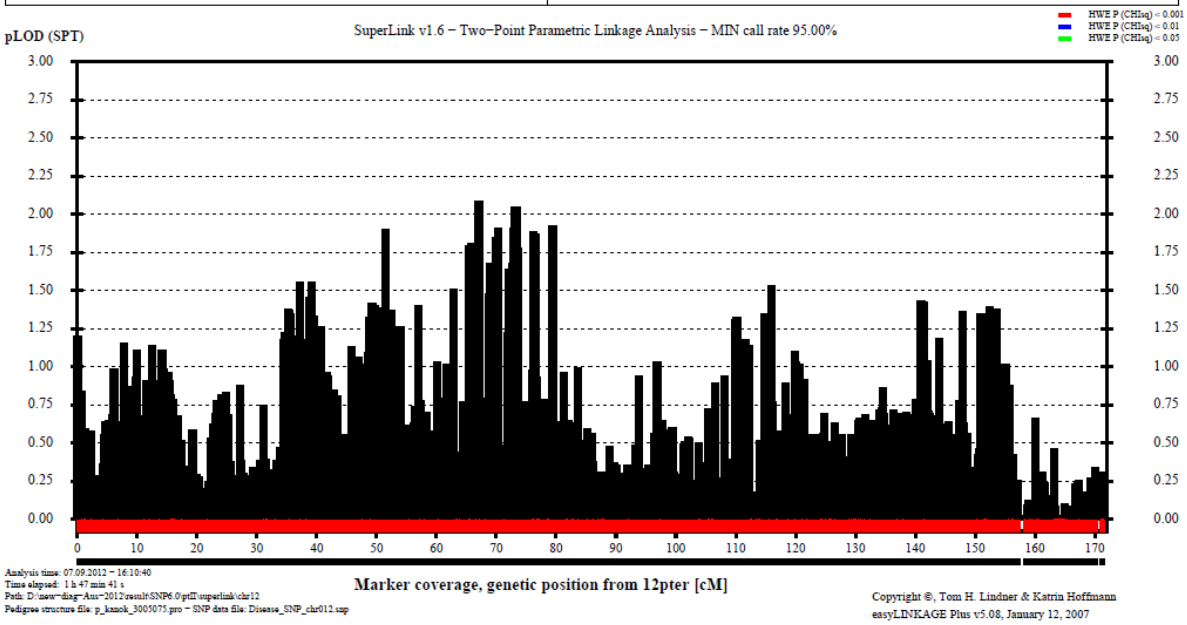


Figure 46 Two-point parametric LOD score was observed on chromosome 12 by using data from genotyping SNP array 6.0

Project:	CHR13	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	23138 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs9510718	13	8.76	2.8102	0.0000
2. rs9507294	13	11.25	2.5443	0.0000
3. rs754116	13	8.18	2.3716	0.0000
4. rs754115	13	8.18	2.3716	0.0000
5. rs9552826	13	8.40	2.2492	0.0000

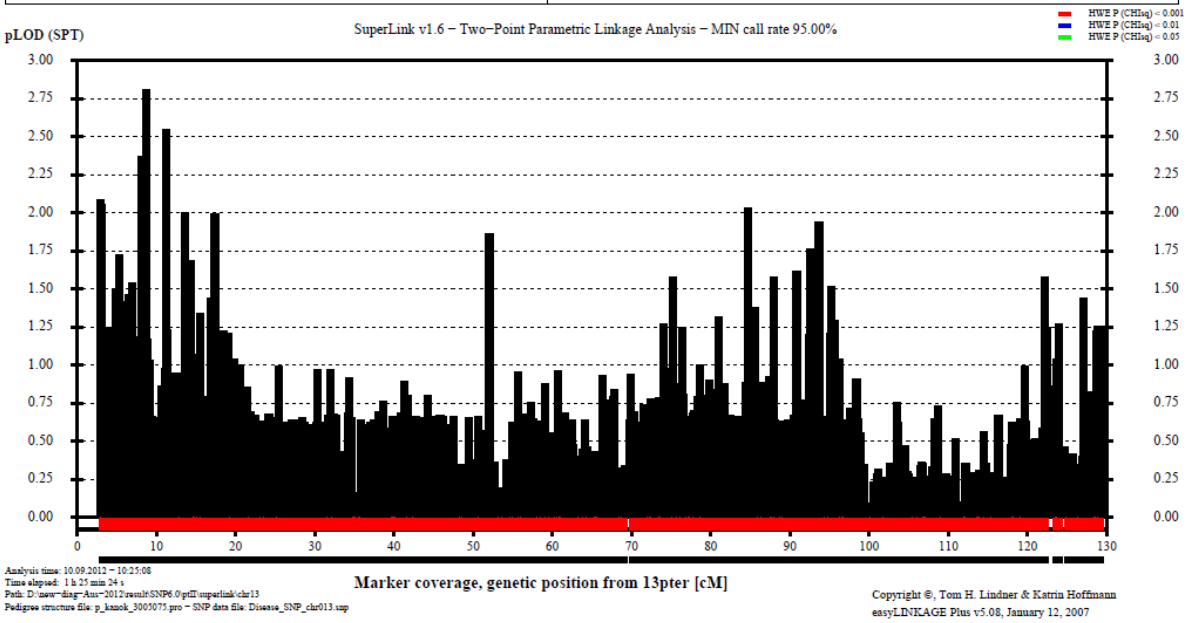


Figure 47 Two-point parametric LOD score was observed on chromosome 13 by using data from genotyping SNP array 6.0

Project:	CHR14	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-a	Disease allele:	0.10 %
Marker positions:	19519 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/wt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1243602	14	6.08	1.5060	0.0000
2. rs8020432	14	34.80	1.3942	0.0000
3. rs986555	14	102.69	1.3177	0.0000
4. rs986555	14	102.69	1.3176	0.0000
5. rs909073	14	13.28	1.3167	0.0000

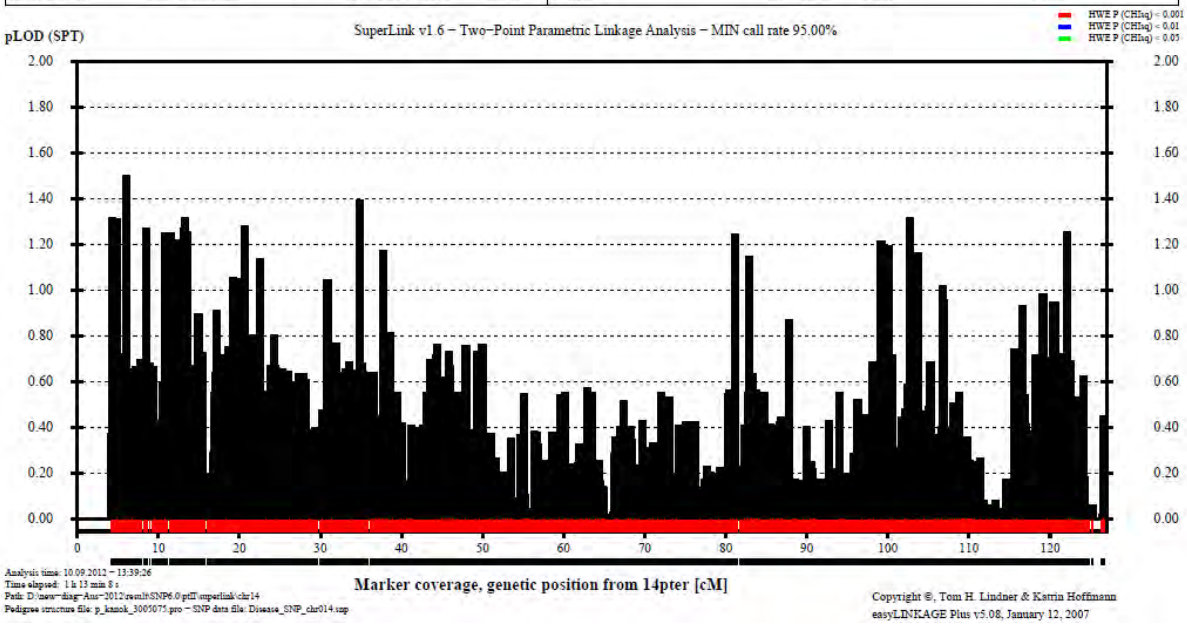
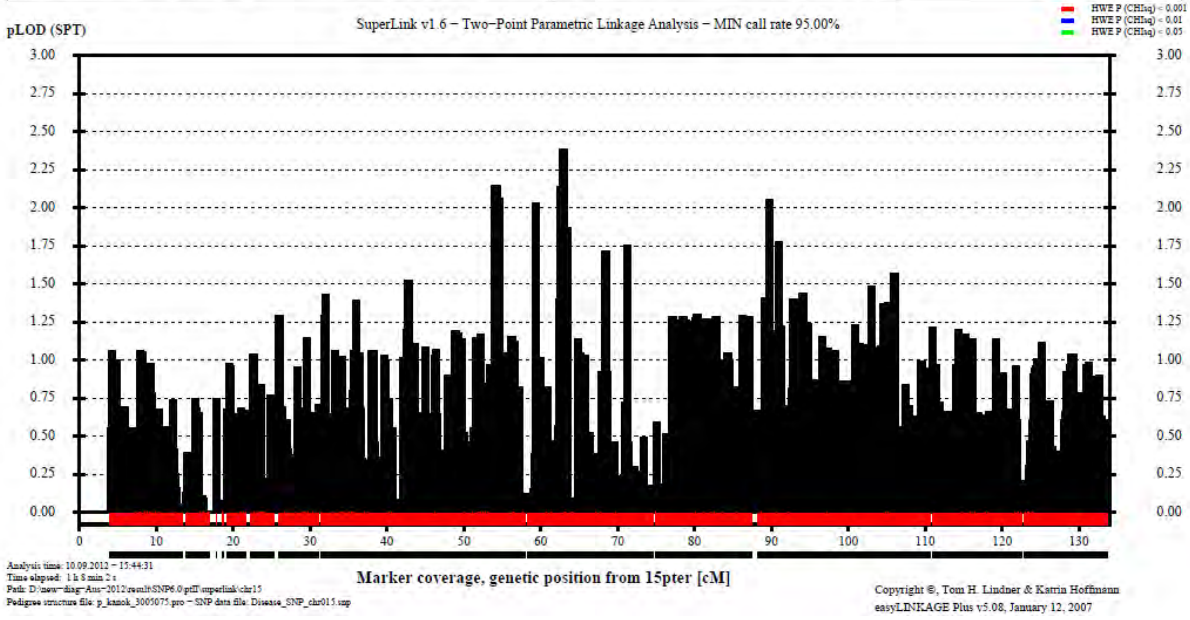


Figure 48 Two-point parametric LOD score was observed on chromosome 14 by using data from genotyping SNP array

Project:	CHR15	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-a	Disease allele:	0.10 %
Marker positions:	18014 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/wt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs12912086	15	62.94	2.3824	0.0000
2. rs7180597	15	62.93	2.2157	0.0000
3. rs12909303	15	62.95	2.2121	0.0000
4. rs34262415	15	54.08	2.1488	0.0000
5. rs12900986	15	54.08	2.1488	0.0000



6.0

Figure 49 Two-point parametric LOD score was observed on chromosome 15 by using data from genotyping SNP array 6.0

Project:	CHR16	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-a	Disease allele:	0.10 %
Marker positions:	18652 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1592560	16	27.89	2.6713	0.0000
2. rs1034392	16	33.02	2.5173	0.0000
3. rs8054134	16	31.04	2.1690	0.0000
4. rs8058552	16	31.04	2.1690	0.0000
5. rs7202772	16	31.04	2.1690	0.0000

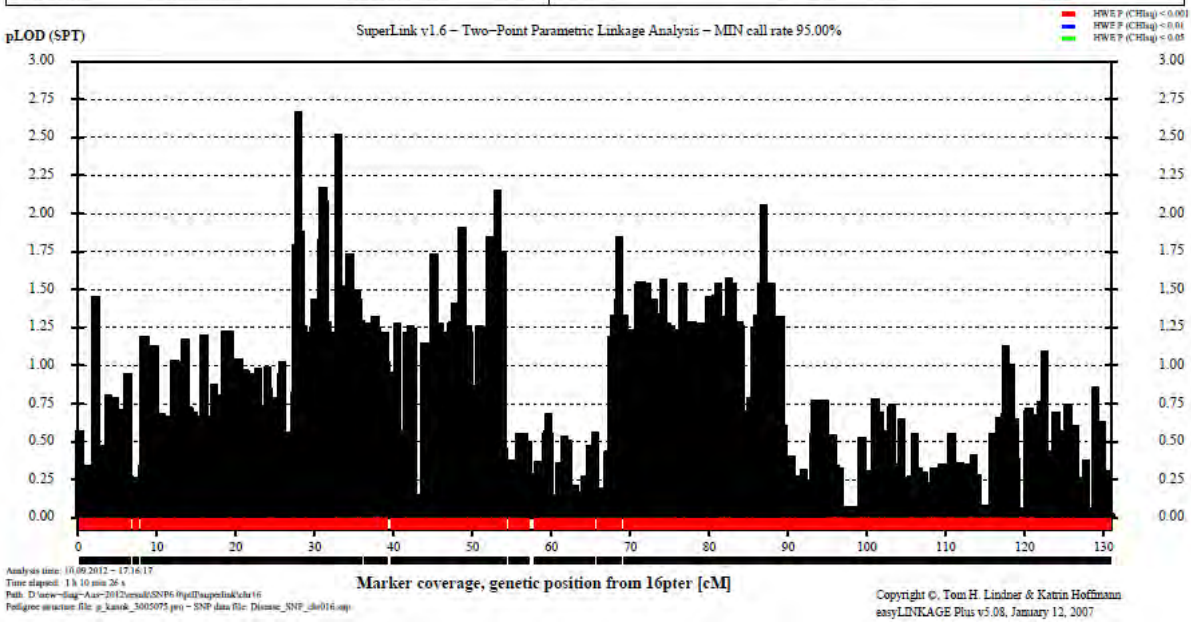


Figure 50 Two-point parametric LOD score was observed on chromosome 16 by using data from genotyping SNP array 6.0

Project:	CHR17	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-a	Disease allele:	0.10 %
Marker positions:	14087 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs12936957	17	80.91	1.9138	0.0000
2. rs2541244	17	80.99	1.9138	0.0000
3. rs2111098	17	81.12	1.9138	0.0000
4. rs171512	17	81.31	1.9138	0.0000
5. rs11079129	17	80.57	1.9014	0.0000

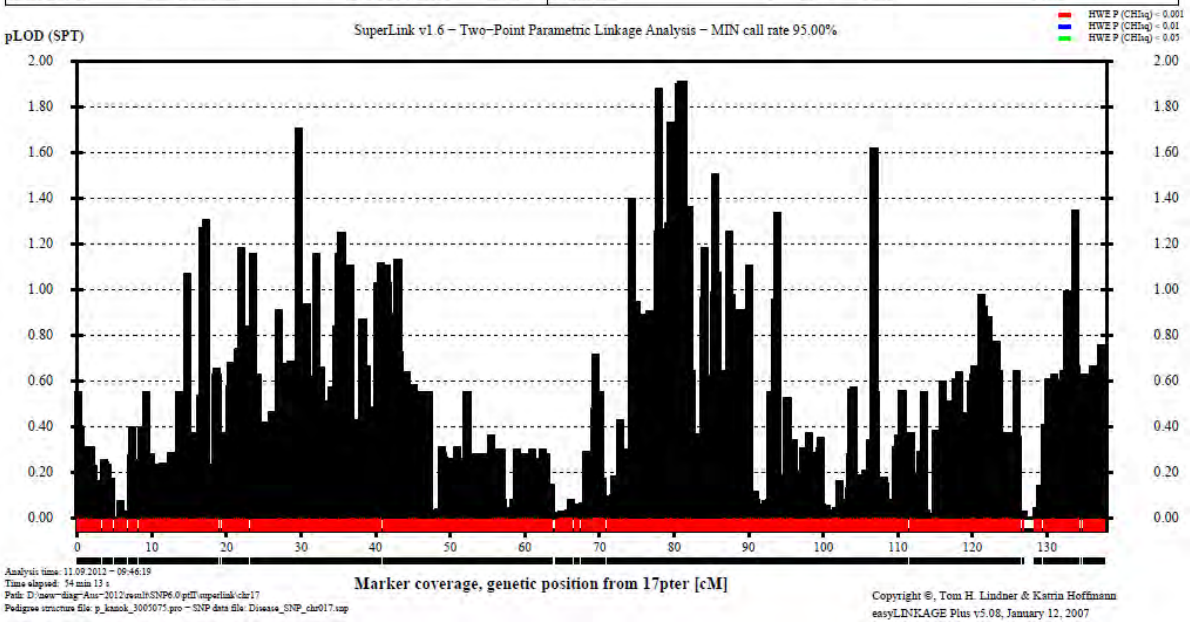


Figure 51 Two-point parametric LOD score was observed on chromosome 17 by using data from genotyping SNP array 6.0

Project:	CHR18	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	17835 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs8082909	18	33.97	2.2081	0.0000
2. rs8088774	18	92.72	1.8638	0.0000
3. rs6567232	18	86.18	1.7901	0.0000
4. rs12969529	18	86.13	1.7842	0.0000
5. rs9945863	18	86.39	1.7522	0.0000

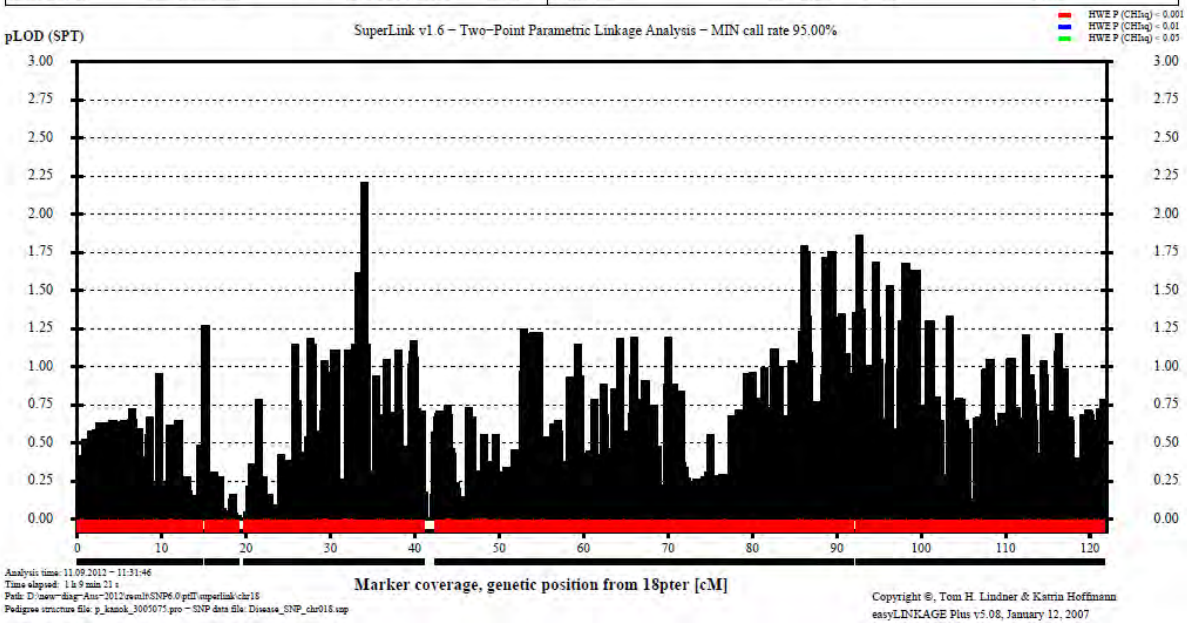


Figure 52 Two-point parametric LOD score was observed on chromosome 18 by using data from genotyping SNP array 6.0

Project:	CHR19	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	8440 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs4801446	19	109.68	1.3589	0.0000
2. rs7250424	19	109.77	1.3163	0.0000
3. rs7254323	19	109.77	1.3163	0.0000
4. rs2215048	19	109.66	1.2816	0.0000
5. rs8108292	19	109.66	1.2816	0.0000

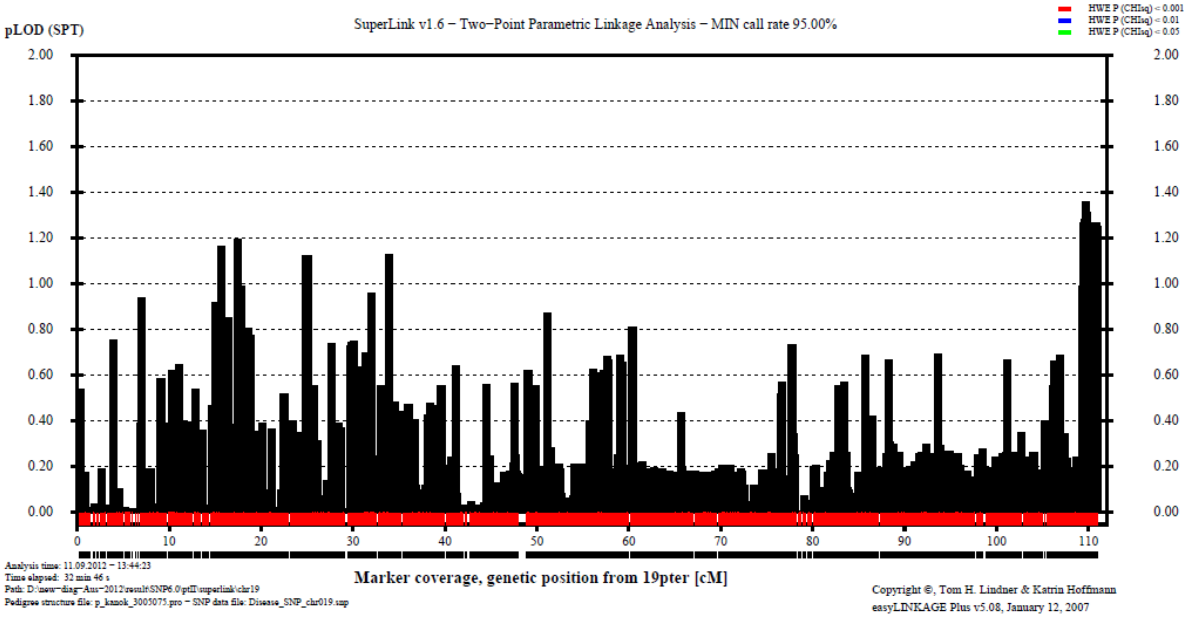


Figure 53 Two-point parametric LOD score was observed on chromosome 19 by using data from genotyping SNP array 6.0

Project:	CHR20	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	15723 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs1045402	20	95.93	2.2359	0.1000
2. rs11696913	20	95.85	1.9608	0.1000
3. rs1078297	20	101.17	1.9544	0.0000
4. rs2064393	20	93.78	1.8597	0.1000
5. rs6099813	20	94.57	1.8025	0.0000

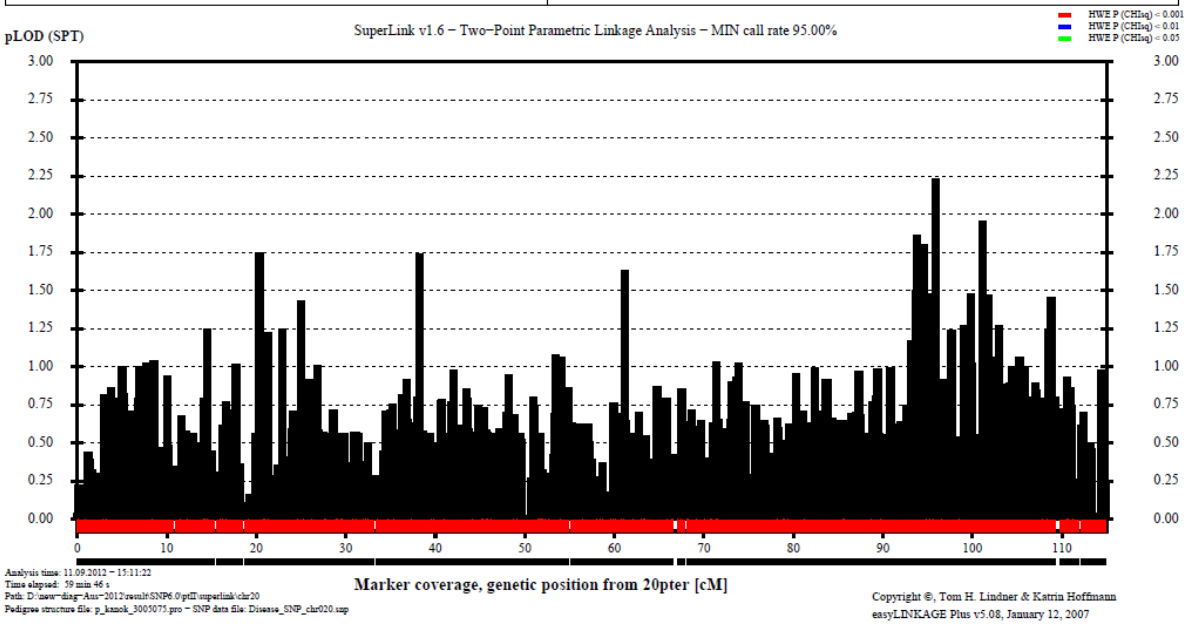


Figure 54 Two-point parametric LOD score was observed on chromosome 20 by using data from genotyping SNP array 6.0

Project:	CHR21	Inheritance:	Dominant
Family name:	TOTALS	Common allele:	99.90 %
Used map:	AFFY 1M deCODE 02122007 v6 (sex-averaged)	Disease allele:	0.10 %
Marker positions:	8966 ok / 0 ? / 0 outside	LC1 PCOPY rate:	0.00 %
Allele frequencies:	All individuals from marker file	LC1 PENET wt/wt:	100.00 %
CALC interval:	Entire chromosome	LC1 PENET mt/mt:	100.00 %

Marker	CHR	cM	LOD	Theta
1. rs200691	21	22.78	1.9584	0.0000
2. rs2210272	21	54.16	1.8401	0.0000
3. rs2825119	21	15.87	1.8022	0.0000
4. rs17275302	21	7.74	1.7743	0.0000
5. rs2824385	21	12.35	1.7743	0.0000

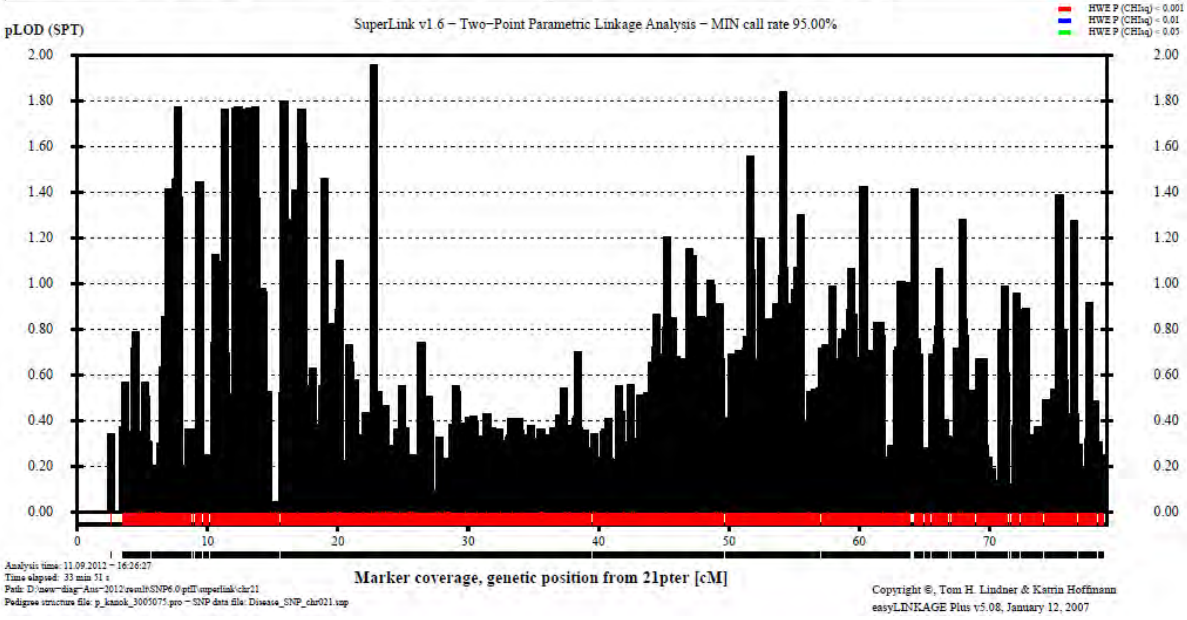


Figure 55 Two-point parametric LOD score was observed on chromosome 21 by using data from genotyping SNP array 6.0

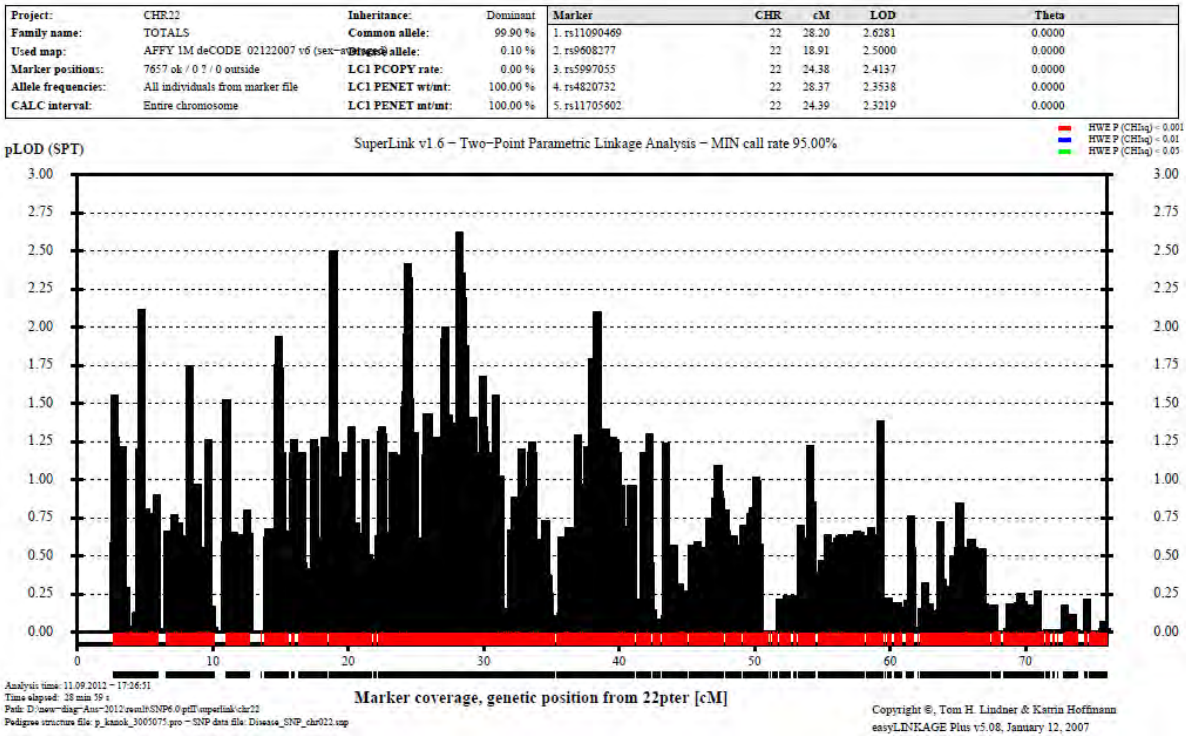


Figure 56 Two-point parametric LOD score was observed on chromosome 22 by using data from genotyping SNP array 6.0

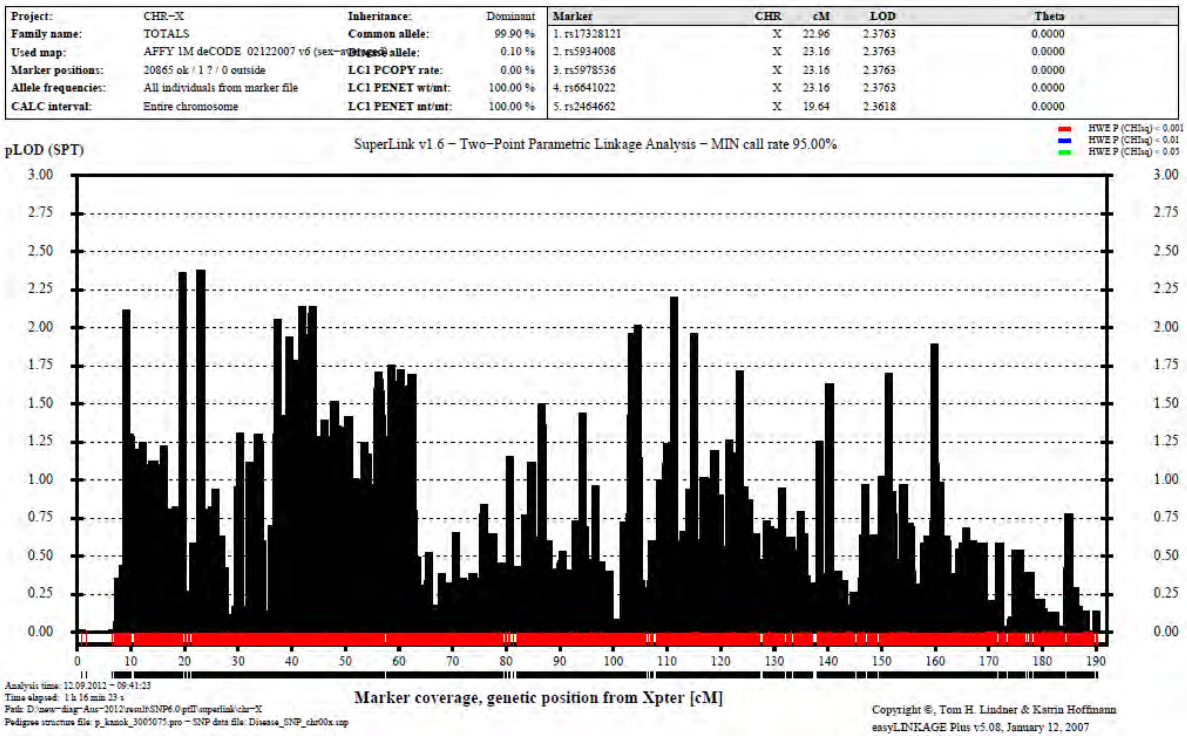


Figure 57 Two-point parametric LOD score was observed on chromosome X by using data from genotyping SNP array 6.0

Table 9 Two-point parametric LOD score result in each chromosome by using SuperLink v1.6

Chromosome	Two-Point Parametric Linkage Analysis			
	SNP array 6.0		Affymetric 10K	
	LOD	Region (cM)	LOD	Region (cM)
1	2.6879	160.85	1.5865	122.26
2	2.2206	67.27	1.2435	242.48
3	2.4918	138.95	1.5501	149.20
4	2.6309	74.29	2.4220	78.04
5	2.8981	101.08	1.3714	111.60
6	2.7871	114.87	1.8204	101.75
7	2.1285	38.40	1.2267	16.53
8	2.8121	118.92	1.2426	113.75
9	2.5447	111.35	1.8106	14.03
10	2.5781	105.91	2.3316	124.29
11	2.5427	48.79	2.2206	48.28
12	2.0867	67.08	1.2857	48.68
13	2.8102	8.76	1.2622	8.24
14	1.5060	6.08	1.0573	19.19
15	2.382	62.94	1.4287	32.07
16	2.6713	27.89	1.5402	86.73
17	1.9138	80.91	1.2294	79.22
18	2.2081	33.97	1.3601	98.20
19	1.3589	109.68	0.7294	27.82
20	2.2359	95.93	1.0393	75.81
21	1.9584	22.78	1.1389	60.33
22	2.6281	28.20	2.0000	27.20
X	2.3763	22.96	1.2422	43.67

Table 10 Selected chromosomes which show high LOD score ≥ 2.5 by using SNP array 6.0

Chromosome	Two-Point Parametric Linkage Analysis by Superlink v1.6			
	LOD	Highest LOD Position (cM)	Selected Region (cM)	Selected Position (NCBI37/hg19)
1	2.6879	160.85	110-170	86,072,408 – 168,332,730
4	2.6309	74.29	50-80	29,903,074 – 69,639,729
5	2.8981	101.08	100-120	82,383,876-114,669,438
6	2.7871	114.87	95-120	89,919,826 – 121,465,011
8	2.8121	118.92	110-120	103,367,599 – 120,186,415
9	2.5447	111.35	90-120	91,344,293 – 116,664,097
10	2.5781	105.91	40-50	16,547,058 – 26,206,856
11	2.5427	48.79	40-65	24,087,296 – 60,037,567
13	2.8102	8.76	2-20	26,695,151 – 28,512,585
16	2.6713	27.89	20-40	7,723,572 - 18,910,039
22	2.6281	28.20	15-35	22,596,259 – 33,611,894

7. Variants identification by using WES

We performed WES in four members from this family including 2 diabetic patients (8310, 8311) and 2 non-diabetic subjects (8101, 8203). The clinical characteristics of selected members are shown in Table 11.

Table 11 The clinical characteristics of selected members who were performed by WES

Clinical data/Member ID	8101	8203	8310	8311
Diabetes status	unaffected	unaffected	affected	affected
Sex	Female	Female	Female	Female
Current Age	96	63	36	30
Age at diagnosis (year)	-	-	23	29
Number of generation	1	2	3	3
BMI (kg/m^2)	15.31	20.78	36.68	35.29
Current treatment	-	-	Metformin HCl	-
FPG (age) [mg/dl]	88 (96)	84 (54)	238 (23)	165 (29)
HbA1c (age) [%]	5.6 (96)	NA	11.1 (23)	7.0 (29)
Total cholesterol [mg/dl]	249	165	202	204
Triglyceride [mg/dl]	169	136	101	128
LDL [mg/dl]	173.2	211	123.8	138.4
HDL [mg/dl]	42	53	58	40

Samples were prepared as an Illumina sequencing library, and in the second step, the sequencing libraries were enriched for the desired target using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencer (Figure 1). From WES, we found 67,802 variants per exome on average. Approximately 19,205 are coding variants per individual (Table 12).

Table 12 Summary of data analysis from whole exome sequencing in selected subjects

Item	8310	8203	8311	8101
Total reads	65,367,484	48,419,904	62,826,614	61,840,608
Total yield (bp)	6,602,115,884	4,890,410,304	6,345,488,014	6,245,901,408
Read length (bp)	101	101	101	101
Target regions (bp)	62,085,286	62,085,286	62,085,286	62,085,286
Average throughput depth of target regions			102.2X	100.6X
Mappable reads	58,521,087	43,304,691	48,019,058	46,161,491
Mappable yield (bp)	5,572,350,908	4,132,926,007	4,748,322,541	4,557,814,606
% Mappable reads	89.50%	89.40%	76.40%	74.60%
On-target reads	41,928,940	31,019,124	34,846,562	32,352,673
On-target yield (bp)	3,270,817,493	2,418,700,395	2,802,455,550	2,611,847,914
% On-target reads	71.60%	71.60%	72.60%	70.10%
% On-target reads	64.10%	64.10%	55.50%	52.30%
% Coverage of target regions (more than 1X)	93.50%	92.60%	93.50%	93.80%
Number of on-target genotypes	58,080,356	57,461,617	58,041,408	58,254,488
% Coverage of target regions (more than 10X)	82.00%	77.80%	85.30%	85.70%
Number of on-target genotypes	50,904,439	48,298,374	52,973,876	53,194,389
Median read depth of target regions	46.0X	34.0X	43.0X	40.0X
Mean read depth of target regions	52.7X	39.0X	45.1X	42.1X
Number of SNPs	64,723	61,409	72,723	76,413
Number of coding SNPs	18,592	18,123	19,849	21,291
Number of synonymous SNPs			10,225	11,050
Number of nonsynonymous SNPs			9,127	9,704
Number of Indels	13,361	12,669	7,272	7,700
Number of coding Indels	526	482	354	392

8. Candidate variants selection

We selected candidate variants based on 4 criteria (i) variants are identified only in diabetic patients and not identified in non-diabetic patients were selected; (ii) heterozygous SNPs were selected because MODY is an autosomal dominant inheritance; (iii) we selected non-synonymous, nonsense, splice site, frame shift deletion and frame shift insertion because most of pathogenic mutations located on exon or splice-site which may effected on protein structure or function; (iv) MODY is a rare Mendelian disorder, variants with MAF < 5% in the 1000 Genome Project and NIEHS Exome Project data and no data in dbSNP135 were included.

By focusing on variants of interest (non-synonymous, nonsense, splice-site or frame shift mutations), which is a heterozygous and were identified only in two diabetic patients (8310, 8311), 296 variants were chosen. Furthermore, we only included variants with a MAF<5% in the 1000 Genome Project and NIEHS Exome Project data. It remained 148 variants, including 63 variants not reported in the database dbSNP135 (Table 13). Then, we only selected novel heterozygous-nonsynonymous variants within linkage loci based on an LOD score $\geq$ 2.5. It remained only 6 variants in 6 genes located in linkage loci including p.P1060L in *WNK2* and p.M920L in *PTCH1* on chromosome 9, p.K129N in *NEBL* on chromosome 10, p.G101C in *PAMR1* and p.K295N in *OR4D11* on chromosome 11 and p.A48T in *GGT5* on chromosome 22 (Table 14).

Table 13 Overview of all selected variants after variants reduction

Criteria filtrations	Number of SNPs which present in affected and not present in nonaffected				
	nsSNP	nonsense	Splice site	Fram_del/ins	Total
Heterozygous SNPs	278	3	15	0	296
1000G < 0.05 or NIEHS $\leq$ 5 %	138	3	7	0	148
No data in dbSNPs135	57	2	4	0	63

Table 14 Number of variants identified in linkage loci

Chromosome	Linkage Loci (NCBI37/hg19)	Identified heterozygous, non-synonymous variants in only diabetic patients (8310,8311) and MAF≤5%	Novel variation (not found in dbSNP 135)
1	86,072,408 – 168,332,730	11	0
4	29,903,074 – 69,639,729	1	0
5	82,383,876-114,669,438	1	0
6	89,919,826 – 121,465,011	2	0
8	103,367,599 – 120,186,415	0	0
9	91,344,293 – 116,664,097	4	2
10	16,547,058 – 26,206,856	2	1
11	24,087,296 – 60,037,567	3	2
13	26,695,151 – 28,512,585	0	0
16	7,723,572 - 18,910,039	0	0
22	22,596,259 – 33,611,894	8	1
Total		32	6

The 6 variants on 6 genes were then used to predict effect of variant on protein function by *in silico* programs (PolyPhen V2.0.23, SIFT, Mutation Taster and VarioWatch). There are 2 variants on 2 genes showed high effect on protein function in at least 3 program including p.M920L in *PTCH1* which present high effect in all programs and p.P1060L in *WNK2* which present high effect in 3 of 4 programs. Moreover, both genes are expressed in pancreatic cells. Therefore, they were chosen to analyse the segregation with diabetes in this family (Table 15).

Table 15 List of novel nonsynonymous variants on the linkage regions and effect of variants on protein function by *in silico* programs

Chromosome	Position	Gene	Variants	dbSNP135/1000G/ NIEHS frequency[a]	PolyPhen 2/ SIFT/Mutation Taster/VarioWatch[b]
9	96,024,208 98,218,653	WNK2 PTCH1	c.C3179T:p.P1060L c.A2758T:p.M920L	-/0/0 -/0/0	+ / + / - / + ++ / + / + / +
10	21,169,816	NEBL	c.A387C:p.K129N	-/-/-	+ / - / + / -
11	35,513,671	PAMR1 OR4D11	c.G301T:p.G101C c.G885C:p.K295N	-/-/- -/-/-	++ / n.a. / + / ± -/ + / - / +
22	59,271,933	GGT5	c.G142A:p.A48T	-/-/-	+ / - / + / -

[a]Protein prediction program including **PolyPhen2**: -benign, + possibly damaging, ++ probably damaging/ **SIFT**: - tolerated, + deleterious/ **Mutation Taster**: - polymorphism, + disease cuasing/ **VarioWatch**: - low risk, ± medium risk, + high risk, ++ very high risk.

[b]Allele frequencies from 1000 Genome project (2011) and NIEHS project.

Abbreviation: Chr, chromosome; 1000G, the 1000 Genome project; NIEHS, National Institute of Environmental Health Sciences (NIEHS) Exome Project; n.a, not analysis.

9. Confirmation of exome sequencing by direct sequencing and segregation analysis

We performed PCR-RFLP and Sanger sequencing to determine segregation of the candidate variants with diabetic member in this family and confirm the exome findings, respectively. We found that p.M920L in *PTCH1* and p.P1060L in *WNK2* partial segregated with diabetes in this family. Both candidate variants were identified in the same sample including sample number 8201, 8208, 8209, 8218, 8219, 8224, 8225, 8310, 8311, 8318, 8321 (**Fig 58**). All samples are diabetes member except the sample number 8318 and 8321 which may be got diabetes in the future. Both variants were validated by Sanger sequencing (**Fig 59, Fig 60**). The other 4 variants were also performed by PCR-RFLP and they did not segregate with diabetes member in this family (Fig 61- Fig 64).

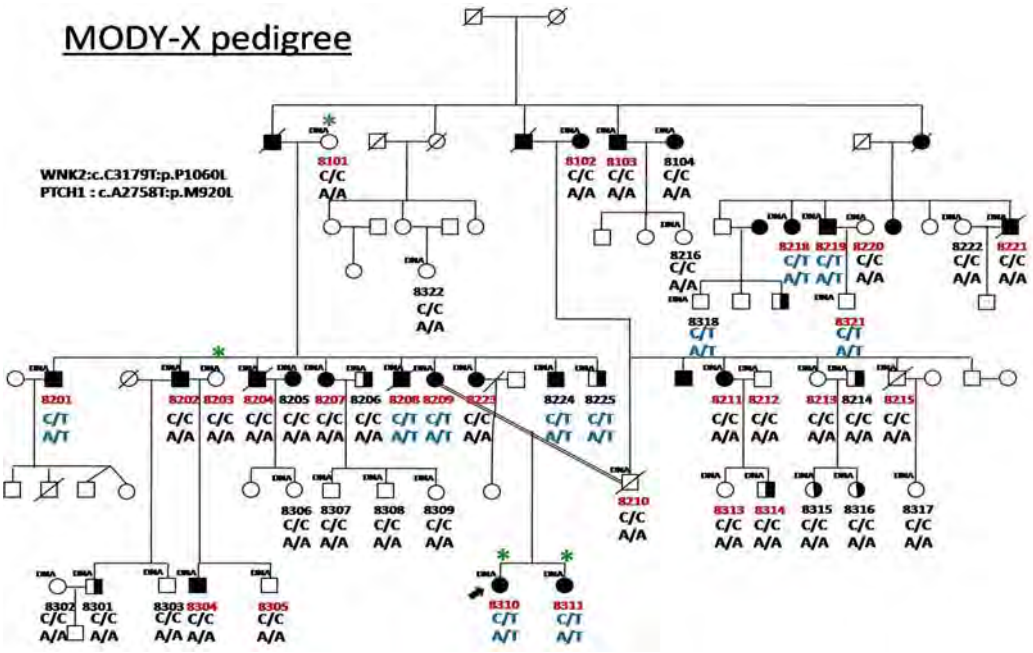


Figure 58 Segregation analyses in MODY-X family. Circles and squares represent female and male participants, respectively. A slash through the symbol indicates that the family member is deceased. Affected with diabetes and unaffected are represented by black symbols and white symbols, respectively. The half-filled symbols indicate individuals with impaired glucose tolerance or impaired fasting glucose. The first row is genotyping result of p.P1060L in WNK2 and the second row is p.M920L in PTCH1 in MODY-X family

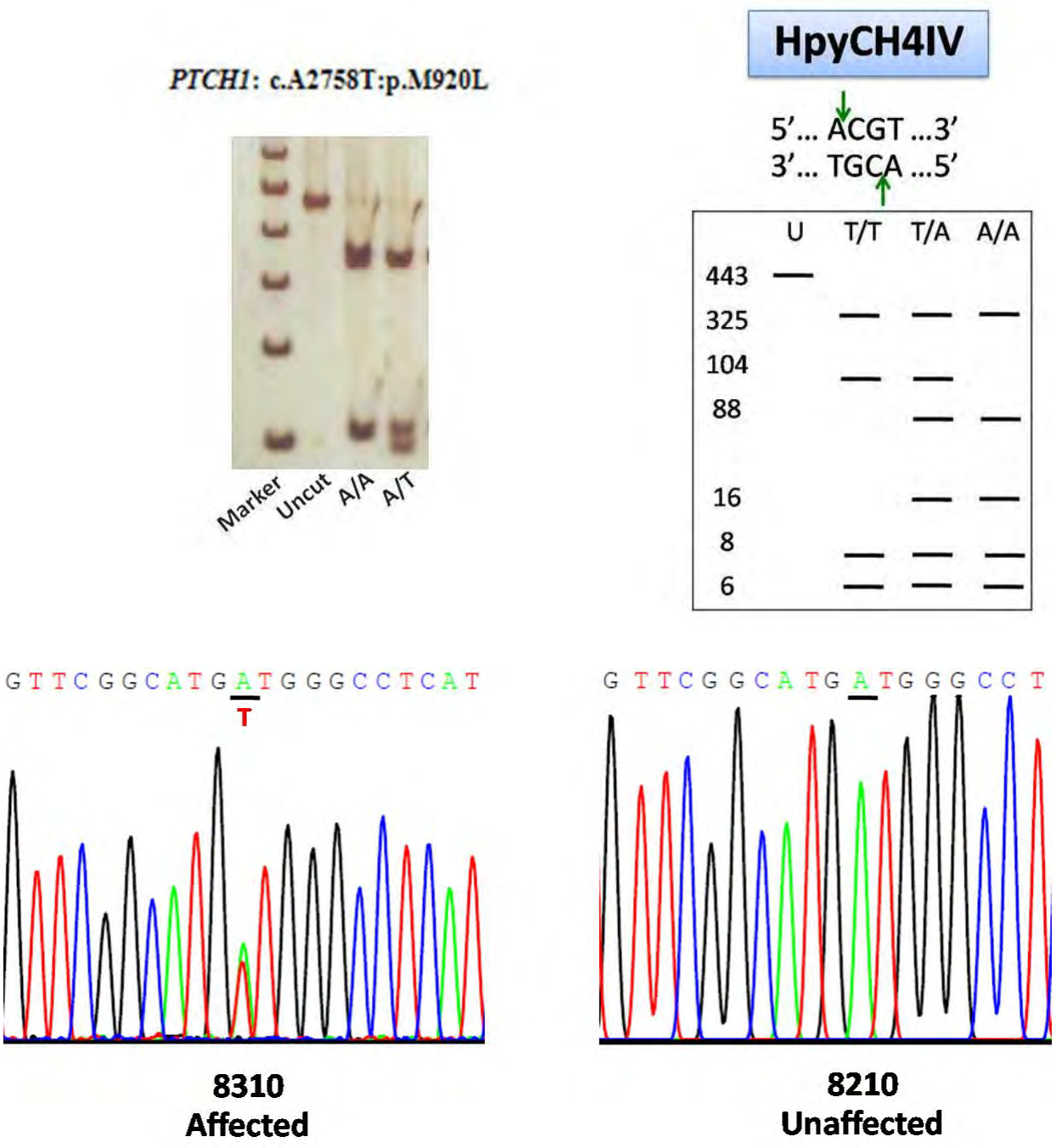


Figure 59 RFLP result and Sanger sequencing chromatograms of p.M920L in *PTCH1* are shown upper panels and bottom panels, respectively. Predicted amino acid changes are indicated upper the sequence. Mutated nucleotides are indicated using black arrow. M, Methionine; L, Leucine.

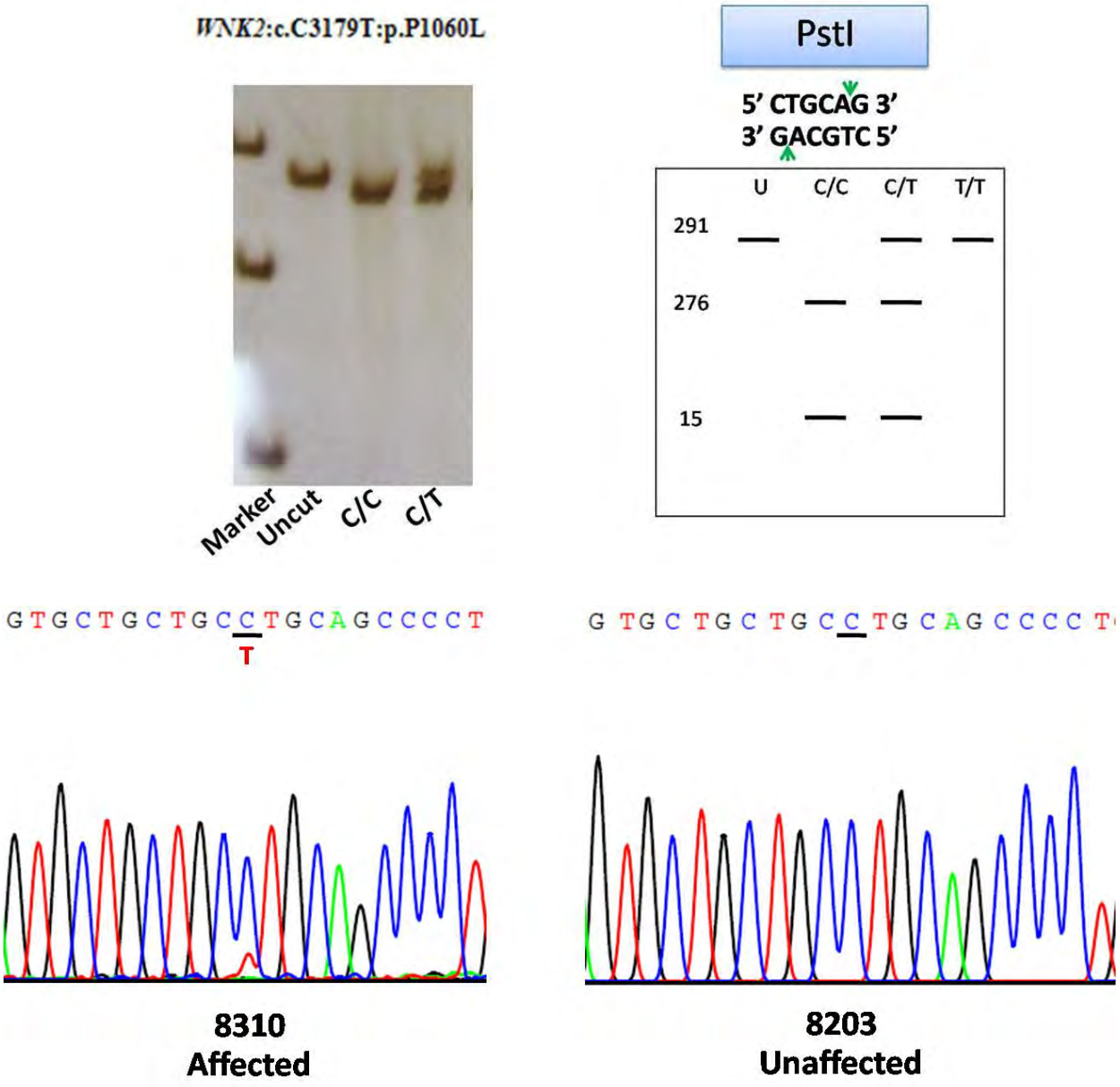


Figure 60 RFLP result and Sanger sequencing chromatograms of p.P1060L in *WNK2* are shown upper panels and bottom panels, respectively. Predicted amino acid changes are indicated upper the sequence. Mutated nucleotides are indicated using black arrow. L, Leucine; P, proline.

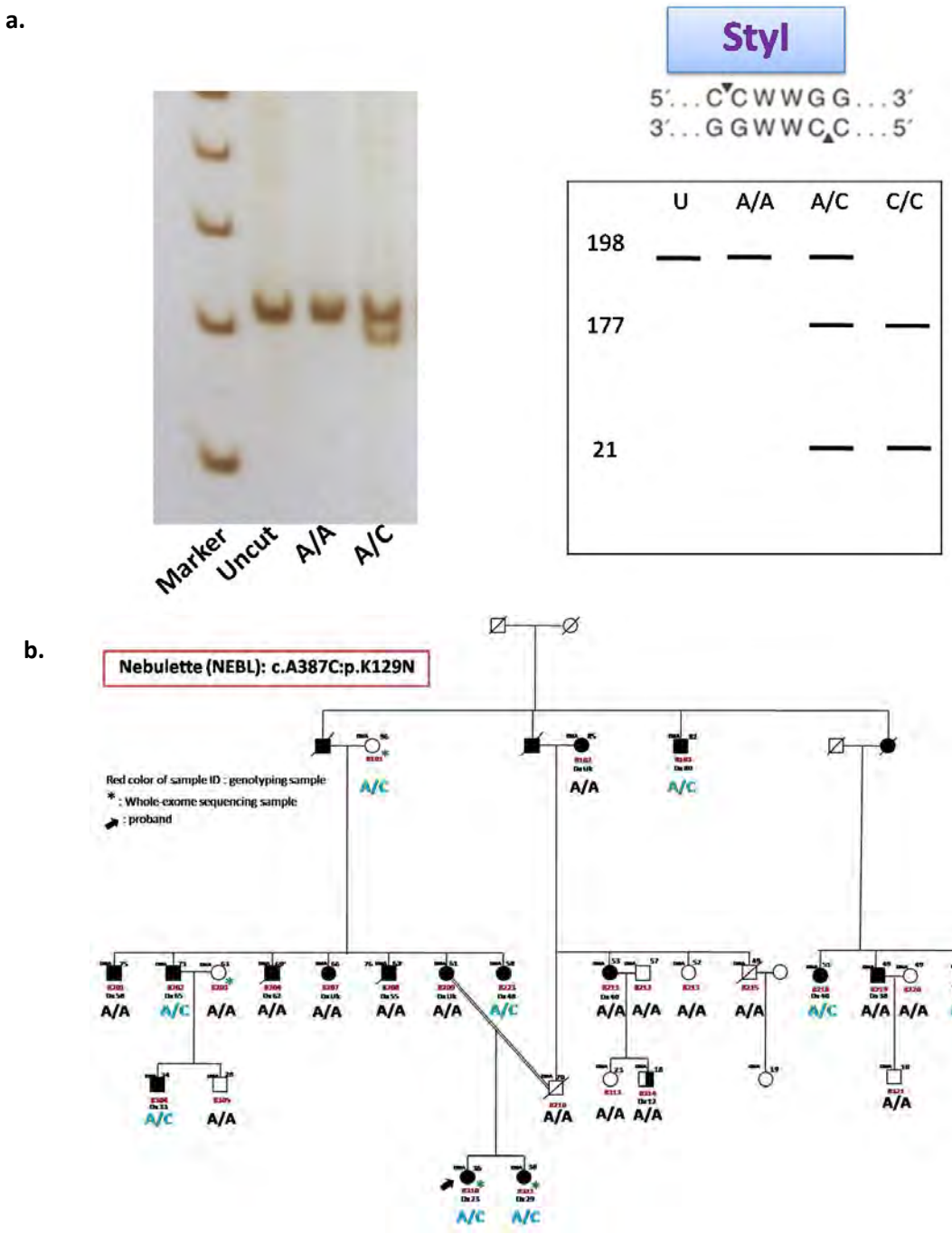


Figure 61 Genotyping result in 27 selected family members. Variant, p. K129N in *NEBL*, was genotyped by RFLP method. **a**; represent genotyping result by using restriction enzyme StyI and pattern of homozygous wild type (A/A) and heterozygous mutant (A/C) was shown. **b**; represent genotyping result in 27 family members who were performed by linkage analysis. K, Lysine; N, Asparagine.

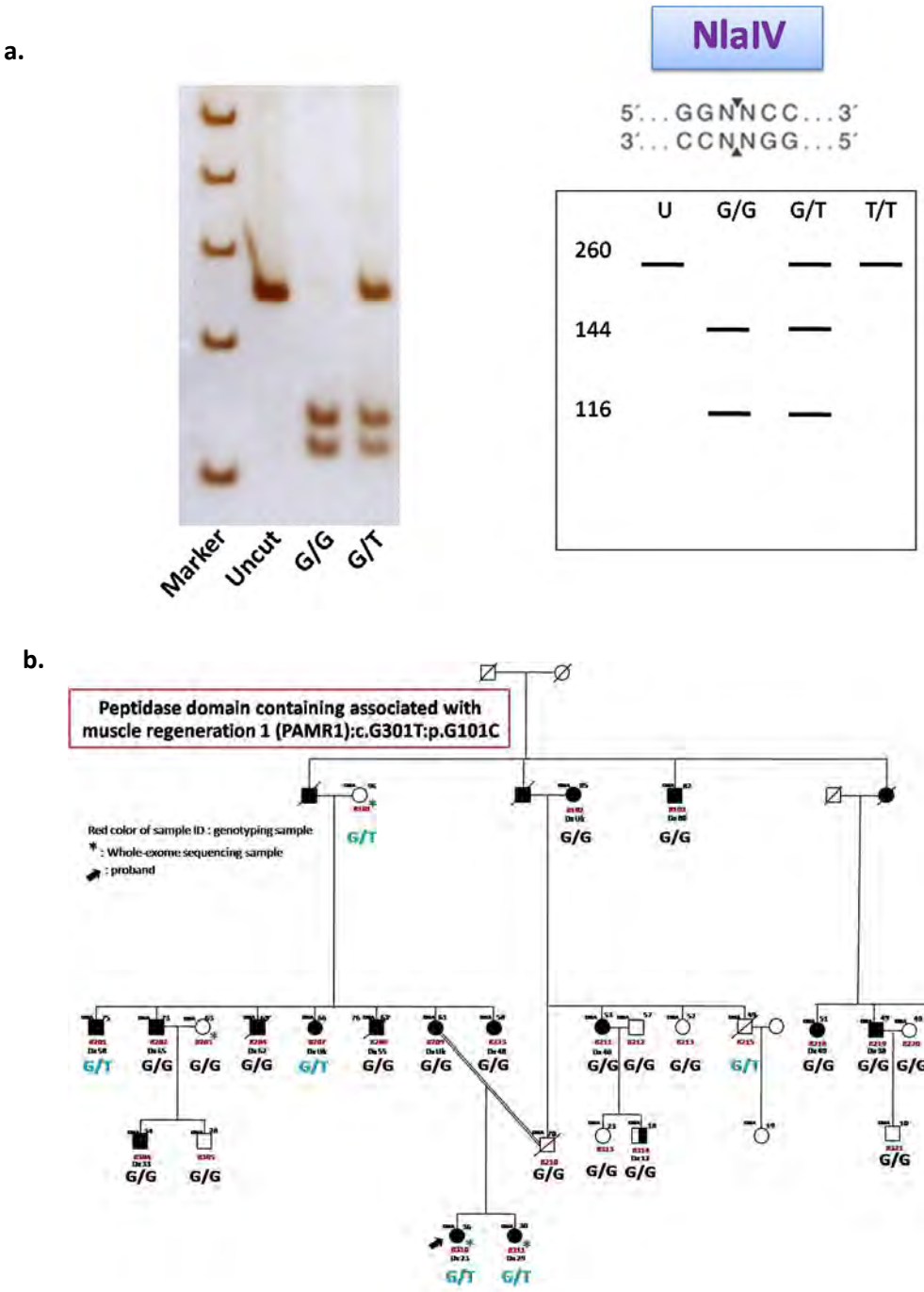


Figure 62 Genotyping result in 27 selected family members. Variant, p. G101C in *PAMR1*, was genotyped by RFLP method. **a**; represent genotyping result by using restriction enzyme NlaIV and pattern of homozygous wild type (G/G) and heterozygous mutant (G/T) was shown. **b**; represent genotyping result in 27 family members who were performed by linkage analysis. G, Glysine; C, Cysteine.

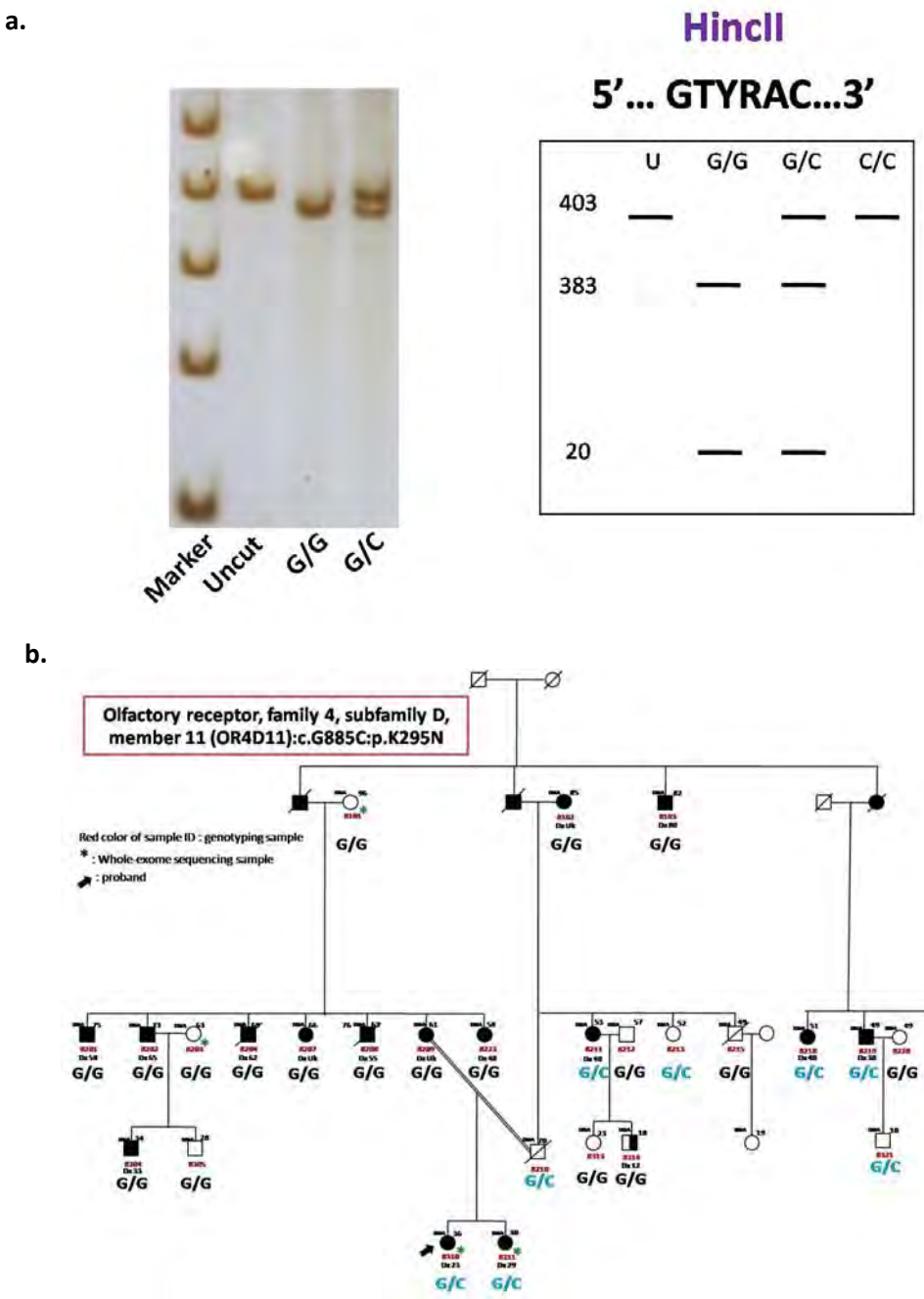


Figure 63 Genotyping result in 27 selected family members. Variant, p. K295N in *OR4D11*, was genotyped by RFLP method. **a**; represent genotyping result by using restriction enzyme HincII and pattern of homozygous wild type (G/G) and heterozygous mutant (G/C) was shown. **b**; represent genotyping result in 27 family members who were performed by linkage analysis. K, Lysine; N, Asparagine.

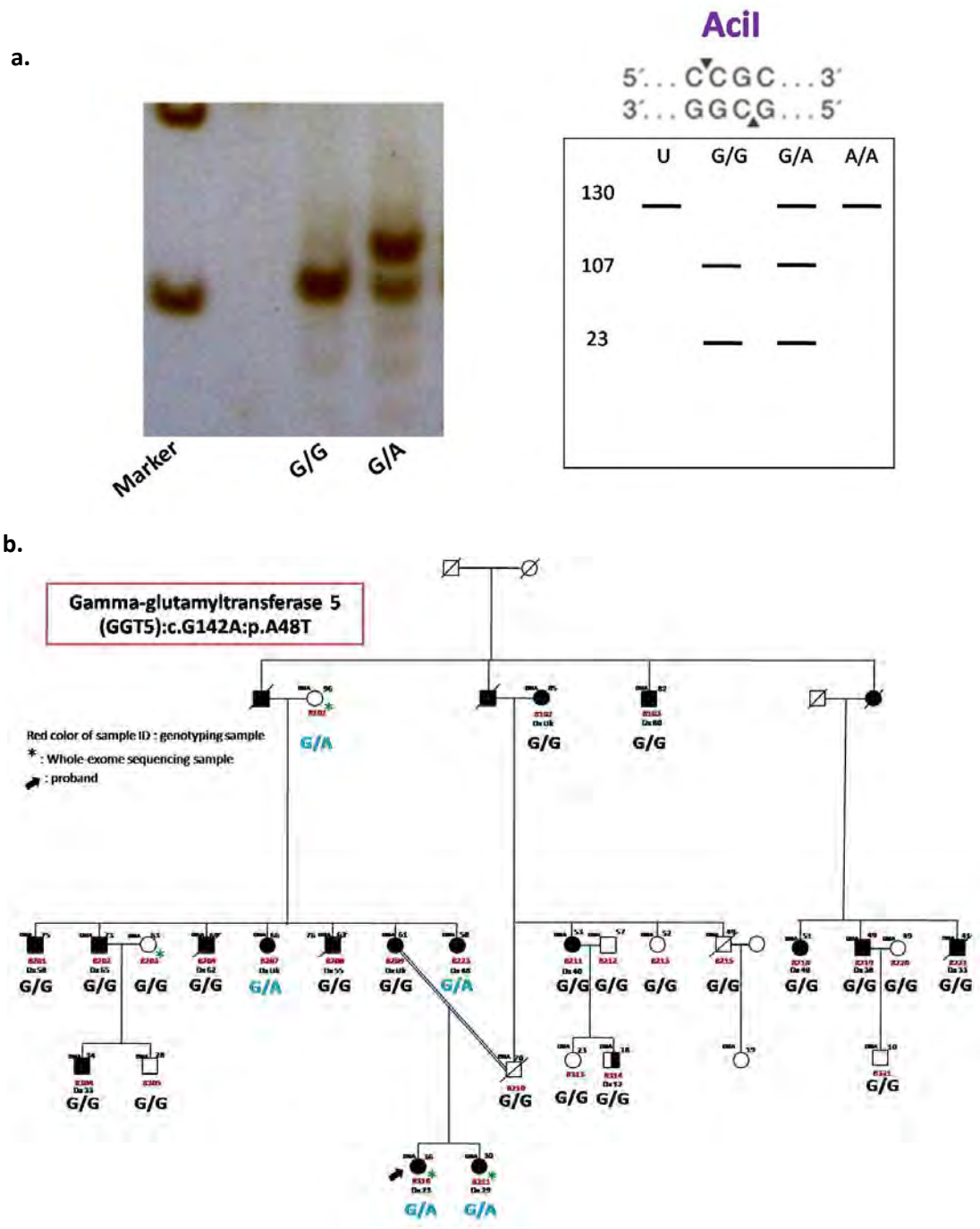


Figure 64 Genotyping result in 27 selected family members. Variant, p. A48T in *GGT5*, was genotyped by RFLP method. **a**; represent genotyping result by using restriction enzyme AciI and pattern of homozygous wild type (G/G) and heterozygous mutant (G/A) was shown. **b**; represent genotyping result in 27 family members who were performed by linkage analysis. A, Alanine; T, Threonine.

10. Genotyping variants in control subject

Two variants (p.M920L in *PTCH1* and p.P1060L in *WNK2* variants) were genotyped in 150 non-diabetic controls and other MODY-X probands by PCR-RFLP method to determine the minor allele frequency (MAF) of candidate genes/variant in controls. Both variants were not identified in 150 controls and other 70 MODY-X probands. Suggest that MAF of both variants are very rare (< 1%) in Thais population.

11. Linkage analysis to confirm significant region by Genehunter program

We performed linkage analysis in the family on chromosome 9 where candidate variants are located by using GeneHunter v2.1r5 by adding variants found in exome sequencing. Multipoint parametric and non-paremetric LOD score were calculated. We found LOD score 2.9062 and 132.5740 of parametric LOD (pLOD) and non-parametric LOD (npLOD) score, respectively (**Fig 65, Fig 66**). High LOD score region is located at the locus of p.M920L in *PTCH1* (at 98 cM) and p.P1060L in *WNK2* (at 96 cM).

				Marker	CHR	cM	pLOD	NPL	npLOD	P val	INFO	Alpha	hLOD
Project:	CHR9	**Inheritance:**	Dominant	1. rs356678	9	98.16	2.9062	24.7088	132.5740	0.0010	0.7734	1.0000	2.9062
Family name:	TOTALS	**Common allele:**	99.90 %	2. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5738	0.0010	0.7734	1.0000	2.9062
Used map:	AFFY 1M deCODE 02122007 v6 (sex–averaged)	**Disease allele:**	0.10 %	3. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5737	0.0010	0.7734	1.0000	2.9062
Marker positions:	40872 ok / 0 ? / 0 outside	**LC1 PCOPY rate:**	0.00 %	4. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5735	0.0010	0.7734	1.0000	2.9062
Allele frequencies:	All individuals from marker file	**LC1 PENET wt/wt:**	100.00 %	5. rs10283758 – rs356660	9	98.16	2.9062	24.7087	132.5732	0.0010	0.7734	1.0000	2.9062
HAP algo / SCFct:	MaxProb / All	**LC1 PENET mt/mt:**	100.00 %										

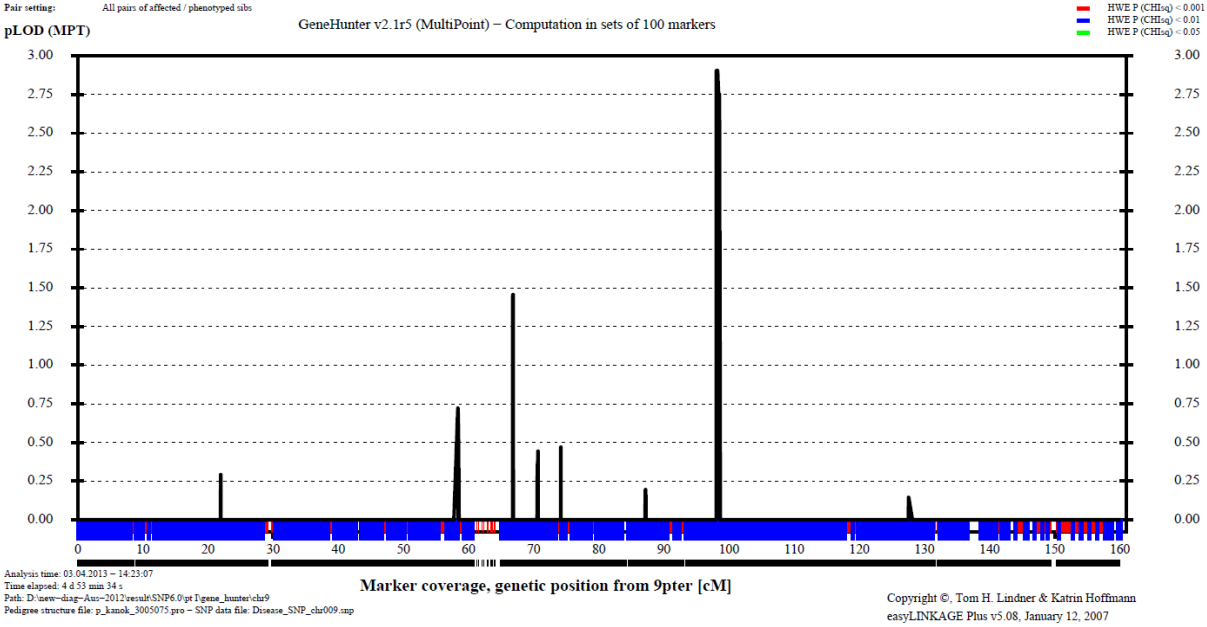


Figure 65 Multipoint parametric LOD scores of chromosome 9

		Marker	CHR	cM	pLOD	NPL	npLOD	P val	INFO	Alpha	hLOD
Project:	CHR9	1. rs356678	9	98.16	2.9062	24.7088	132.5740	0.0010	0.7734	1.0000	2.9062
Family name:	TOTALS	2. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5738	0.0010	0.7734	1.0000	2.9062
Used map:	AFFY 1M deCODE 02122007 v6 (sex–averaged)	3. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5737	0.0010	0.7734	1.0000	2.9062
Marker positions:	40872 ok / 0 ? / 0 outside	4. rs10283758 – rs356660	9	98.16	2.9062	24.7088	132.5735	0.0010	0.7734	1.0000	2.9062
Allele frequencies:	All individuals from marker file	5. rs10283758 – rs356660	9	98.16	2.9062	24.7087	132.5732	0.0010	0.7734	1.0000	2.9062
HAP algo / SCFct:	MaxProb / All										

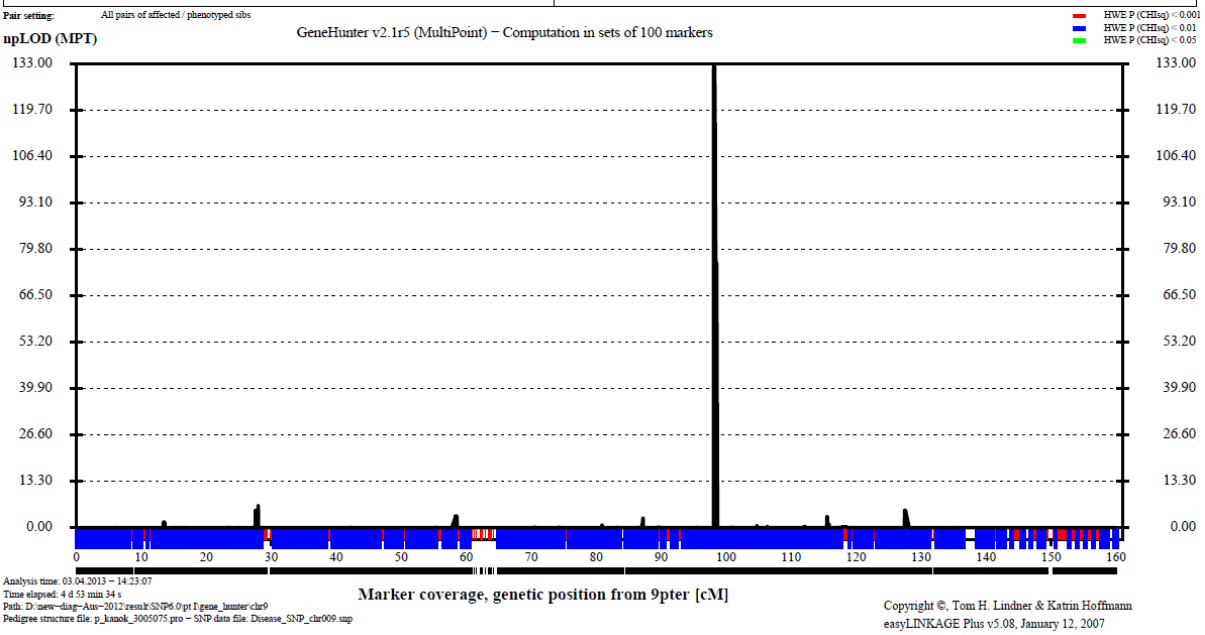


Figure 66 Multipoint non-parametric LOD scores of chromosome 9

Discussion

From several study, this is the first study to identify novel MODY gene by using the combination of linkage analysis and WES. Principle of this tactic is identification linkage regions first and then identified variants from WES located in linkage regions are selected. This method reduces the number of interested variants which are identified from WES and the time consumptions for selecting interested variants from the large data.

Here, we identified 2 candidate variants of p.M920L in *PTCH1* and p.P1060L in *WNK2* in the linkage region located on chromosome 9. Both variants were detected in the same sample ID including in 9 affected family members (8201, 8208, 8209, 8218, 8219, 8224, 8225, 8310, and 8311) and 2 unaffected family members (8318 and 8321). Moreover, they are rare in control and not identified in other MODY-X probands. Both variants are expressed in pancreatic cells and were predicted by PolyPhen2, VarioWatch, SIFT and Mutation Taster to be detrimental to the protein function or structure.

PTCH1, located at 9q22.32, is a receptor for sonic hedgehog (SHH), indian hedgehog (IHH) and desert hedgehog (DHH). Hedgehogs (Hhs) is a intercellular signaling molecule which regulate the mammalian tissue development. Binding of hedgehogs (Hhs) to PTCH1 results in release of the membrane-bound protein smoothened (Smo) and then leads to transcriptional activation of target genes. The function of PTCH1 has been reported that binding of hedgehogs (Hhs) to PTCH1 receptor stimulates insulin gene expression in differentiated β-cells. Therefore, defective Hg signaling in pancreas may induce a pathogenesis of diabetes.

WNK2, located at 9q22.31, is a cytoplasmic serine-threonine kinase. This protein plays an important role in the regulation of electrolyte homeostasis, cell signaling, survival, and proliferation. It acts as an activator and inhibitor of sodium-coupled chloride cotransporters and potassium-

coupled chloride cotransporters respectively. Moreover, it activates SLC12A2, SCNN1A, SCNN1B, SCNN1D and SGK1 and inhibits SLC12A5. However, the function of WNK2 on pancreatic cells has not been identified.

Our study has several limitations. First, this family composed of large number of affected members. Second, father and mother of proband are relative (consanguineous family). Third, diabetes passed through the father family from biparental. Therefore, this family is very complicated for using linkage analysis to identified significant LOD score (LOD>3). To compensate for this limitation, we therefore collected chromosome which showed high LOD score $\geq$ 2.5. Two candidate variants were identified from this study. Although these variants are segregated with only 9 diabetic members and two non-diabetic members in this family but they are novel and rare variants. This study, we did not conduct functional analysis to investigate effect of variants on pancreatic β-cells function. Therefore, we cannot definitely conclude that they are causative gene of MODY. Further functional study is needed to provide knowledge of these proteins on pancreatic β-cells function.

Acknowledgement

This work was supported by Siriraj Graduate Thesis Scholarship, Siriraj Grant for Research and Development, Mahidol University Grant and Thailand Research Fund.

3. **กิจกรรมอื่น ๆที่เกี่ยวข้อง**

- **ผลงานวิจัยที่ตีพิมพ์ในวารสารวิชาการระดับนานาชาติ**

- Pratipanawatr T, Rawdaree P, Chetthakul T, Bunnag P, Ngarmukos C, Benjasuratwong Y, Leelawatana R, Kosachunhanun N, Plengvidhya N, Deerochanawong C, Suwanwalaikorn S, Krittiyawong S, Mongkolsomlit S, Komoltri C; TDR Working Group. Smoking and death in Thai diabetic patients: the Thailand Diabetic Registry cohort. J Med Assoc Thai. 2013 Mar;96(3):280-7.
- Tangjittipokin W, Chongjarean N, Plengvidhya N, Homsanit M, Yenchitsomanus PT. Transcription factor 7-like 2 (TCF7L2) variations associated with earlier age-onset of type 2 diabetes in Thai patients. J Genet. 2012 Aug;91(2):251-5.
- Plengvidhya N, Chanprasert K, Tangjittipokin W, Thongnoppakhun W, Yenchitsomanus PT. Identification of copy number variation of CAPN10 in Thais with type 2 diabetes by multiplex PCR and denaturing high performance liquid chromatography (DHPLC). Gene. 2012 Sep 15;506(2):383-6.
- Kooptiwut S, Plengvidhya N, Chukijrungroat T, Sujjitjoon J, Semprasert N, Furuta H, Yenchitsomanus PT. Defective PAX4 R192H transcriptional repressor activities associated with maturity onset diabetes of the young and early onset-age of type 2 diabetes. J Diabetes Complications. 2012 Jul-Aug;26(4):343-7.
- Plengvidhya N, Tangjittipokin W, Ninareeman Binnima, Nalinee Chongjarean, Chanprasert K, Yenchitsomanus PT. Identified Candidate Genes Causing Maturity-Onset Diabetes of the Young by using Combination of Linkage Analysis and Whole-Exome Sequencing (WES) (In prep)
- Nattachet Plengvidhya, Watip Tangjittipokin, Nalinee Chongjarean, Kanjana Chanprasert, Jatuporn Sujjitjoon, Pa-thai Yenchitsomanus. Analysis of sequence variations of HNF1A in Thai MODY (In prep)

- **ผลงานวิจัยอื่น ๆ**

- นำเสนอผลงานแบบโปสเตอร์ งานประชุมวิชาการศิริราช (2555) คณะแพทยศาสตร์ศิริราชพยาบาล โรงพยาบาลศิริราช เรื่อง Mutation Analysis of Gene Causing Maturity-Onset Diabetes of the Young (MODY) in Thai by Next Generation Sequencing
- นำเสนอผลงานแบบโปสเตอร์ งานประชุม Molecular Medicine Conference (MMC2012) คณะแพทยศาสตร์ศิริราชพยาบาลโรงพยาบาลศิริราช เรื่อง Pax4 Double Variations, R192H and P321H, Impaired Survival of Pancreatic Beta cell Cultured in High Glucose Medium
- นำเสนอผลงานแบบออรัล งานประชุมวิชาการศิริราช (2555) คณะแพทยศาสตร์ศิริราชพยาบาล โรงพยาบาลศิริราช เรื่อง Testosterone Protects High Glucose-Induced Pancreatic cell Apoptosis via Suppression of Oxidative Stress and Endoplasmic Stress (ER) Stress
- นำเสนอผลงานแบบออรัล เรื่อง Impaired Survival of Diabetes-Associated Paired box 4 Mutants under a High -Glucose Condition
- นำเสนอผลงานการวิจัยแบบปากเปล่าในการประชุมนานาชาติ The 12th Symposium on Molecular Diabetology in Asia (The Study Group of Molecular Diabetology in Asia) เรื่อง Genome-wide Linkage Analysis Identified A Novel Chromosomal Region Linked to Diabetes in Thai MODY Family: a preliminary report ณ เมืองปูซาน ประเทศเกาหลี ระหว่างวันที่ 17-20 ตุลาคม 2553
- นำเสนอผลงานการวิจัยเรื่อง Replication of Genome-Wide Association Signals of Type 2 Diabetes in Thai population ในการประชุมวิชาการ 72nd scientific sessions of American diabetes association ประจำปี 2555 ณ Philadelphia, United state วันที่ 8-12 June 2012
- ส่งผลงานการวิจัยในการประชุมนานาชาติ 9th IDF-WPR Congress / 4th AASD Scientific Meeting ณ ประเทศญี่ปุ่น เรื่อง
    - Genetic study of Maturity onset diabetes of the young (MODY) by linkage analysis and exome sequencing
    - A type 2 diabetes associated PAX4 R192H has a defect in transcriptional regulation

- **จำนวนและรายละเอียดการได้รับเชิญไปเป็นวิทยากร**

- บรรยายหัวข้อเรื่อง Who needs MODY screening ในการอบรมวิชาการโรคต่อมไร้ท่อในเวชปฏิบัติ ครั้งที่ 24 สมาคมต่อมไร้ท่อแห่งประเทศไทย ณ ห้องประชุมสยามมกุฎราชกุมาร อาคารเฉลิมพระบารมี เมื่อวันที่ 15 กรกฎาคม 2552
- บรรยายหัวข้อเรื่อง Outpatient management of diabetes :Interactive case discussion ใน update in internal medicine 2009 ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล ณ ห้องประชุมตรีเพ็ชร์ อาคาร 100 ปี ตึกสมเด็จพระศรีนครินทร์ ชั้น 15 โรงพยาบาลศิริราช เมื่อวันที่ 26 พฤศจิกายน 2552
- เป็นผู้ร่วมอภิปรายในการประชุม The 15th congress of the ASEAN Federation of the Endocrine societies (AFES2009) เรื่อง Molecular diabetes and metabolic syndrome in Asia ณ ศูนย์การประชุมแห่งชาติสิริกิติ์ เมื่อวันที่ 29 พฤศจิกายน 2552
- บรรยายหัวข้อเรื่อง What is the other specific type of diabetes? ในการอบรมวิชาการโรคต่อมไร้ท่อในเวชปฏิบัติครั้งที่ 2สมาคมต่อมไร้ท่อแห่งประเทศไทย ณ ห้องประชุมสยามกุฎราชกุมาร อาคารเฉลิมพระบารมี 50 ปี เมื่อวันที่ 25 กรกฎาคม 2555
- นำเสนอผลงาน Research Highlights เรื่อง Novel Adiponectin Variants Identified in Type 2 diabetic patients Reveal Multimerization and Secretion Defects ในการประชุมวิชาการประจำปีครั้งที่ 28 ราชวิทยาลัยอายุรแพทย์ ณ ห้องชลบุรี ระยอง โรงแรมแอมบาสเดอร์ เมื่อวันที่ 28 เมษายน 2555
- บรรยายหัวข้อเรื่องสถานส่งเสริมการวิจัยกับงานวิจัยของคณะแพทยศาสตร์ศิริราชพยาบาล ณ ห้องประชุมอทิตยาทรกิติคุณ ตึกสยามินทร์ ชั้น 7 เมื่อวันที่ 22 กันยายน 2554
- บรรยายหัวข้อเรื่อง Benefit of DPP4 inhibitor in clinical treatment ในการประชุมวิชาการประจำปี 2554 ณ โรงพยาบาลสมุทรปราการ ห้องประชุมสำนักงานอาคารผู้ป่วยนอก ชั้น 4 เมื่อวันที่ 14กรกฎาคม 2554
- เป็นผู้ดำเนินการอภิปรายในการอบรมวิชาการโรคต่อมไร้ท่อในเวชปฏิบัติ ครั้งที่ 26 สมาคมต่อมไร้ท่อแห่งประเทศไทย ณ ห้องถภุสยามกุฎราชกุมาร อาคารเฉลิมพระบารมี 50 ปี เมื่อวันที่ 27 กรกฎาคม 2554

- **การเชื่อมโยงกับนักวิชาการอื่น ๆทั้งในและต่างประเทศ**

  Associate Professor Alessandro Doria, M.D., Ph.D.

  Section on Genetics and Epidemiology

  Joslin Diabetes Centre, Harvard Medical School

  Boston, MA USA

- **การเชื่อมโยงทางวิชาการภายในสภาบันเดียวกัน**
  - มีการประชุมร่วมกันกับผู้ร่วมวิจัยภาควิชาวิทยาภูมิคุ้มกัน,ภาควิชาสรีรวิทยาและหน่วยอณูเวชศาสตร์ คณะแพทยศาสตรศิริราชพยาบาล ทุกสัปดาห์
  - มีการประชุมปรึกษาหารือกับหน่วนต่อมไร้ท่อ ภาควิชาอายุรศาสตร์ เกี่ยวกับวิธีการเก็บตัวอย่างตรวจ
  - ประชุมกับภาควิชาเวชศาสตร์ป้องกัน เกี่ยวกับการรวบรวมอาสาสมัครปกติ

ลงนาม.....

รศ.นพ.ณัฐเชษฐ์ เปล่งวิทยา

(หัวหน้าโครงการวิจัยผู้รับทุน)