A PREDICTIVE MODEL FOR TYPE 2 DIABETES MELLITUS BASED ON GENOMIC AND PHENOTYPIC RISK FACTORS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HÜSAMETTİN GÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
MEDICAL INFORMATICS

JANUARY 2014

A PREDICTIVE MODEL FOR TYPE 2 DIABETES MELLITUS BASED ON GENOMIC AND PHENOTYPIC RISK FACTORS

Submitted by **Hüsamettin Gül** in partial fulfillment of the requirements for the degree of **doctor of Philosophy in the Department of Medical Informatics**,
Middle East Technical University by,

Prof. Dr. Nazife Baykal ____________________
Director, Informatics Institute

Assist. Prof. Dr. Yeşim Aydın Son ____________________
Head of Department, Health Informatics

Assist. Prof. Dr. Yeşim Aydın Son ____________________
Supervisor, Department of Health Informatics, METU

Examining Committee Members

Prof. Dr. Melih BABAOĞLU ____________________
MED, HACETTEPE

Assist. Prof. Dr. Yeşim Aydın Son ____________________
HI, METU

Assist. Prof. Dr. Tuğba TaşkayaTemizel ____________________
IS, METU

Assist. Prof. Dr. Aybar Can Acar ____________________
HI, METU

Prof. Dr. Ümit YAŞAR ____________________
MED, HACETTEPE

Date: 03.01.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Hüsamettin Gül

Signature:

ABSTRACT

A PREDICTIVE MODEL FOR TYPE 2 DIABETES MELLITUS BASED ON GENOMIC AND PHENOTYPIC RISK FACTORS

Gül, Hüsamettin
Ph.D., Department of Medical Informatics
Supervisor: Assist. Prof. Dr. Yeşim Aydın Son

January 2014, 128 pages

Despite the rise in type 2 diabetes (T2D) prevalence worldwide, we do not have a method for early T2D risk prediction. Phenotype variables only contribute to risk prediction near the onset or after the development of T2D. The predictive ability of genetic models has been found to be little or negligible so far. T2D has mostly genetic background but the genetic loci identified so far account for only a small fraction (10%) of the overall heritable risk. In this study, we used data from The Nurses' Health Study and Health Professionals' Follow-up Study cohorts to develop a better and early risk prediction method for T2D by using binary logistic regression. Phenotypic variables yielded 70.7% overall correctness and an area under curve (AUC) of 0.77. With regard to genotype, 798 single nucleotide polymorphisms (SNPs) with P values lower than 1.0E-3, yielded 90.0% correctness and an AUC of 0.965. This is the highest score in literature, even including the scores obtained with phenotypic variables. The additive contributions of phenotype and genotype increased the overall correctness to 92.9%, and AUC to 0.980. Our results showed that the genotype could be used to obtain a higher score, which could enable early risk prediction. These findings present new possibilities for genome-wide association study (GWAS) analysis in terms of discovering missing heritability. Changes in diet and lifestyle due to early risk prediction using genotype could result in a healthier population. These results should be confirmed by follow-up studies.

Key words: Diabetes, genome-wide association study, METU-SNP, binary logistic regression, ROC curve, personalized medicine

ÖZ

TİP 2 DİYABET İÇİN GENOMİK VE FENOTİPİK RİSK FAKTÖRLERİNE DAYALI PREDİKTİF BİR MODEL

Gül, Hüsamettin
Doktora, Tıp Bilişimi Anabilim Dalı
Tez Yöneticisi: Yard. Doç. Dr. Yeşim Aydın Son

Ocak 2014, 128 sayfa

Tip 2 Diyabet yaygınlığı dünya çapında artmasına karşılık, T2D için erken risk tahminine yönelik bir metoda sahip değiliz. Fenotip değişkenleri ancak T2D'nin başlangıcında ya da gelişiminden sonra risk tahminine katkıda bulunmaktadır. Genetik modellerin ise şu ana kadar tahmin kabiliyeti küçük ya da ihmal edilebilir olarak bulunmuştur. T2D çoğunlukla genetik temele sahiptir, fakat günümüze kadar tanımlanan genetik bölgeler genetik mirasın ancak %10'unu açıklamaktadır. Biz bu çalışmada, "Hemşireler Sağlık Çalışması (NHS)" ve "Sağlık Çalışanları İzleme Çalışması (HPFS)" nin verileri ile ikili lojistik regresyon analizi metodunu kullanarak daha iyi ve erken risk tahmini yapabilecek bir metot geliştirmeye çalıştık. Fenotip değişkenleri, %70.7 tahmin değeri ve 0.77 eğri altında kalan alan değeri oluşturdu. Genotip ise, P değeri 1.0E-3'tek küçük 798 adet tek nükleotid polimorfizmi (SNP) kullanarak %90 tahmin doğruluğu ve 0.965 eğri altında kalan alan değeri oluşturdu. Bu değer, fenotip değişkenleri ile bile elde edilen değerden daha yüksek, literatürdeki en yüksek değerdir. Fenotip ve genotip değişkenlerinin birlikte oluşturdukları tahmin değeri ise %92.9 ve eğri altında kalan alan 0.98'dir. Bizim bulgularımız, genotip tabanlı metotların yüksek tahmin değeri elde etmek ve erken risk tahmini için kullanılabileceğini göstermektedir. Bu bulgular, genetik olarak geçen risklerin ortaya çıkarılması suretiyle genom çaplı ilişkilendirme çalışmalarına yeni imkanlar sağlamaktadır. Genotip verileri ile erken tanı sayesinde diyet ve yaşamsal değişiklikler yapılarak daha sağlıklı bir toplum meydana gelebilir. Bu çalışmanın sonuçları takip çalışmaları ile doğrulanmalıdır.

Anahtar Kelimeler: Diyabet, genom çaplı ilişikilendirme çalışması, METU-SNP, ikili lojistik regresyon, ROC eğrisi, bireyselleştirilmiş tedavi

To My Family

ACKNOWLEDGEMENTS

Medical informatics is very important for me to find better solution to the current medical problems. As a pharmacologist, medical informatics is also important for me in the aspects of personalized medicine. Experimental diabetes is one of my earliear research area. I always incline my ears to the news and literature on diabetes. I am interested in most of medical research method but not molecular (DNA, etc) area so far, thanks for Assist. Prof. Dr. Yeşim Aydın Son provided me this opportunity.

It was not easy to perform this study without help of teachers and friends that helped me to achieve it. I am grateful to them for making this possible. First of all, I am thankful to Prof. Dr. Nazife Baykal, my first advisor. She has supported me from the beginning of study. I am very much indebted to my supervisor Assist. Prof. Dr. Yeşim Aydın Son. I achieved this study with her support, interest and guidance. Next, I thank to my friend Y.Yener Tuncel, in METU Bİoinformatics who showed me several excel functions just before his death that were very valuable for me. I always remember him, God rest his soul. I thank my friend Cengizhan Açıkel who supported me about statistical analysis. I also thanks to friends who presented their works during weekly meeting that I learned a lot of thing from them. Next, I need to express my gratitude to my examining committee members Prof. Dr. Melih Babaoglu, Assist. Prof. Dr. Tuğba Taşkaya Temizel, Assist. Prof. Dr. Aybar Can Acar and Prof. Dr. Ümit Yaşar for their valuable comments and contributions. Finally, I am so grateful to my family, my wife and childrens for saving me from their time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Motivation

In this thesis, we have presented an accurate risk prediction method for type 2 diabetes, in which risk SNP panels (genotype) and phenotype are integrated.

1.2 What is Diabetes

Diabetes is characterized with high levels of blood glucose. Glucose is taken from nutrients. Insulin, a hormone made in the pancreas, helps to convert blood glucose into energy and lower blood glucose level [1].

If pancreas does not make enough insulin or because the cells in the muscles, liver, and fat do not use insulin properly, or both, as a result, the amount of glucose in the blood increases while the cells are starved for energy. Persistent high blood glucose level, also called hyperglycemia, damages nerves and blood vessels, which can lead to complications such as heart disease, stroke, kidney disease, blindness, nerve problems, gum infections, and amputation.

There are several types of diabetes. The two main types of diabetes are called type 1 and type 2. A third form of diabetes is called gestational diabetes.

Type 1 diabetes, previously called juvenile diabetes, is generally diagnosed in children, teenagers, and young adults. In this type of diabetes, the pancreas no longer could produce insulin. Insulin-producing beta cells are destroyed or not functional. Patients need insulin treatment. Type I diabetic patients comprise five percent of all diabetic patients.

Type 2 diabetes (T2D) is also called adult-onset diabetes. It is the most common type of diabetes. Nearly 95% of diabetic patients are T2D. T2D could develop at any age, but mainly after 30. T2D usually begins with insulin resistance in peripheral tissues, which muscle, liver, and fat cells do not use insulin properly. As a result, the body needs more insulin to help glucose enter cells for energy production. Initially, the pancreas produces more insulin, but by the time, the insulin secretion by pancreatic beta cells is dysregulated, and eventually it loses the ability to secrete enough insulin in response to high glucose level.

Type 2 diabetes (T2D) is a major public health concern, and its prevalence is increasing at an alarming rate in parallel with rising obesity rates worldwide. The highest incidences of T2D are seen in developing countries where 80% of diabetes deaths occur [2, 3]. There is also recent evidence to show that the age of onset has decreased and cases of T2D in adolescents and children have been reported [4]. Although this rise in diabetes prevalence can be mostly attributed to changes in diet and lifestyle, there is strong evidence of a genetic basis for T2D [5]. For example, a study in Danish twins estimated the T2D concordance rate in dizygotic twins as 43% compared with 63% in monozygotic twins [6, 7], and the relative risk of T2D for a sibling is approximately four- to six-fold higher than that of the general population [8].

It is estimated that 371 million people are already affected with T2D and projected to reach 552 million by 2030 [9]. Its increasing prevalence is a serious concern in many countries. T2D affects approximately 21 million individuals in the U.S. or almost 10% of the U.S. adult population. Because diabetes is determined by both genetic and environmental factors, a better understanding of the etiology of diabetes requires a careful investigation of gene-environment interactions. Few studies have been conducted to analyze these interactions so far. One of the most known study is GENEVA Genes and Environment Initiatives in Type 2 Diabetes which is performed among nurses and health professionals [10].

1.3 Genetics of Diabetes

The success of the completion of human genome (sequencing) project, followed by the start of GWAS held out the hope that personalized medicine would be realized within the near future. Prior to the GWAS studies, the importance of genetic factors in the etiology of T2D had been well established through family and twin studies [5, 11]. The primary methods to identify susceptibility loci for diseases or phenotypic traits were linkage analysis and candidate gene association studies. Linkage analysis is useful for identifying familial genetic variants that have large effects and was successfully used to discover several causal mutations for the monogenic forms of diabetes mellitus, such as maturity-onset diabetes of the young (MODY) [8].

A significant breakthrough in understanding the genetic basis of complex traits of T2D was facilitated by GWAS. GWAS is a powerful method to detect genetic variations that predispose to a disease. In GWAS, the entire genomes of individuals with and without the disorder of interest (i.e., cases and controls) are screened for a large number of common SNPs. These studies have been facilitated by several recent developments including completion of the Human Genome Project and the International HapMap project. Several million SNPs were discovered and confirmed by the International HapMap project and have been deposited in a public database [9]. The underlying pattern of the inheritance of genetic variation was defined and as quantified by LD. Two SNPs with strong LD are thought to be coinherited more frequently than SNPs with weak LD. Using this correlation structure, association analyses can be made in a more efficient and cost-effective manner by using a smaller subset of SNPs or “tag” SNPs to capture most of the remaining common genetic variations.

Type-2 diabetes is a complex disease characterized by a number of environmental and genetic factors that contribute at varying degrees to the final phenotype. Genetics and environmental factors interact with each other. Deciphering the genetic background of T2D could increase our knowledge on the pathogenesis and identifying new targets for drug development to successfully personalizing clinical disease prediction, prognosis and treatment. Several genes have been described from genome-wide association studies (GWAS) on T2D so far, to identify the gene targets that have been assessed to-date stem from the rapid growth of literature on this issue. A considerable number of the proposed genes seem to be related to beta-cell development and function, but there are several genes identified as "diabetes-genes" whose underlying pathway linked to diabetes remains poorly understood. Despite the increasing numbers of identified genetic markers, a large proportion of the observed type-2 diabetes heritability remains unexplained.

1.4 What is SNP?

The human genome has an array of nearly 3 billion letters from the set of [12] representing nucleotides Adenine, Cytosine, Guanine and Thymine. The nucleotide sequence does not differ across the populations in more than 99% of the positions of the whole genome. However, individuals possess genetic variations in about 1% of their genomic sequences. Among those variations, the most frequently observed are changes at single nucleotide level, called Single Nucleotide Polymorphisms (SNPs), when occurred in over 1% of a given population. SNP is one of the important genetic investigation area. SNP (snip) is a DNA sequence variation accrues when a single nucleotide-A, T, C or G- in the genome differs between members of a biological species or paired chromosomes in an individual. SNPs comprise >90% of all of the polymorphisms.

AAGC**C**TA There two alleles: C and T
AAGC**T**TA

SNPs might be important for humans susceptibility to diseases and respond to pathogens, chemicals, drugs, vaccines. SNPs might be the key enablers in realizing the concept of personalized medicine.

Recent developments in genotyping technologies, public access to whole genome and other genetic information and the start of the International HapMap Project have facilitated the implementation of SNP based GWAS [12, 13].

1.5 Literature Review: The Need for Early Risk Prediction using Genotype Based Method for Type 2 Diabetes

The development of high-throughput genotyping technologies along with statistical and computational software has allowed remarkable progress over the past decade in the "genome-wide" search for genetic associations. GWAS have dramatically increased the number of known T2D susceptibility loci. The analysis of related quantitative traits has uncovered new loci associated with T2D and potential pathways for therapeutic intervention. Since the first GWAS for T2D identified novel susceptibility loci in 2007, approximately 40 T2D susceptibility loci have been identified so far, and most of them were through GWAS [14].

Prior to the accumulation of GWAS data, a genetic predisposition to insulin resistance had been considered to play a dominant role in development of T2D, especially in populations of European origin. However the results obtained from early GWAS, emphasize the crucial role of the pancreatic beta cells in the onset of T2D, and a genetic predisposition for reduced beta-cell function might be the major reason for susceptibility to T2D.

In fact, for most of the T2D susceptibility loci identified so far, the causal variants and molecular mechanisms for diabetes risk were unknown. Disease-associated SNPs are usually annotated by the gene in closest proximity; however, the protein encoded by that gene may not have a causative role in the development of T2D in humans.

The SLC30A8 encodes ZnT-8, which transports zinc from the cytoplasm into secretory vesicles for insulin storage and secretion [15]. A therapeutic agent that enhances the intracellular function of this transporter could theoretically increase insulin secretion and lower blood glucose levels. In addition, other T2D susceptibility variants confirmed by GWAS include variants within the genes PPARG and KCNJ11 that encode targets of the established oral hypoglycemic agents, thiazolidinediones and sulphonylureas, respectively [16, 17]. Therefore, elucidating the mechanisms by which each susceptibility locus contributes to T2D will improve our understanding of the pathophysiology of T2D and will provide new and useful information for the development of new drugs for the treatment and/or prevention of T2D.

Development of genotype-based prediction will help us for early prediction, identification, and prevention of T2D. Translation of new findings from GWAS to the clinic is the most attractive aspects of genome research. One of the potential clinical applications is the development of genetically based personalized susceptibility profiles via prediction, early identification, and prevention of T2D or its complications.

The development of T2D is caused by a combination of lifestyle and genetic factors [5, 18]. Some of the risk factors such as diet and obesity are under personal control, but genetic factors are not [19]. Although the rise in T2D prevalence can be mostly attributed to changes in diet and lifestyle, there is strong evidence of a genetic basis for T2D [5]. However, genetic risk factors have been found to have less predictive value when compared to phenotype variables such as body mass index (BMI), familial diabetes history, blood pressure and cholesterol [20, 21]. Furthermore, additive contribution of genetic studies using single nucleotide polymorphism (SNP) to phenotype variables was found almost negligible in several studies [11, 20-26]. Numerous genetic and non-genetic risk factors interact in the causation of T2D, the predictive ability of genetic models will likely remain modest.

Approximately T2D susceptibly 40 variants have been identified so far, many of which were discovered through GWAS [25]. However, the genetic loci identified till now account for only a small fraction (approximately 10%) of the overall heritable risk for T2D [26]. There is likely to be many additional signals with minimal effect and low frequency that would be discovered through ongoing iterations of the genome-wide approach. Uncovering the missing heritability is essential to the progress of T2D genetic studies and to the translation of genetic information into clinical practice.

At present, the clinical use of genetic testing for T2D prediction in adults is not recommended due to the low predictive power. Phenotype based risk factors have higher predictive ability, in which AUC is between 0.70-0.90 but for patients over 45 when the reversibility of the factors might not be possible. However, we need a model to predict risk score for T2D earlier. Pre-diabetic individuals usually remain undiagnosed and untreated. Identifying new methods using genotype for screening and prediction of risk factors are very important. If we predict risk factors earlier, it may help patients by changing lifestyle modification about preventable risk factors such as obesity [27].

Genome-wide association studies (GWAS) has been widely used to investigate the role of genotypic profiles in the molecular etiology of diseases. Although many studies has been conducted to uncover heritability of T2D, only small proportion of genetic heritability was explained by the variants identified. Thorough GWAS, 44 susceptibility loci were identified as genome-wide significant associations with T2D so far [28]. While the current T2D risk variants explained up to 5–10% of the genetic basis of T2D, much of the genetic basis still remains unexplained [29].

In most studies the logistic regression is used for the analysis of genetic variables. However, the maximum number of SNPs analyzed only goes up to 42 SNP and C-statistics (area under curve, AUC) for genotype was under than 0.60 [11, 20-26]. When we were performing GWAS analysis of NHS and HPFS data, we realized that sensitivity, specificity, and C-statistics increased when the number of SNPs in the analysis also increased. We took the advantage of the GWAS data in the study to expand our research to hundreds of SNPs, and examine 798 associated SNP, with P values lower than 1.0E-3. Including high number SNPs resulted with the the highest prediction risk scores and AUC for T2D reported so far in the literature. Predictive performance of SNP profiles was even higher than the predictive models based on the phenotype. Overall we have presented the importance of genome wide analysis of genotypes for the prediction of T2D which were previously disregarded when small set of SNPs investigated in the studies.

1.6 Prioritization

Although the current rise in T2D prevalence is driven mainly by changes in life-style, complex genetic determinants are widely considered to contribute to the inherent susceptibility of this disease. The pathogenesis of T2D is heterogeneous, suggesting that the contribution from individual genetic factors is modest. Linkage analysis and the candidate gene approach were the primary methods to link genotype and phenotype before the development of genome wide association studies (GWAS). Although these techniques can detect rare genetic variants that strongly influence disease susceptibility, they are not suitable to identify variants that have a smaller effect on disease susceptibility. Therefore, the discovery of novel T2D susceptible loci has been challenging, and a more powerful strategy was needed to overcome this difficulty. Prioritization of the SNPs that is most relevant with the disease emerged as one of the promising methods to overcome these difficulties.

There are various studies investigating the relations between SNP and disease, including diabetes [27, 30-33]. Some of them use not only p value of SNPs but also uses prioritization algorithms to identify statistically and biologicaly relevant SNPs with diabetes. Previously, a SNP prioritization tool was developed by METU Informatics group called METU-SNP for this purpose. METU-SNP has some favorable features over the others [34].

The METU-SNP software [34], performs analytical hierarchical process (AHP) for SNP prioritization and calculates a combined p-value for the genes. In GWAS analysis, the determination of the statistical significance of SNPs by calculating p-values of association is performed as a first step. Depending on user's choice, three different methods can be used to calculate p-values: (1) uncorrected, (2) Bonferroni and (3) False Discovery Rate. P value threshold could be set by user and depending on the threshold. SNPs are labeled as significant by METU-SNP software.

The second step of GWAS is performed by calculating the combined p-values to reveal statistically significant (enriched) genes and pathways as described previously [34, 35]. Fisher's combination test is applied to combine p-values of all SNPs within a gene, where the statistics for combining K SNPs is given by

$ZF = -2\sum_{i=1}^{K} lnPi$ which follows $\chi 2K2$ distribution.

In order to determine the overrepresentation of significantly associated genes among all genes in a pathway, the hypergeometric test (Fisher's exact test) has been used. Assuming that total number of genes is N, the number of genes that are significantly associated with the disease is S and the number of genes in the pathway is m; p-value of observing k-significant genes in the pathway is calculated by:

It is important to note that when describing an association, it has become standard practice to refer to the identified signal by the closest gene(s) name(s); but this does not necessarily mean that the gene itself is causal.

1.7 Binary Logistic Regression Models

A major strength of regression is that it easily provides an opportunity to include interactions. Among the other advantages of regression analyses are explicit parametric models, stable algorithms for parameter estimation, easy incorporation of covariates such as age, sex, and ethnic origin and wide availability of reliable and well-documented software. Some of the disadvantages failure to deliver spare solutions, and the hierarchical nature of the model selection requiring detection of main effects before detecting interaction.

Binomial (or binary) logistic regression is a form of regression, which is used when the dependent is a dichotomy and the independents are of any type. Logistic regression uses binomial probability theory, does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, and in general has no stringent requirements, and a linear combination of the predictors is linked to the mean of a binary outcome variable by the logit function.

The primary distinction between a logistic regression model and a linear regression model is that the outcome variable in logistic regression is binary or dichotomous. The logistic regression model is simply a non-linear transformation of the linear regression. The goal of logistic regression analysis is the same as that of any model building techniques used in statistics: to find the best fitting and most parsimonious, yet biologically reasonable model to describe the relationship between a response variable and a set of independent variables. In logistic regression, the method of maximum likelihood estimation (MLE) is used to estimate the unknown parameters, which maximizes the probability of obtaining the observed data.

Logistic regression involves fitting an equation of the to the data using the following formulae for binary data,

$$logit\{ Y = 1 \mid x \} = \ln \left(\frac{\mathrm{P(Y=1|}x}{1-\mathrm{P(Y=1|}x} \right) \qquad \text{Equation 1}$$

$$logit\ (P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k \qquad \text{Equation 2}$$

Classification table tells us how many of the cases where the observed values of the dependent variable were 1 or 0 respectively have been correctly predicted. In a perfect model, all cases will be on the diagonal and the overall percent correct will be 100%.

Logistic regression has many analogies to linear regression: logit coefficients correspond to b coefficients in the logistic regression equation, the standardized logit coefficients correspond to beta weights, and the Wald statistic, a pseudo R2 statistic, is available to summarize the strength of the relationship. The success of the logistic regression can be assessed by looking at the classification table, showing correct and incorrect classifications of the dependent. In addition, goodness-of-fit tests such as model chi- square are available as indicators of model appropriateness, as is the Wald statistic to test the significance of individual independent variables. The EXP(B) value indicates the increase in odds from a one unit increase in the selected variable.

$$P = \frac{exp^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k)}}{1 + exp^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k)}} \qquad \text{Equation 3}$$

P, the probability that a case is in a particular category, exp, the base of natural logarithms (~2.72), β, the constant of the equation, β_0, the coefficient of the predictor variables.

There is an ample spectrum of different statistical approaches for detecting interaction; logistic regression is probably the most popular one among genetic epidemiologists and geneticists. As logistic regression measures the relationship between a categorical dependent variable and one or more independent variables by using probability, it is used extensively in numerous disciplines, including the medical and social science fields. Logistic regression is generally used to predict whether a patient has a given disease (e.g. diabetes), based on observed characteristics of the patient (age, gender, body mass index, results of various blood tests, etc.).

LR can play an important role as statistical tools in large-scale genetic association studies where unknown interactions exist among true risk-associated SNPs with marginal effects and in the presence of a significant number of noise SNPs. The primary goal of using logistic regression in this study was to identify SNPs that may increase or decrease susceptibility to disease. This was achieved by quantifying how much each SNP contributes to the predictive accuracy of these methods by measuring its predictive importance. Finding that a SNP helps differentiate between cases and controls is an indication that the SNP either contributes to the phenotype or is in linkage disequilibrium with SNPs contributing to the phenotype.

In addition, we also realized that BLR has been used extensively in genotype studies but these studies used only several SNPs (i.e. 40 SNPs). However, our SNPs selected from 934,940 SNPs and represented nearly all genomes as explained in the following sections. Furthermore, our genotypic results have the highest score to predict the risk factor of diabetes in the literature. Therefore, we thought that BLR was effective methods for this purpose. For finite number of SNP, it is easy to perform BLR, but we used as high as 798 SNPs which not tried before.

CHAPTER 2

MATERIALS AND METHODS

2.1 Genotyping and Phenotype Data

Data were taken from the study which is a part of the GENEVA, funded by the trans-NIH Genes, Environment, and Health Initiative (GEI). The overarching goal of this initiative was to identify novel genetic factors that contribute to T2D through large-scale genome-wide association studies of well-characterized cohorts of nurses and health professionals. Genotyping was performed at the Broad Institute of MIT and Harvard, a GENEVA genotyping center. Data cleaning and harmonization were done at the GEI-funded GENEVA Coordinating Center at the University of Washington [10].

The Nurses' Health Study (NHS) and Health Professionals' Follow-up Study (HPFS) are well-characterized cohorts of nurses and health professionals, which conducted to identify novel genetic factors that contribute to T2D through large-scale genome-wide association studies and to investigate the role of environmental exposures on the development T2D. NHS and HPFS cohorts are part of the Gene Environment Association Studies initiative (GENEVA, http://www.genevastudy.org). The NHS was established in 1976 and the HPFS study was started in 1986. Participants of NHS and HPFS study completed a mailed questionnaire on their medical history and lifestyle. Blood samples were collected in 1989-1990 for NHS and 1993-1995 for HPFS. Genotyping was completed in December 2008 for NHS and in March 2009 for HPFS. The lifestyle factors, including smoking, menopausal status and postmenopausal hormone therapy, and body weight, have been updated by validated questionnaires every 2 years.

We have only used white, type 2 diabetic patients' data in our analysis. We have excluded the cases with other type of diabetes and races. The summary of the case and controls were given in the Table 2.1.

Participants meeting the following criteria were excluded from the study: 1) those with other types of diabetes (65 NHS, 68 HPFS); 2) those belonging to races other than white (61 NHS, 100 HPFS); 3) HapMap controls (45 NHS, 29 HPFS), and 4) first-degree relatives (15 NHS, 14 HPFS). The final sample included 3,248 (1,769 controls and 1,479 cases) for NHS and 2,391 (1,277 controls and 1,114 cases) for HPFS. The current analysis includes single nucleotide polymorphisms (SNPs) mapped to chromosomes 1 through 23, as annotated based on the Affymetrix Genome-wide Human SNP Array 6.0 (GeneChip 6.0).

The Nurses’ Health Study (NHS) cohort was established in 1976 when 121,700 female registered nurses aged 30 to 55 years and residing in 11 U.S. states completed a mailed questionnaire on their medical history and lifestyle characteristics. The women have since received follow-up questionnaires biennially to update information on exposures and newly diagnosed illnesses. Starting in 1980, on a 2-4 year cycle, dietary information has been updated using validated semi-quantitative food frequency questionnaires. Between 1989 and 1990, a blood sample was requested from all active participants in NHS and collected from 32,826 women. The cases and controls for the NHS Type 2 Diabetes (T2D) project were selected among those with a blood sample using a "nested" case-control study design. Cases of T2D were identified by self-report on biennial follow-up questionnaires and confirmed by a medical record-validated supplementary questionnaire. Controls were defined as those free of diabetes at the time of diagnosis of the case. The case-control sampling was carried out for prevalent (diagnosed before blood collection) and incident diabetes cases (diagnosed after blood collection and before June 1, 2004). DNA was extracted from white blood cells using the Qiagen "QIAamp" blood protocol and all samples were processed in the same laboratory. The genotyping was done at the Broad Center for Genotyping and Analysis (CGA) using the Affymetrix Genome-Wide Human 6.0 array.

The Health Professionals Follow-up Study (HPFS) was initiated in 1986 when 51,529 male health professionals between 40 and 75 years of age years and residing in 50 U.S. states completed a food frequency questionnaire (FFQ) and a medical history questionnaire. The participants have been followed with repeated questionnaires on lifestyle and health every 2 years and FFQs every 4 years. Between 1993 and 1994, a blood sample was requested from all active participants in the HPFS and collected from 18,225 men. Cases of T2D were identified by self-report on biennial follow-up questionnaires and confirmed by a medical record-validated supplementary questionnaire. Controls were defined as those free of diabetes at the time of diagnosis of the case. The case-control sampling was carried out for prevalent diabetes cases (diagnosed before blood collection) and incident cases (diagnosed after blood collection and before June 1, 2004). Subsequently, cases were divided into two categories, T2D and diabetes of uncertain type [10].

Table 2.1 Characteristics of the case and controls.

	NHS (female)	HPFS (male)	Total
Control	1769	1277	3046
Case (T2D)	1479	1114	2593
Other type of diabetes *	65	68	133
Other than white race *	61	100	161
HapMap control *	45	29	74
First degree relatives *	15	14	29
Total	3434	2603	6036

* Excluded from the study.

2.2 Phenotypic Dataset Description

We used phenotypic variables obtained from dbGAP. This dataset represents variables that were selected from the Nurses' Health Study (NHS, all female) and the Health Professionals Follow-up Study (HPFS - male) to determine if dietary and life-style habits effect the development of Type 2 Diabetes. The variables describe medical history (3 variables), intake of e.g. alcohol (1 variable) and nutrients (6 variables), smoking (1 variable), exercise habits (1 variable) and body measurements (3 variables), menopause status (1 variable), and general socio-demographic status (5 variables).

2.2.1 Study Inclusion/Exclusion Criteria

The study was performed using Nurses' Health Study or Health Professionals Follow-up Study cohort subjects.

Cases: Type 2 diabetes mellitus

Controls: no diabetes mellitus

We excluded other type diabetes (i.e., type I diabetes, gestational diabetes), person other than white race, HapMap control and first-degree relatives from the raw data.

2.2.2 Molecular Data

Type: Whole Genome Genotyping

Vendor/Platform: AFFYMETRIX AFFY_6.0

Number of Oligos/SNPs: 934940

SNP Batch Id: 52074

SNPs that met any of the following criteria are excluded from the analysis: 1) minor allele frequencies (MAF) <0.05; 2) call rate <95%; 3) P for Hardy-Weinburg equilibrium (HWE) <0.001; and 4) missing rates 0.1.

Before frequency and genotyping pruning, there are 909,622 SNPs, 5 of 6041 individuals removed for low genotyping (MIND >0.1), 308,275 heterozygous haploid genotypes set to missing, 45,179 markers to be excluded based on HWE test (p <= 0.001), total genotyping rate in remaining individuals is 0.96. 50,080 SNPs failed missingness test (GENO >0.1), 229,277 SNPs failed frequency test (MAF <0.05), after frequency and genotyping pruning, there are 642,576 SNPs; after filtering, 2593 cases, 3046 controls and 397 missing person.

2.3 Analysis Steps

We used METU-SNP analysis software to calculate AHP score. It has preprocessing, association, prioritization, and selection tools. Since we have binary data (.bim, .bed and .fam instead of .ped and .map), we started from association step. However, cases and controls was not defined in the existing .fam file, we described them by ourselves using phenotype files. The processing steps were described below.

- **Merging** data files (NHS and HPFS data);
  (command: plink --bfile NHS --bmerge HPFS.bed HPFS.bim HPFS.fam --make-bed --out diab)
- **Filtering** files for QC using plink software;

  (command: plink --bfile filename **--geno 0.1 --hwe 0.001 --mind 0.1 --maf 0.05** --make-bed --out newfilename)
- Creating **.fam** file according to case and controls (obtained from phenotype data),
- *Plink* analysis was performed and p-values obtained.
- Creating **.adjusted** file for analysis using plink software;

  (command: plink --bfile filename --assoc --adjust --out association)

- **Converting** Affymetrix data format to reference snp (rsid) data format before prioritization step (i.e. SNP_A-8319564 to rs11121467)
- **Prioritization** steps by METU-SNP software and obtaining AHP score.
- Gene databases were constructed and SNPs, which have significant p-value, were mapped to genes. "webgestalt" website (http://bioinfo.vanderbilt.edu/webgestalt)
- **Mapping** SNPs and genes according to chromosome, location, odd ratio, minor and major bases of SNPs, MAF, p-value
- **Interpretation** of results with literature
- Phenotype and genotype data were combined
- Binary logistic regression analysis was performed by SPSS ver 15.0.
- Genotype features were analyzed with binary logistic regression
- 886 SNPs with p value lower than 1.0E-3 were extracted from raw data and analyzed with binary logistic regression (after elimination of SNPs that had >50missing allele),
- Phenotype and genotype features were analyzed with binary logistic regression,
- ROC curve was constructed for phenotype and genotype,

2.4 SNP Selection

We have selected 798 SNPs amongst 934,940 SNPs. SNP selection method is presented in Figure 2.1.

909,622 --> 642,576 SNPs (preprocessing)

↓

886 SNPs (p<1.0E-3)

After elimination of SNPs ↓ which have high missing allele

798 SNPs

Figure 2.1 SNP Selection Method for BLR Analysis

2.5 Extraction of SNP Data

SNPs was extracted with the following command from raw data;

```
>plink --bfile data --snps snp1, snp2, ... --recode --out data1
```

It should be noted that SNPs should be in chromosomal and location order.

2.6 Software

2.6.1 PLINK

PLINK version 1.07 was used to analyze genome-wide data (http://pngu.mgh.harvard.edu/~purcell/plink). There were methodological advances, including statistical tools to analyze SNP data such as PLINK that were made freely available, facilitating the design, analysis, and interpretation of the large amounts of data being produced [36]. When performing such large numbers of association tests, the importance of stringent significance thresholds was recognized, i.e. minor allele frequency, missingness rate etc. that will be described below. We used PLINK to obtain the significance level (P value), frequency, and odds ratio of SNPs.

2.6.2 R Software

R is a free software environment for statistical computing and graphics (http://www.r-project.org). The R language is widely used among statisticians and data miners for developing statistical software and data analysis. The capabilities of R are extended through user-created packages, which allow specialized statistical techniques, graphical devices, import/export capabilities, reporting tools, etc. We used R programming to plot the QQ graphics, Manhattan plot, and graphics of distribution densities.

2.6.3 AMELIA

We used Amelia for data imputation of missing allele [37]. 886 SNPs was selected for analysis which had lower p values than 0.001 (1.0E-3). 88 of 886 SNPs were eliminated since their missing allele number was greater than 50, after elimination these SNPs 798 SNPs remained as summarized in Figure 2.1.

The SNP rs10739592 with the lowest p value (2.08E-14) and one of the highest OR (1.34), and MAF (0.49) is not excluded from the study even though it had missing allele number of 99/5639, which is greater than 50 (patients). Therefore, we filled the missing value of rs10739592 by Amelia. The results of imputation is validated by comparing before and after p-values of SNPs and observing the distribution density of the original data set and the imputed data set. We have compared the p-values before and after the imputation to observed the influence of filling the p-value, which were 2.08E-14 before and 3.13E-14 respectively, Thus, filling the missing allele seems had no major effect on the p-value. The details of the imputation with Amelia is given in Appendix A. Imputed allele rate was 0.14%.

2.6.4 SPSS

SPSS is used for both conventional statistical analysis (i.e. Student t test where appropriate) and the binary logistic regression analysis.

2.6.4.1 Binary Logistic Regression Functions

Logistic regression is widely used to model independent binary response data in medical and epidemiologic studies. Many methods have been proposed in regression models for variable selection. Classical methods for variable selection include forward selection, backward elimination, and stepwise regression.

The binary logistic regression (BLR) is used for variable reduction and also presented to be an efficient method to identify the risk SNPs associated with T2D. The relation between genotype and/or phenotype variables and T2D are evaluated.

The SPSS version 15.0 software for BLR is used. We performed binary logistic regression (BLR) using NHS and HPFS genotype and phenotype data via SPSS to test associations of the genotype and phenotype risk scores with diabetes. We coded genotypes for common allele homozygote, heterozygote, and rare allele homozygote separately for analysis. We evaluated model discrimination using C-statistics (the areas under receiver operating characteristic curves, ROC-AUCs) which were calculated for the predicted risk of the logistic regression model. Significance of the difference between the areas under two independent ROC curves was calculated according to Hanley and McNeil (1982) using http://vassarstats.net/ website [38].

2.6.4.1.1 The Wald statistic

The Wald statistic and associated probabilities provide an index of the significance of each predictor in the equation. The Wald statistic has a chi-square distribution. The simplest way to assess Wald is to take the significance values and if less than .05 reject the null hypothesis as the variable does make a significant contribution.

Wald χ2 statistics are used to test the significance of individual coefficients in the model and are calculated as follows:

$$Walds\ Statistics = \left[\frac{Coefficient}{SE\ of\ Coefficient}\right]^2 \qquad \text{Equation 4}$$

Each Wald statistic is compared with a χ2 distribution with 1 degree of freedom. Wald statistics are easy to calculate.

We found that for four phenotype variables are the most important and their coefficients are given in Table 2.3.

Table 2.2 Example of constant, Wald, and P values in Binary Logistic Regression Analysis

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	FAMDB	1.132	.064	308.641	1	.000	3.102
	HBP	.862	.066	171.634	1	.000	2.368
	CHOL	.556	.071	60.395	1	.000	1.743
	BMI	1.351	.061	487.412	1	.000	3.860
	Constant	-1.579	.054	853.081	1	.000	.206

a Variable(s) entered on step 1: FAMDB, HBP, CHOL, BMI.

As noted above, high Wald value is proportional to the significance level variables. In this example, we calculate probability as;

$$P = \frac{exp^{(-1,579+famdb*1,132+hbp*0,862+chol*0,556+bmi*1,351)}}{1+exp^{(-1,579+famdb*1,132+hbp*0,862+chol*0,556+bmi*1,351)}} \qquad \text{Equation 5}$$

- If a person has FAMDB (exist; 1), HBP (exist; 1), CHOL (exist; 1), and BMI (exist; 1) so the risk probability of this person is 0.911
- If a person has FAMDB (not exist; 0), HBP (not exist; 0), CHOL (not exist; 0), and BMI (not exist; 0) so the risk probability of this person is 0.171
- If a person has FAMDB (exist; 1), HBP (not exist; 0), CHOL (not exist; 0), and BMI (not exist; 0) so the risk probability of this person is 0.390
- If a person has FAMDB (not exist; 0), HBP (not exist; 0), CHOL (not exist; 0), and BMI (exist; 1) so the risk probability of this person is 0.443 and so on.

Wald Statistics for FAMDB $\left[\frac{1,132}{0,064}\right]^2$ =308,641 etc.

2.6.4.1.2 Method Types in BLR

Method selection allows us to specify how independent variables are entered into the analysis. We can construct a variety of regression models from the same set of variables using different methods. However, methods other than ENTER were found to be time consuming. For example, while 5639 rows and 798 columns data took ~30 min to analyze using ENTER method, where as it was 10 days for Forward Likelihood Ratio method. In addition, prediction score was higher by using more SNPs with ENTER method. However, if we want to reduce SNP number by eliminating of less contribution, we can also use ENTER method with some minor modification as showed in result section. Briefly, after performing ENTER method we can choose SNPs which have p-value less than 0.05, in the "Variables in the Equation" table in SPSS output. We obtained 76.6% prediction score and 0.852±0.005 AUC with 193 SNP. This score is higher than the score of 114 SNPs, which remained in Forward LR method that AUC was 0.825±0.005 and overall percentage was 74.4%. The detail of analysis were given in results section and discussed in discussion.

- ENTER: A procedure for variable selection in which all variables in a block are entered in a single step.
- Forward Selection (Conditional): Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on conditional parameter estimates.
- Forward Selection (Likelihood Ratio): Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on the maximum partial likelihood estimates.
- Forward Selection (Wald): Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of the Wald statistic.
- Backward Elimination (Conditional): Backward stepwise selection. Removal testing is based on the probability of the likelihood-ratio statistic based on conditional parameter estimates.
- Backward Elimination (Likelihood Ratio): Backward stepwise selection. Removal testing is based on the probability of the likelihood-ratio statistic based on the maximum partial likelihood estimates.
- Backward Elimination (Wald): Backward stepwise selection. Removal testing is based on the probability of the Wald statistic.

2.6.4.1.3 Nagelkerke R^2

It is used to measure the usefulness of the model and that are similar to the coefficient of determination (R^2) in linear regression [39]. The Cox & Snell and the Nagelkerke R^2 are two such statistics. The maximum value that the Cox & Snell R^2 attains is less than 1. The Nagelkerke R^2 is an adjusted version of the Cox & Snell R^2 and covers the full range from 0 to 1, and therefore it is often preferred. The R^2 statistics do not measure the goodness of fit of the model but indicate how useful the explanatory variables are in predicting the response variable and can be referred to as measures of effect size. If Nagelkerke R^2 is greater than 0.5, which indicates that, the model is useful in predicting case.

2.6.4.1.4 Asymptotic Significance (Asymp. Sig.) in ROC Analysis

The significance level based on the asymptotic distribution of a test statistic. Typically, a value of less than 0.05 is considered significant. The asymptotic significance is based on the assumption that the data set is large. If the data set is small or poorly distributed, this may not be a good indication of significance.

2.6.4.2 Population Attributable Risk (PAR)

We used PAR to understand the contribution and the risk of the SNPs on the development of diabetes. PAR was calculated by using the following formulae [40].

$PAR = (X-1)/X$ Equation 6

$X = (1-f)^2 + 2f(1-f)\gamma+f^2\gamma^2$ Equation 7

Where f is the frequency and γ is the estimated odd ratio of the risk allele.

2.6.4.3 Net Reclassification Improvement (NRI %)

NRI was calculated manually as a ratio of sum of the difference in control and diabetic case to the population. For example, if we add variable for BLR analysis and this variable cause 100 control and 5 diabetic case is predicted more correctly, assuming total sample 1000, so NRI is (100+50)*100/1000= 15%.

CHAPTER 3

RESULTS

3.1 General Results of Genome-wide Association Study

PLINK analysis revealed 34,289 SNPs that has individual *p*-value smaller than 0.05. The genomic locations of the SNPs are identified to map the coding SNPs to their related genes. Several genes identified to have more than one associated SNP, which are strongly indicator of potential loci associated with T2D. Distribution p-values after GWAS is summarized in Figure 3.1. Detailed list of P values, MAF, Odds ratios, and corresponding SNPs and genes are given in Appendix B.

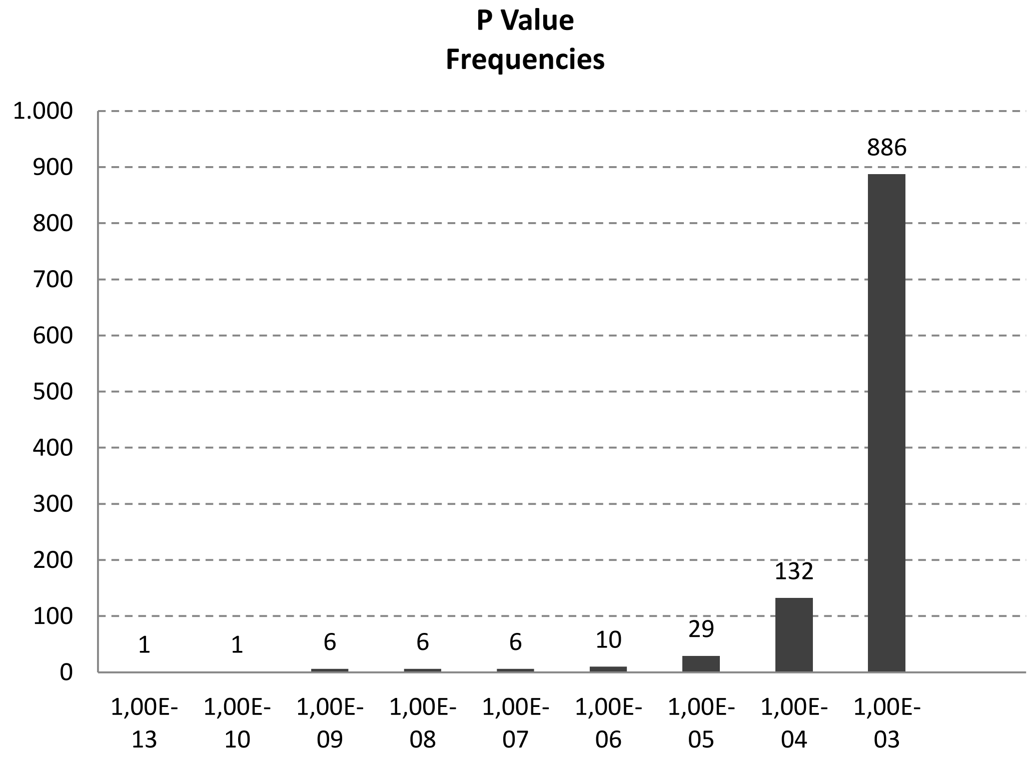


Figure 3.1 P value distribution of 886 SNPs.

An illustration of a Manhattan plot depicting several strongly associated risk loci is given in Figure 3.2. Each dot represents a SNP, with the X-axis showing genomic location and Y-axis showing association level.

Additionally, Manhattan plot of chromosome 9 and 10 which have strong association signals on them are given separately in Figure 3.3.

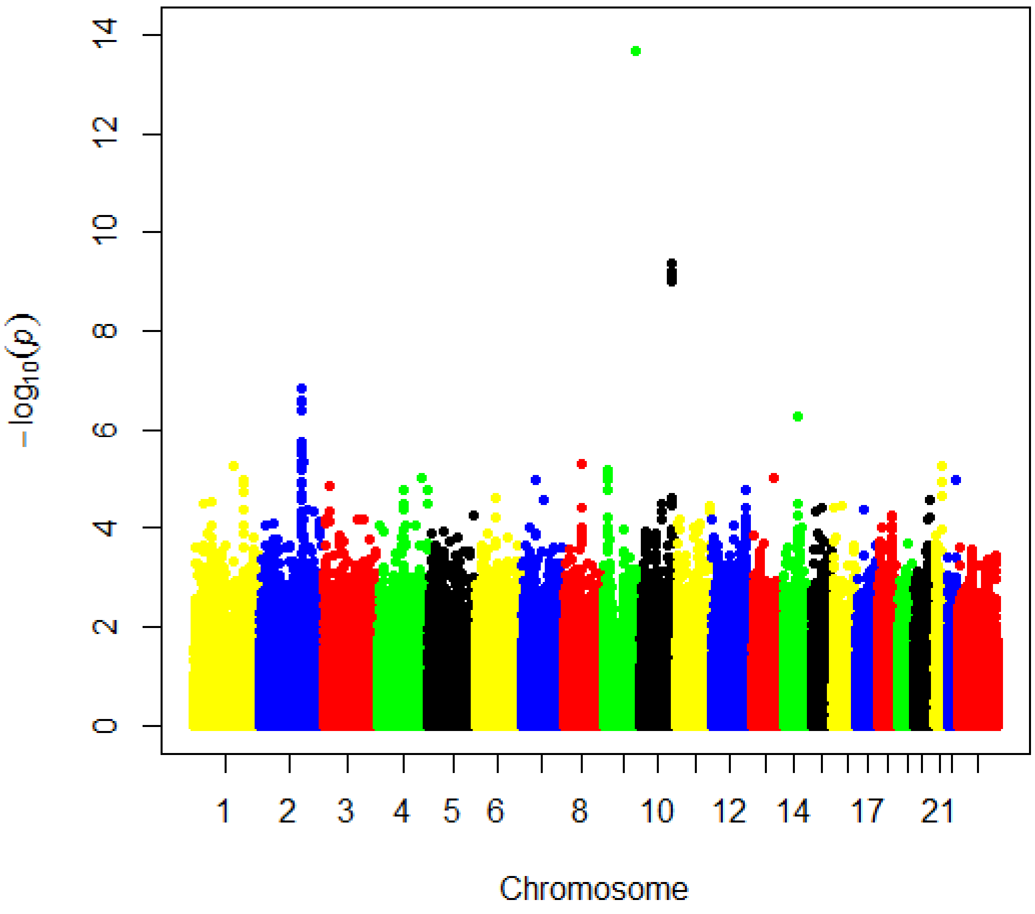


Figure 3.2 Manhattan Plot of the Pointwise P-values for the 642,576 SNP loci of the NHS and HPFS dataset.

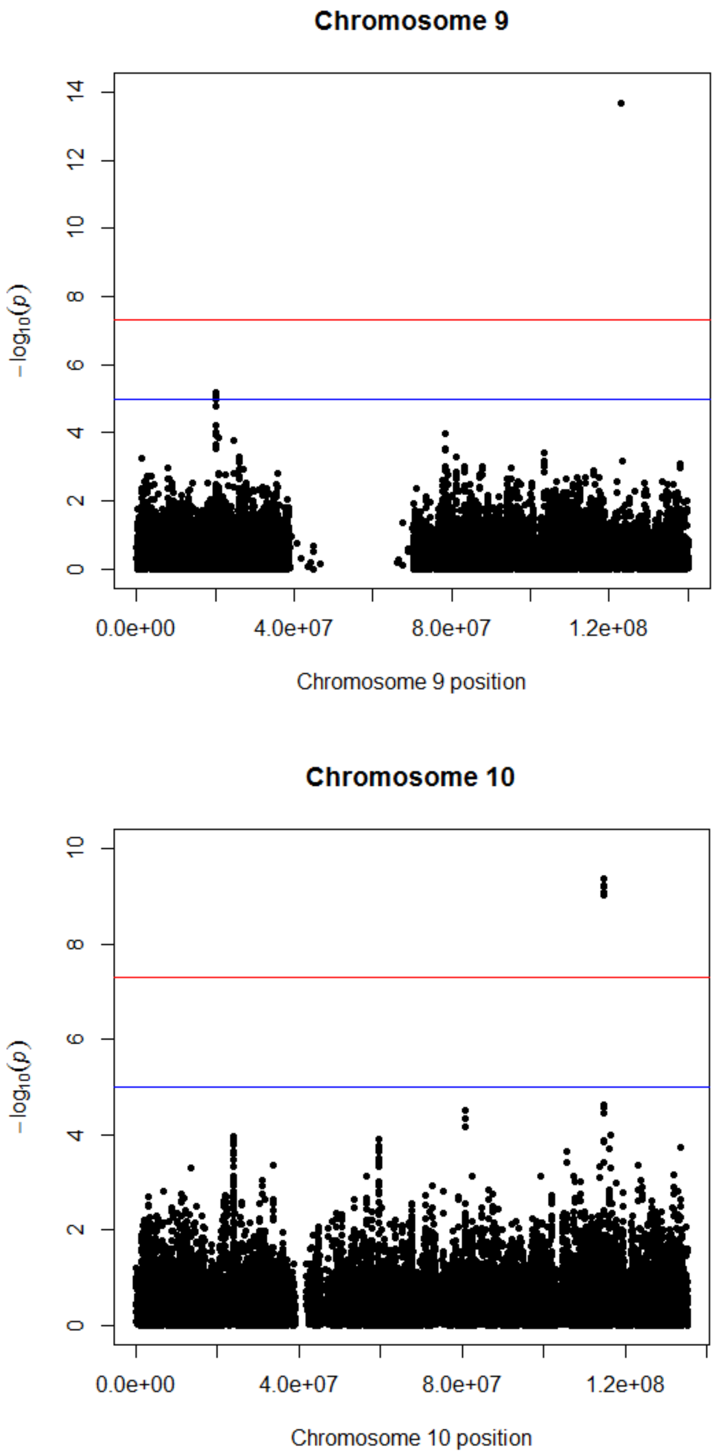


Figure 3.3 Manhattan plot of chromosome 9 and 10 in detail in general (NHS+HPFS) GWAS analysis.

Quantile-quantile plots of SNP P values in (NHS+HPFS) GWAS analysis is examined in order to set the p-value threshold as in Figure 3.4. Detaching point from the expected –log10, which was approximetly 1.0E-3 is set as the p-value threshold for selecting the associated SNPs in further analysis.

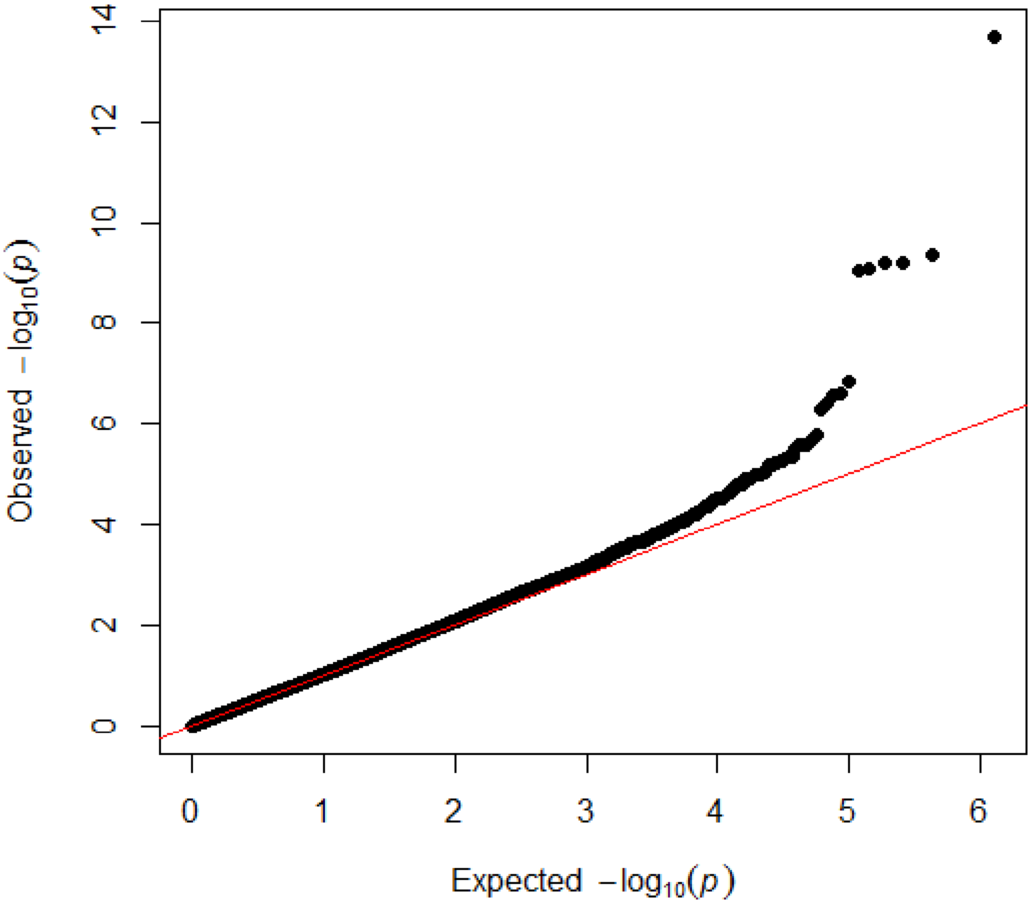


Figure 3.4 Quantile-quantile plots of SNP P values in (NHS+HPFS) GWAS analysis. The x-axis is –log10 of the expected P values and the y-axis is –log10 of the observed P values. Detaching point from the expected –log10 is nearly 1.0E-3.

3.2 Analysis of Individual Data Sets and Sex Based Association Results

When we analyzed male and female participants separately, the change in p-value association was significant in male.

3.2.1 GWAS Results of Nurses Health Study

The results of female participants is summarized in Figures 3.5, 3.6 and 3.7 as shown.

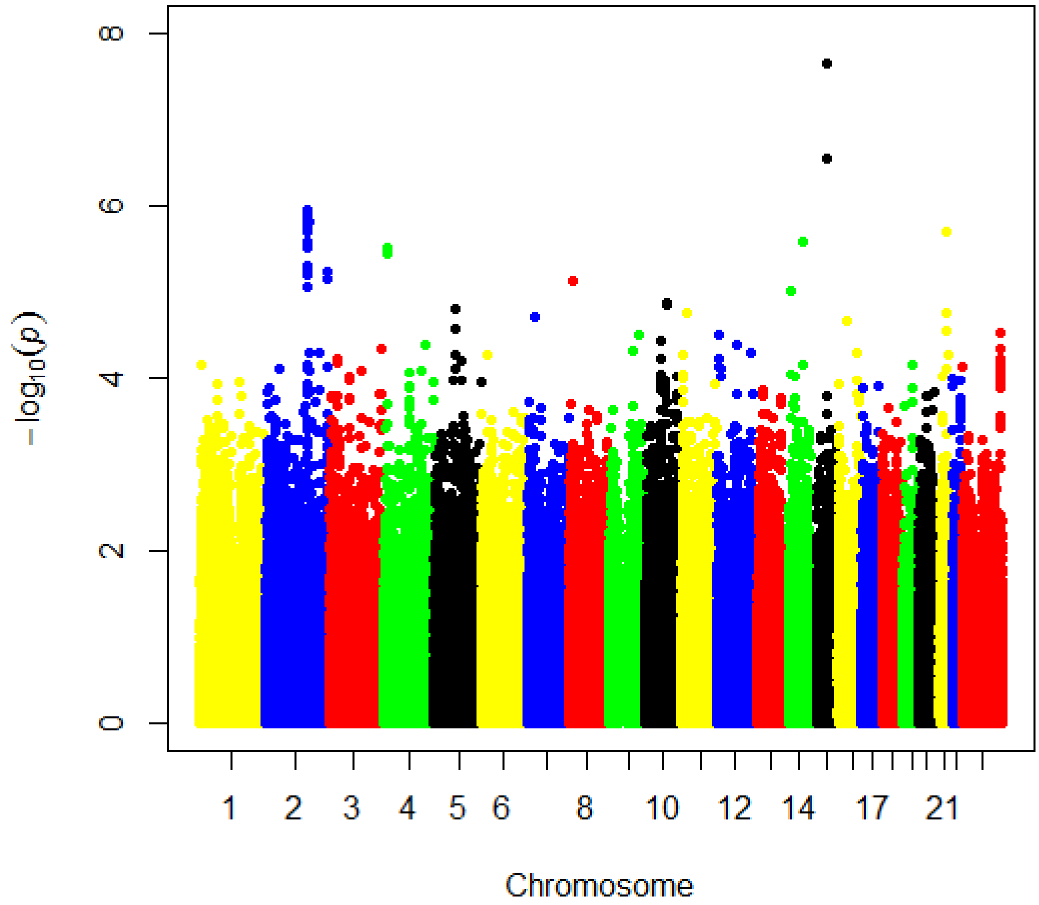
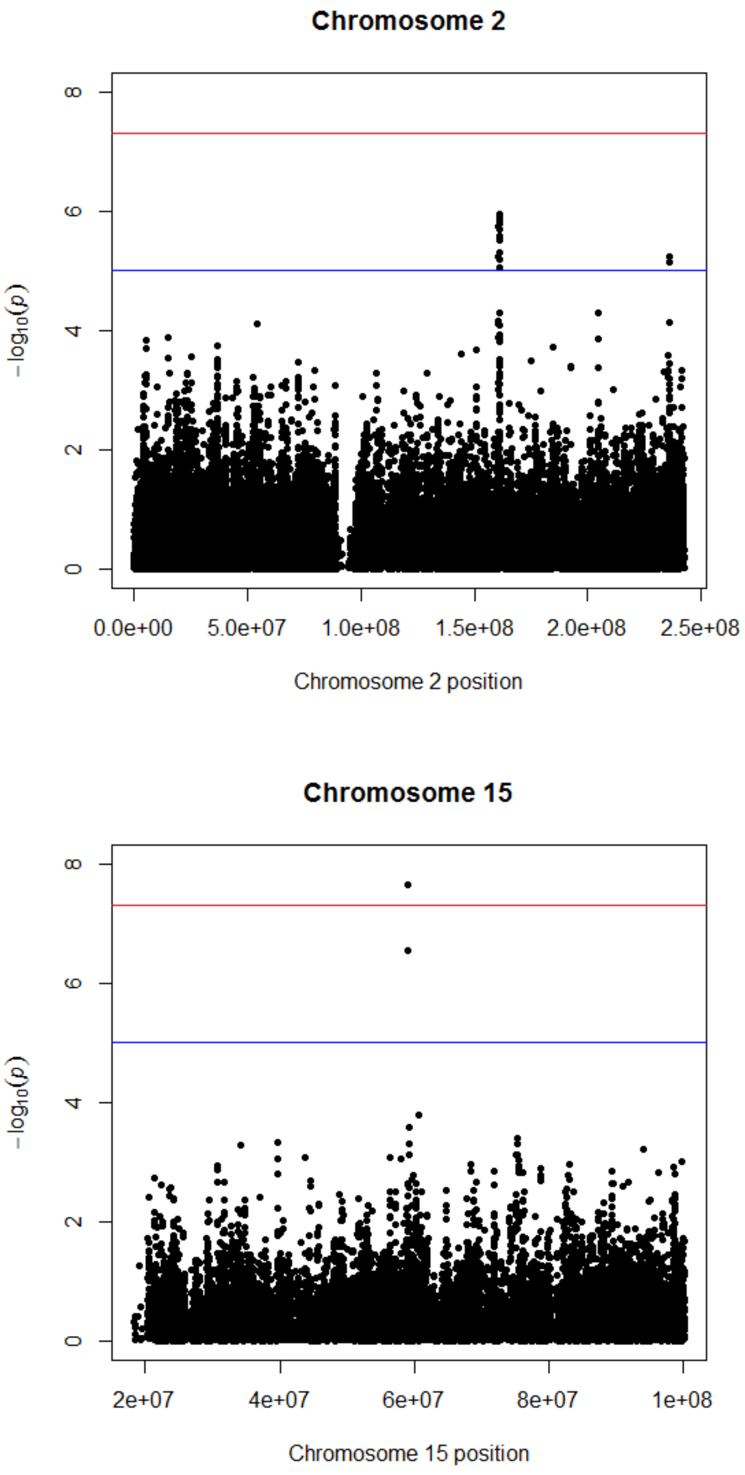


Figure 3.5 Manhattan plot of NHS GWAS results. In the contrary of general GWAS analysis and male participants, SNPs with lowest P value were lower than male participants. While male participants have strong signal on chromosome 9 and 10, female participants have strong signal on chromosome 2 and 15 as shown below.



Chromosome 2

Chromosome 15

Figure 3.6 Manhattan plot of chromosome 15 and 2 in detail in NHS GWAS analysis.

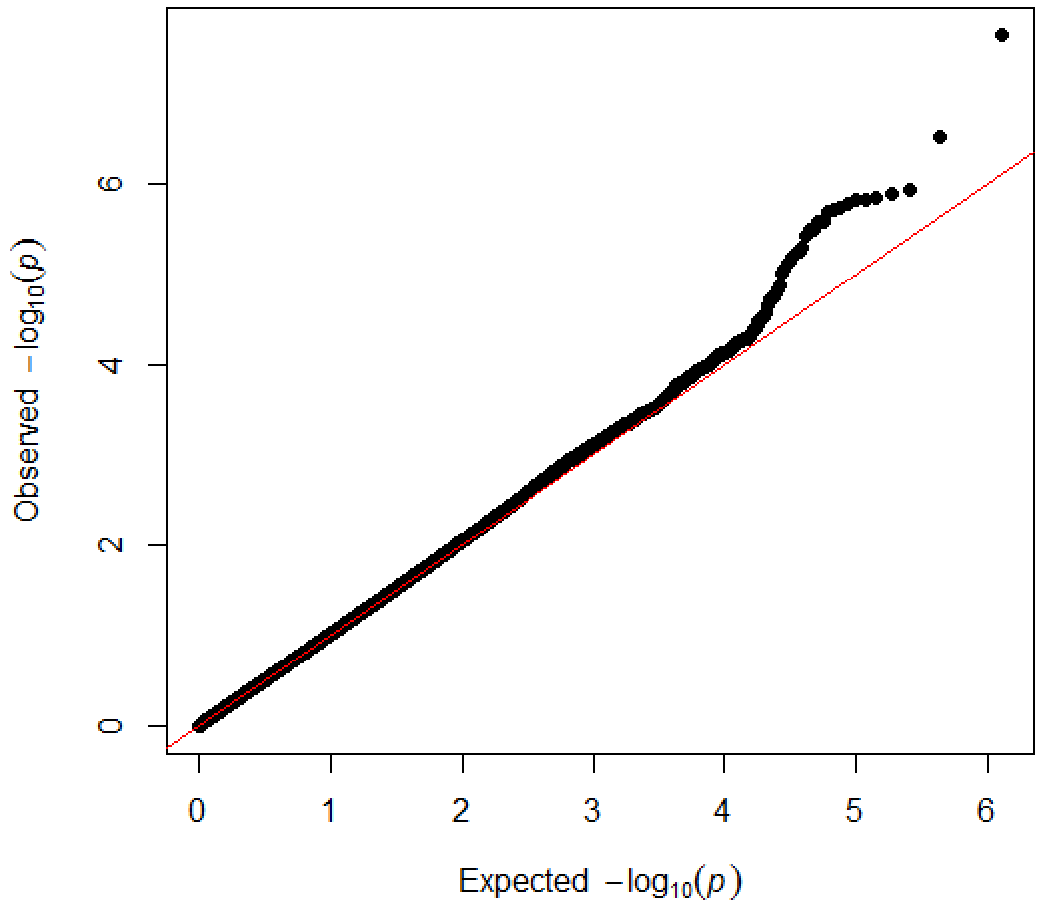


Figure 3.7 QQ plot of NHS (all female) case and controls showing expected and observed p values of SNPs. The most significant p values of SNPs showed detaching from observed curve line (right dots). While detaching point from the expected was around 3 (-log P), in female participants it was around 4. This is important point, since the number of SNPs between 3 and 4 is 604. Since SNP number is important which affecting prediction score, the threshold P level for choosing SNP is important.

3.2.2 GWAS Results of HPFS

The results of male participants is summarized in Figures 3.8, 3.9 and 3.10 as shown.

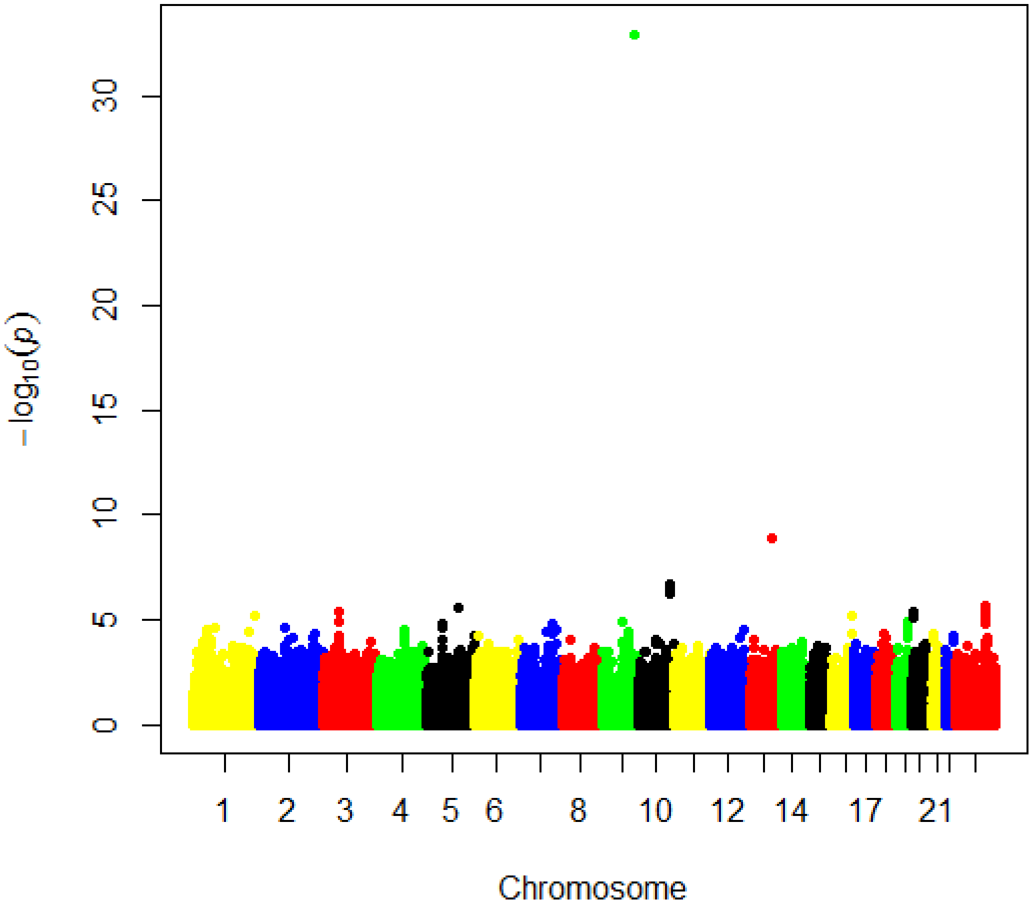


Figure 3.8 Manhattan plot of HPFS GWAS results.

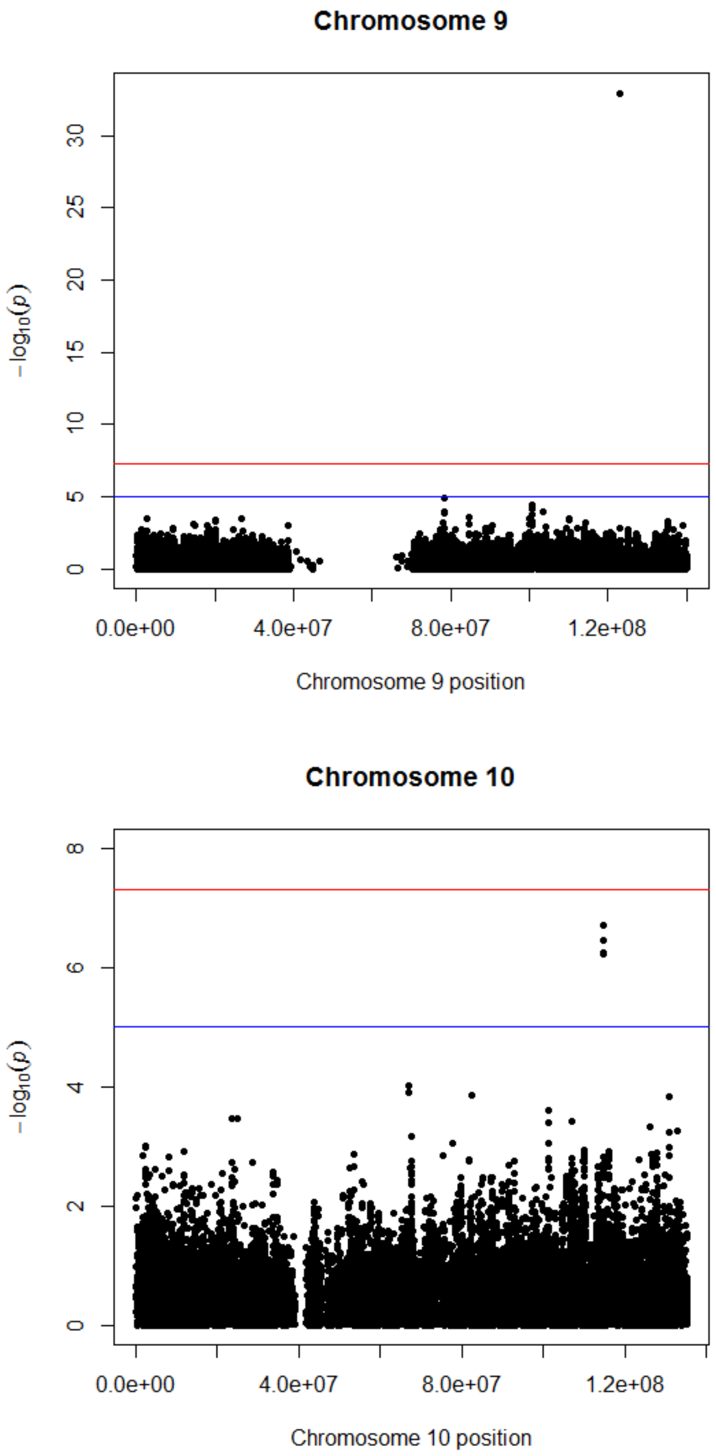


Figure 3.9 Manhattan plot of chromosome 9 and 10 in detail in HPFS GWAS results.

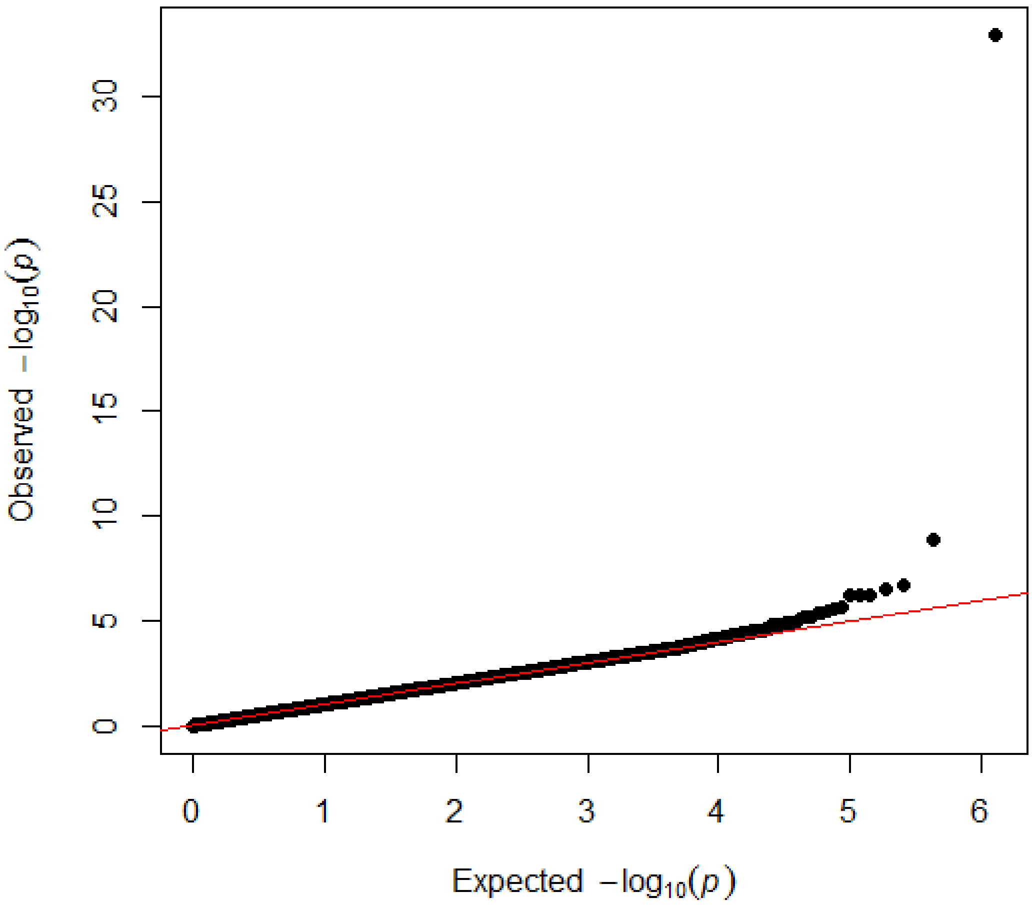


Figure 3.10 QQ plot of HPFS case and controls showing expected and observed p values of SNPs. The most significant p values of SNPs showed detaching from observed curve line (right dots)

3.3 Biological Interpretation of the GWAS Results

Previously, number of SNPs related with T2D risk have been reported in the literature, which are on the chromosome 1. One of these loci is chromosome 1q21-q23. Within this region, T2D was associated with a common single nucleotide polymorphisms that marked an extended linkage disequilibrium block, including the liver pyruvate kinase gene (*PKLR*) [41]. Genes near to *PKLR* (*HCN3*, *CLK2*, *SCAMP3*, and *FDPS*) were also investigated. Location of these nearby genes are given in Table 3.1.

Table 3.1 Genes in close proximity to the *PKLR* on chromosome 1.

Row	**Chr**	**StartPosition**	**EndPosition**	**Entrez ID**	**HUGO id**	**ENSEMBLE id**
1	1	69055	70108	79501	*OR4F5*	ENSG00000177693
2	1	860260	879955	148398	*SAMD11*	ENSG00000187634
....						
1169	1	155204243	155214488	2629	*GBA*	ENSG00000177628
1170	1	155216996	155225274	10712	*FAM189B*	ENSG00000160767
1171	1	155225770	155232221	10067	*SCAMP3*	ENSG00000116521
1172	1	155232659	155248282	1196	*CLK2*	ENSG00000176444
1173	1	155247374	155259639	57657	*HCN3*	ENSG00000143630
1174	1	155259086	155271225	5313	*PKLR*	ENSG00000143627
1175	1	155278539	155290457	2224	*FDPS*	ENSG00000160752
1176	1	155290687	155300905	23623	*RUSC1*	ENSG00000160753
1177	1	155305059	155532484	55870	*ASH1L*	ENSG00000116539
1178	1	155579996	155584758	55154	*MSTO1*	ENSG00000125459
1179	1	155629237	155658791	55249	*YY1AP1*	ENSG00000163374
1180	1	155657751	155708803	7818	*DAP3*	ENSG00000132676

The GWAS results presented previously identified several SNPs, which are listed in Table 3.2, mapped to the *ASH1L* gene (*ASH1L* gene (ash1 (absent, small, or homeotic)-like (Drosophila)), with potential association with increased risk of T2D. This gene is also at very close position to the previously found genes in the literature.

Table 3.2 SNPs mapping to ASH1L gene analyzed in the study and their p-values.

SNPs for ASH1L gene	Chr	Position	A1	A2	P-value
rs11264363	1	153584932	G	C	0.001
rs12041534	1	153673720	T	C	0.003
rs12724079	1	153700566	T	C	0.003
rs1325908	1	153679928	C	A	0.003
rs11264375	1	153690689	C	T	0.003
rs10908470	1	153793637	G	T	0.004
rs11264381	1	153789196	C	T	0.004
rs5005770	1	153611667	G	A	0.004

A1: minor allele, A2: major allele, ASH1L (gene name) : ash1 (absent, small, or homeotic)-like (Drosophila)

Both *ASH1L* and *PKLR* genes have been investigated previously by the "International Type 2 Diabetes 1q Consortium" for their association with SNPs in T2D [42]. Our findings about the *ASH1L* gene confirms previous studies and show the functionalities of METU-SNP. We have also found new candidate genes, which were previously not reported, such as two candidate genes *PLOD1* and *CAPZB*, which are shown below in Table 3.3.

Table 3.3 Potential new candidate gene for diabetes

rsid	AHP_score	Chr	Position	P value	HUGO_id
rs2336381	0.445599	1	12009024	9.00E-04	PLOD1
rs7529705	0.445599	1	19720092	5.09E-04	CAPZB
rs10492998	0.445599	1	19772847	8.10E-04	

PLOD1	procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1
CAPZB	capping protein (actin filament) muscle Z-line, beta

In consistent with the findings of "Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium" [15], we also found strong signals in chromosome 2 related with T2D. It is interesting that this signal is more apparent in females than male cases, whereas it is the otherway for *TCF7L2*, where the signal is more dominant in males than female cases. Gender differences in GWAS analysis was not strongly noticed in previous studies [15]. Additionally, some of the SNPs mapping to binding motif, single stranded interacting protein 1 (*RBMS1*) gene, were found to have significant association (lower p-value) in NHS (female) study, but did not reached significance level in HPFS with male cases. This finding implicates that the results of GWAS results should be carefully evaluated according to gender. The details of *TCF7L2* and *RBMS1* gene analysis is given in Appendix C.

In GWAS analysis, the higher patient number is desirable. It could be possible to find the lowest p value. However, this approach may not be suitable to find specific markers for specific conditions. For example, some markers could be dominant in male whereas some of them in female. According to our knowledge, this issue has not been noticed in detail so far. TCF7L2 gene is one of the most important location in GWAS analysis of diabetes. We also found TCF7L2 statistically significant genes showing risk of diabetes. We found 19 SNPs related with TCF7L2 gene. However, as it could be noticed below, male patients are more susceptible to diabetes according to their p values of TCF7L2 gene.

In addition, other SNPs on TCF7L2 gene (rs12255372 [43], rs7901695 [44], rs4506565 [45], rs10885409 [46] and rs11196205) [47] have been mentioned in the literature for their association with T2D. We have additionally found rs12243326, rs4132670, rs11196208 as additional candidate variations during our analysis.

3.4 The Detail Analysis of SNP rs10739592

rs10739592 has been revealed with the lowest p-value in our analysis which was not reported previously. When we have explored it in detail, we have revealed that this SNP was significantly associated with G allele only in male cases. While its p-value in general is 2.08E-14, the significance increases to 1.19E-33 in males. We do not have further information about this SNP. It is not mapped to any known gene. But, "*RAB14:* GTPase Rab14" gene is located in its proximal region and "*GSN*: Gelsolin isoform b" gene is located in the distal region of rs10739592 reported by the Haploview analysis. The details of the Haploview analysis and distribution density of rs10739592 in control and diabetic cases is given in Appendix D.

3.5 Binary Logistic Regression Analysis of Phenotype Variables

Before binary logistic regression, in order to define the phenotype variables with potential effect on T2D, first we have performed conventional statistical analysis of the phenotype variables between control and diabetic patients. Further information about the statistical analysis of the phenotype variables is given in Appendix E.

Next, we have analyzed phenotype variables by BLR. The result of analysis is summarized in Table 3.4. The most significant phenotypic variables were found to be BMI, familial diabetes history and high blood pressure. Gender, age, activity, polyunsaturated fat intake, magnesium intake, and trans fat intake were not found significant for T2D risk.

BMI had the lowest p-value (5.21E-108) and highest odds ratio (3.86). At the start point the overall prediction correctness percent was 54%, when we add BMI as a parameter, prediction accuracy increased to 68.0%, which means net reclassification index of BMI was 13.99%. Therefore, the most important variables following BMI were familial diabetes history, high blood pressure, and cholesterol. When we combined four phenotype variables it yielded 16.7% NRI, 70.7% overall prediction accuracy, 0.77 AUC and the combined p-value was 1.56E-187.

The classification table is a method to evaluate the predictive accuracy of the logistic regression model. In this table the observed values for the dependent outcome and the predicted values (at a user defined cut-off value, for example p=0.50) are cross classified. Classification table cutoff value could be between 0 and 1 which will be used during the classification.

Table 3.4 Phenotype features by the aspects of NRI, overall prediction, AUC, P value and odds ratio.

Phenotype	NRI %	Overall Prediction %	AUC	P value	Odds ratio
Start level	n.a.	54	n.a.	n.a.	n.a.
Body mass index (BMI)	13.99	68.0	0.677	5.21E-108	3.86
Familial diabetes history (FAMDB)	9.70	63.7	0.625	4.32E-69	3.10
High Blood Pressure (HBP)	9.68	63.7	0.623	3.25E-39	2.37
Cholesterol (CHOL)	4.40	58.4	0.564	7.76E-15	1.74
Four phenotypes (BMI+FAMDB+HBP+CHOL)	16.7	70.7	0.770	1.56E-187	n.a.
rs10739592	2.84	56.9	0.552	2.08E-14	1.34

n.a., not applicable.

Table 3.5 AUC for four phenotype variables (BMI, FAMDB, HBP, and CHOL).

Test Result Variable(s)	Area	Std. Error (a)	Asymp -totic Sig.(b)	Asymptotic 95% Confidence Interval	
				Upper Bound	Lower Bound
BMI	.677	.007	.000	.663	.692
FAMDB	.625	.008	.000	.610	.639
HBP	.623	.008	.000	.609	.638
CHOL	.564	.008	.000	.549	.579
BMI+FAMDB+HBP+ CHOL	.770	.006	.000	.758	.782

The test results variable(s): Phenotype has at least one tie between the positive actual state and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption, b. Null hypothesis: true area = 0.5

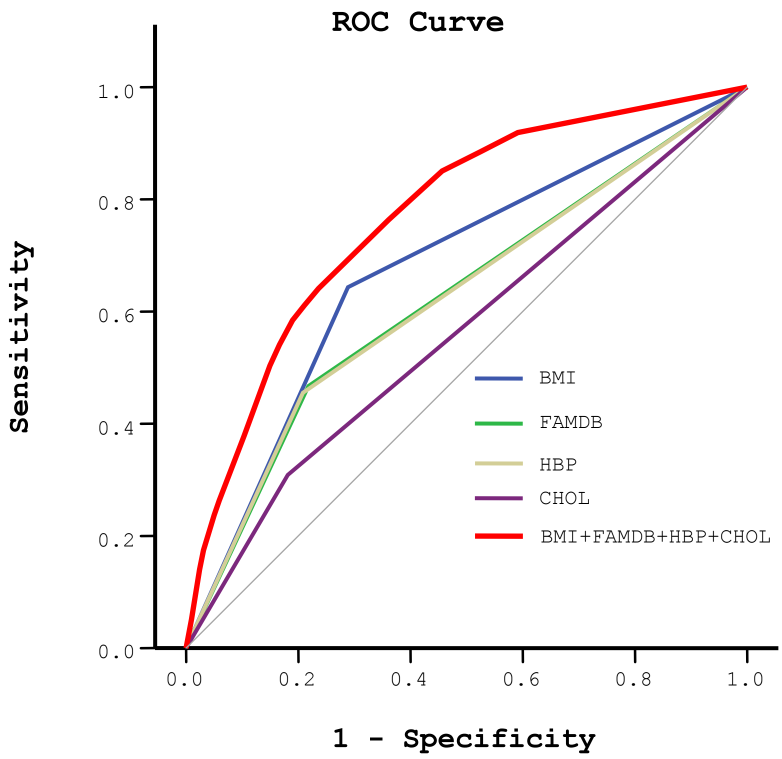


Figure 3.11 ROC curve for four phenotype variables (BMI, FAMDB, HBP, and CHOL).

3.6 Body Mass Index (BMI) Phenotype Analysis

Since BMI was the most important phenotype variable, we investigated its contribution in more detail. Actual BMI variable was continuous but we converted it to binary form. We used Youden Index (YI) for conversion as explained below.

Table 3.6 Body mass index values of male and female in control and diabetic case.

	Male	n	Female	n	Average	n
Control	25.21 ± 2.82	1277	25.39 ± 4.83	1769	25.31 ± 4.11	3046
Diabetes	27.89 ± 4.14 [a]	1114	29.91 ± 5.76 [b]	1479	29.04 ± 5.22 [c]	2593
Average	26.45 ± 3.74	2391	27.44 ± 5.73	3248	27.03 ± 5.01	5639

[a] Independent sample *t* test, 3.72E-115, [b] Independent sample *t* test, 1.52E-68

[c] Independent sample *t* test, p< 1.85E-174

When we have performed Independent Sample t test for BMI, P value was 1.94E-182. However, it is not preferred to perform binary logistic regression with continuous variables, so we converted BMI into binary data. The Youden Index (YI= Sensitivity + Specificity − 1) is used to determine threshold level for BMI conversion from continous to binary form. The value which maximizes YI was selected as a threshold, and it was found to be different for for male and female, 27.1 and 26.3 respectively as presented in Table 3.7 and 3.8. YI of training and test groups are similar and not different from each other. The details of YI analysis is given in Appendix F.

Table 3.7 Youden Index for male in whole cases (n=5639).

Threshold	25	26	27	28	**27.1**	26.3
Positive Predictive Value	0.571	0.625	0.680	0.733	0.693	0.637
Negative Predictive Value	0.709	0.677	0.659	0.634	0.657	0.671
Likelihood Ratio +	1.523	1.910	2.430	3.149	2.586	2.011
Likelihood Ratio -	0.471	0.546	0.593	0.661	0.598	0.562
Sensitivity	0.766	0.636	0.539	0.429	0.522	0.608
Specificity	0.497	0.667	0.778	0.864	0.798	0.698
YI index	0.263	0.303	0.317	0.293	**0.320**	0.305

Table 3.8 Youden Index for female in whole cases (n=5639).

Threshold	25	26	**26.3**	27	28	27.1
Positive Predictive Value	0.603	0.634	0.642	0.656	0.671	0.656
Negative Predictive Value	0.762	0.741	0.739	0.713	0.682	0.711
Likelihood Ratio +	1.815	2.073	2.144	2.282	2.438	2.279
Likelihood Ratio -	0.373	0.418	0.421	0.481	0.558	0.486
Sensitivity	0.789	0.729	0.720	0.658	0.573	0.653
Specificity	0.565	0.648	0.664	0.712	0.765	0.713
YI index	0.354	0.377	**0.384**	0.370	0.338	0.367

3.7 The Effects of Other Phenotype Variables on Prediction Rate and AUC

We have selected only four phenotypes, BMI, FAMDB, HBP, and CHOL to test the effects of phenotype variables on prediction rate and AUC. We have also tested other phenotype variables, (such as activity, smoking, and alcohol), on prediction rate and AUC. Although the latter three phenotype variables were found significantly related with diabetic status, the contribution of these variables to the classification and the AUC were negligible. Alcohol increases the prediction rate only 0.2%. The increase in prediction rate, and in AUC was also small for smoking and activity. In addition, activity and alcohol are continuous variables which makes BLR analysis complicated. Alcohol, smoking, and activity increased overall prediction rate only by 0.8%, and AUC only by 0.6% when added onto the first four variables selected. Because their contribution is negligible, we continued our analysis with BMI, FAMDB, HBP, and CHOL as representatives of phenotype variables for subsequent BLR analysis. The details of binary logistic regression analysis of phenoytpe variables is given in Appendix G.

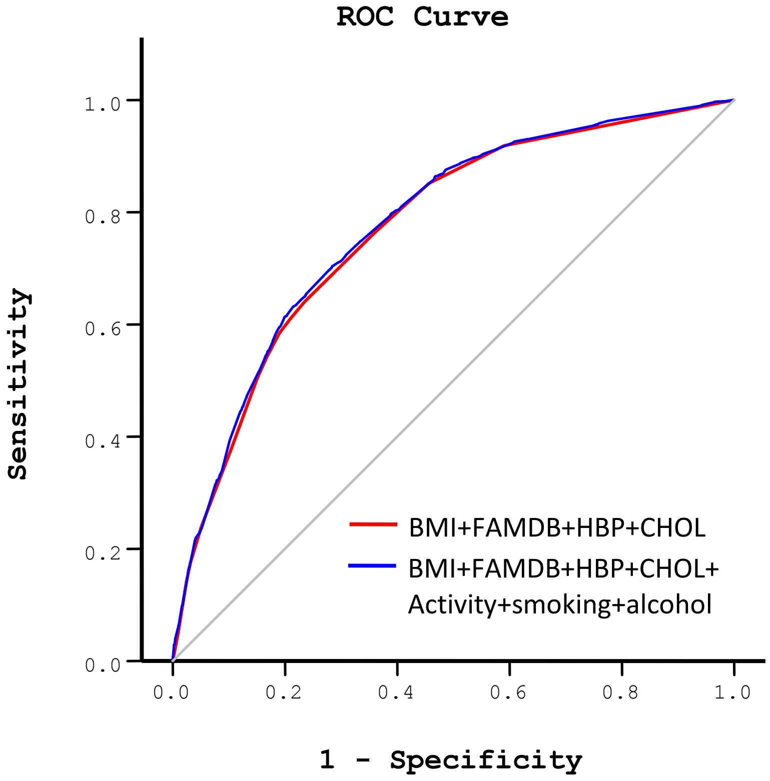


Figure 3.12 Comparison of ROC curves for four and seven phenotype variables (red line; BMI, FAMDB, and CHOL, HBP) and additional three variables (blue line; four variables plus activity, smoking, and alcohol).

3.8 BLR Analysis of Genotype

First, 886 SNPs which have p-value lower than 1.0E-3 are selected for further studies, and eliminated some of them which had high number of missing allele data. Selection and elimination criteria were explained in method section. The list of SNPs included in the analysis were given in the Appendix A.

The p-value distrubution and the chromosomal locations of the selected SNPs are represented in Figures 3.13 and 3.14, respectively. Manathan plot in Figure 3.2 revealed that the chromosomes 2, 1, 12, 10 and 3 are the most important amongst the chromosomes which carries higher number of significantly associated SNPs, indicating potential loci for T2D.

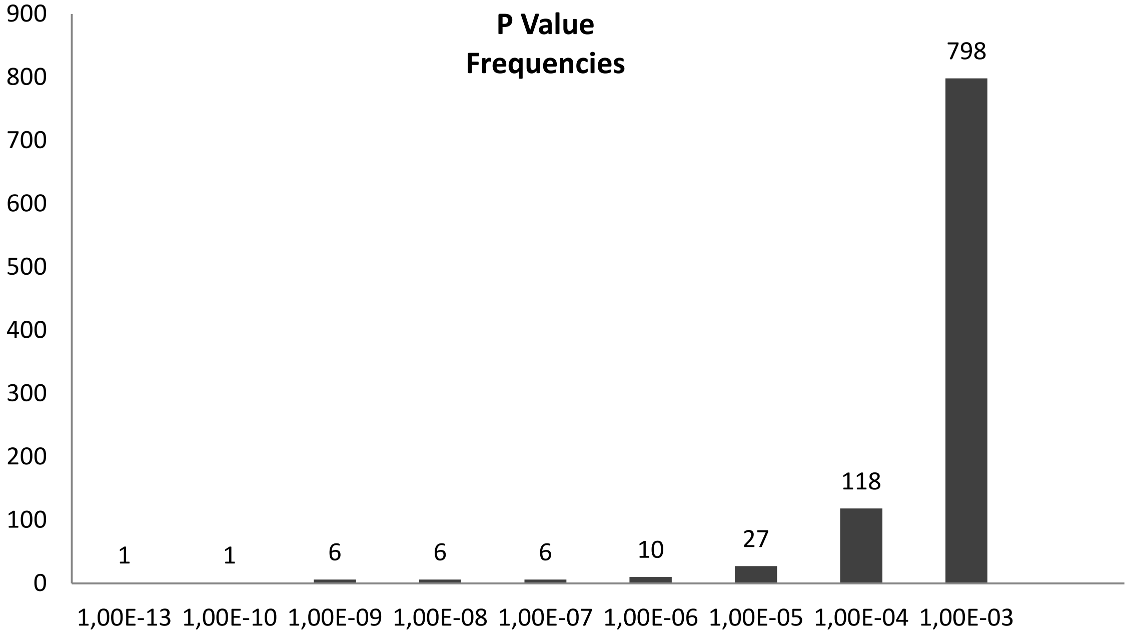


Figure 3.13 Cumulative frequency of P values of 798 SNPs.

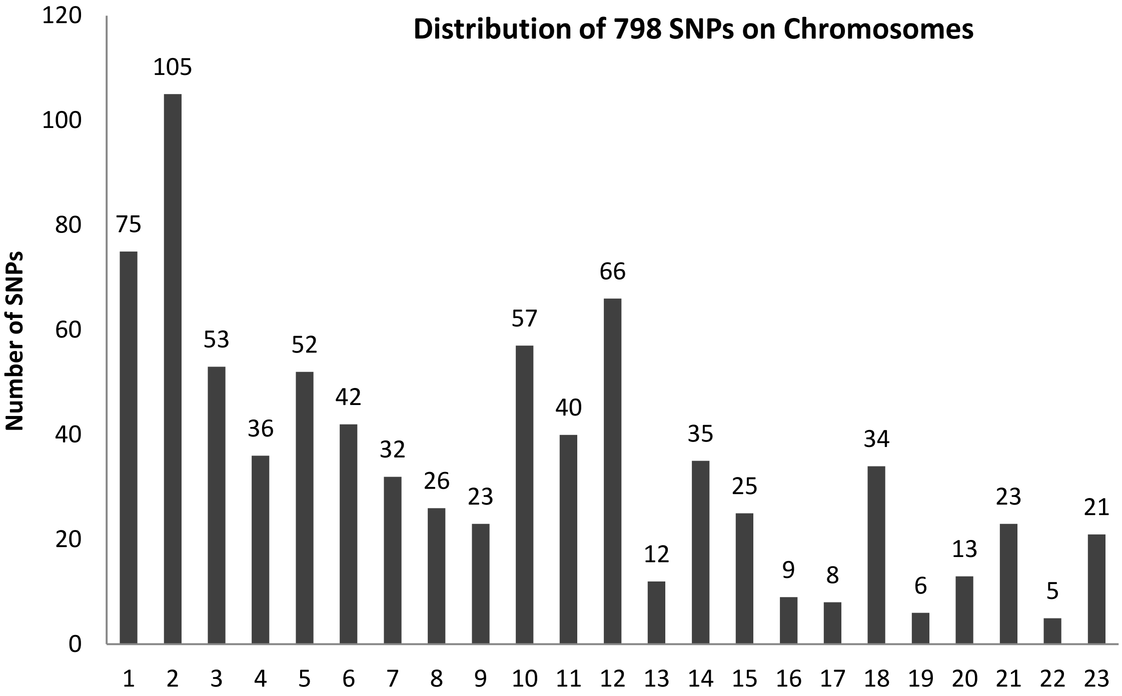


Figure 3.14 Distribution of 798 SNPs on the Chromosomes

3.8.1 The Contribution of Each Chromosome to the Prediction of the Diabetes Risk

The 798 SNPs selected based on the p-value threshold for the BLR analysis is used to investigate the contribution of each chromosome for risk prediction of diabetes. This was first reported study in which hundreds of SNPs are used for T2D classificaiton. The overall prediction rate was between the range of 54.8% and 63.1%, with an AUC range of 0.55 and 0.68. The details of binary logistic regression analysis of each chromosome are given in Appendix H.

3.8.2 BLR Analysis with 798 selected SNPs

We analyzed 798 selected SNPs with BLR. Classification table is given in Table 3.9, AUC and ROC curve calculations are given in Table 3.10 and Figure 3.15.

Table 3.9 Classification Table of the 798 SNPs obtained with BLR analysis [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2762	284	90.7
		Diab	282	2311	89.1
Overall percentage					90.0

a. The cut value is 0.5

Table 3.10 Area Under the Curve for 798 SNPs.

Area	Std. Error (a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Upper Bound	Lower Bound
.965	.002	.000	.961	.969

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

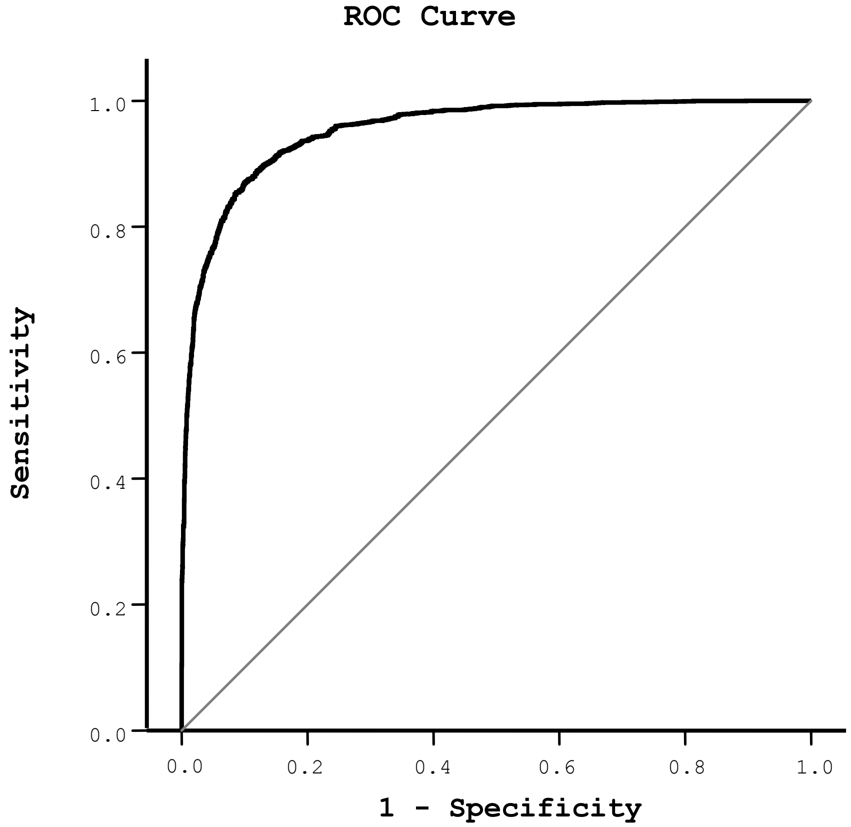


Figure 3.15 ROC curve of 798 SNPs. This was first reported study in which hundreds of SNPs are used for T2D classification which yielded AUC of 0.965.

3.8.3 Genotype Analysis in Training and Test Groups

Our dataset comprised 5639 data sets (3046 control and 2593 diabetes). We divided our data set into two groups randomly using SPSS, one is comprises 80% of dataset which is used as control set, the other is test dataset comprises 20% of dataset and used as a validation group. Control and validation datasets were compared using chi square test to determine equality of datasets to each other on the context of phenotype variables. Training and test groups were demographically, phenotypically and genotypically balanced, so statistically were not different from each other. Initially, we had 798 SNPs with p-value lower than 1.0E-3. We performed binary logistic regression using 798 SNPs with ENTER method. Then, we chose 225 SNPs, since not to exceed 5 events per variable, from ENTER method based on significance level obtained in SPSS from the "variables in the equation" table and performed binary logistic regression in validation group using 225 SNPs. We performed binary logistic regression analysis in three samplings with different training and test groups. The 225 SNPs selected in each sampling only overlapped at 66.67% and 62.67%, between samplings 1 and 2, 1 and 3 respectively. We did not find statistical difference amongst the groups for the predictive performance. Therefore, no further additional sampling is done. Although overall prediction and AUC is a bit higher in training group than test group, this difference is reasonable and comes from the number of SNPs used. The details of binary logistic regression analysis of training and test groups is given in Appendix I.

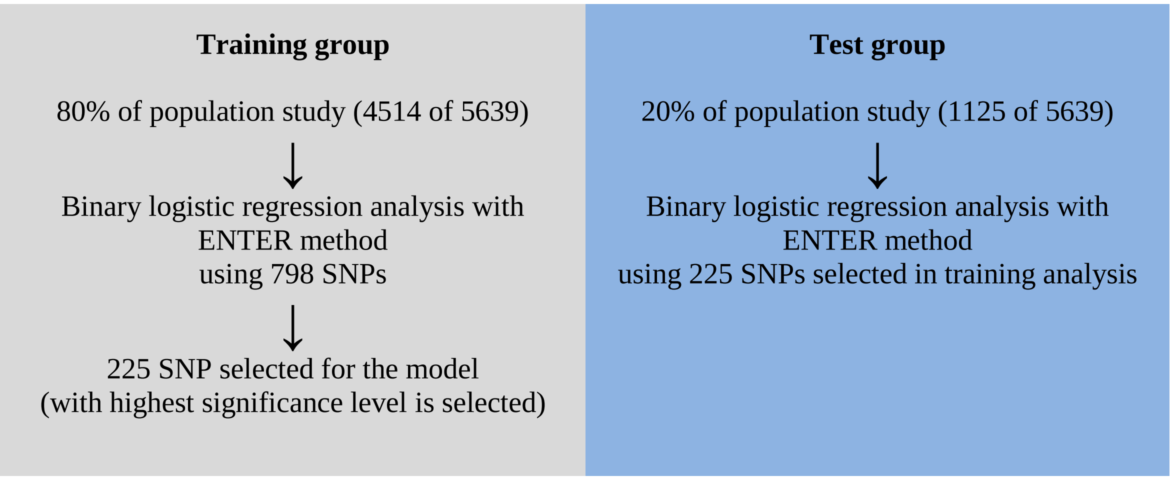

Training group
80% of population study (4514 of 5639)
Binary logistic regression analysis with ENTER method using 798 SNPs
225 SNP selected for the model (with highest significance level is selected)
Test group
20% of population study (1125 of 5639)
Binary logistic regression analysis with ENTER method using 225 SNPs selected in training analysis

Figure 3.16 Schematic representation of analysis of training and test groups

Table 3.11 The results of binary logistic regression analysis of training groups.

Control Groups (80 % of population)	NPV	PPV	Overall prediction	AUC	Statistic
Sampling 1	94.05	92.18	93.19	0.981	No significant difference
Sampling 2	95.05	95.51	93.89	0.984	
Sampling 3	94.72	93.07	93.95	0.985	

Table 3.12 The results of binary logistic regression analysis of test groups.

Validation Groups (20 % of population)	NPV	PPV	Overall prediction	AUC	Statistic
Sampling 1	90.22	87.91	89.14	0.957	No significant difference
Sampling 2	91.03	89.87	90.49	0.958	
Sampling 3	91.67	86.83	89.51	0.962	

3.8.4 BLR Analysis with Integrated Phenotype and Genotype Data

The comparison results of BLR analysis genotype and phenotype were given in Figure 3.17. While genotype analysis (798 SNPs) yielded 90% prediction power, phenotype analysis was only 77%. The additive contributions of phenotype and genotype increased the overall correctness from 90% to 92.9%, and AUC to 0.980. Net reclassification improvement of integrating phenotype data with genotype was 2.9%. Therefore, genotypic variables were found sufficient to achieve high prediction correctness without phenotype data.

Table 3.13 Classification table for genotype (798 SNPs) plus phenotype (BMI, FAMDB, CHOL, and HBP).

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2841	205	93.3
		Diab	194	2399	92.5
Overall percentage					92.9

a. The cut value is 0.5

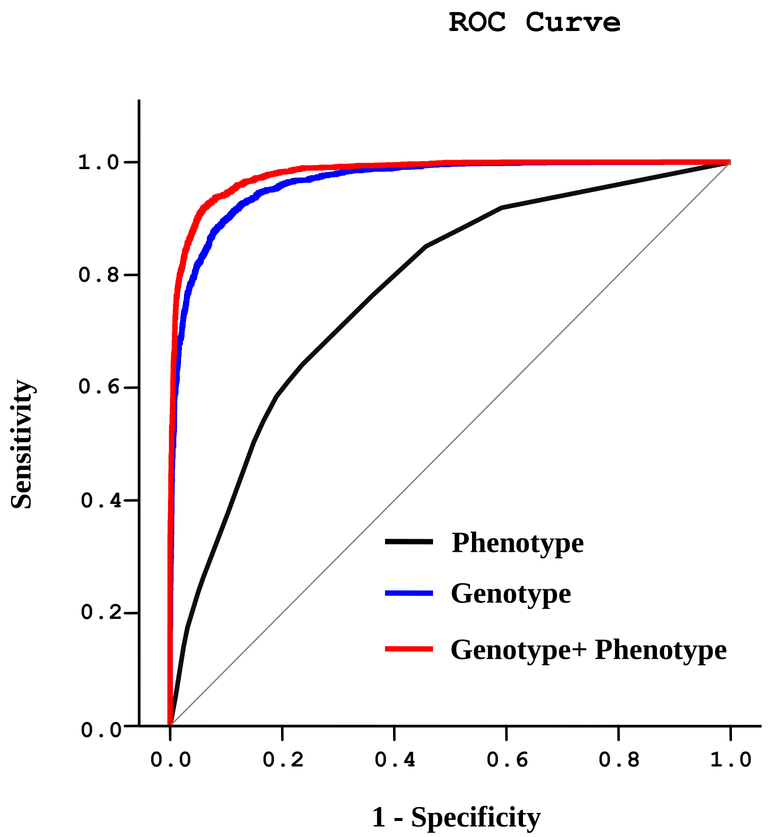


Figure 3.17 ROC Curve of genotype (798 SNPs), phenotype (BMI, FAMDB, CHOL, and HBP), and genotype plus phenotype.

Table 3.14 Area Under the Curve for genotype data, phenotype data, and integrated genotype and phenotype data

Test Result Variable(s)	Area	Std. Error [a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Upper Bound	Lower Bound
Genotype (798 SNPs)	.965 [c]	.002	.000	.961	.969
Phenotype	.770	.006	.000	.758	.782
Genotype plus Phenotype	.980 [d]	.001	.000	.978	.983

The test result variable(s): Phenotype has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

[a] Under the nonparametric assumption, [b] Null hypothesis: true area = 0.5

[c] p<0.001 vs phenotype, [d] p<0.001 vs phenotype, and genotype

3.8.5 Comparison of Genotypic Variables Depending on P values of SNPs in BLR Analysis

We wanted to determine the contribution of SNPs according to their P value. Thus, we grouped 780 SNPs as a P value lower than 1.0E-6, between 1.0E-06 and 1.0E-05, etc. The results were shown below. We realized that lowest P value might be important but not sufficient for prediction of diabetes in our study, so we should increase SNP numbers at least towards P value of 1.0 E-3. The details of binary logistic regression analysis of groups depending on P value is given in Appendix K.

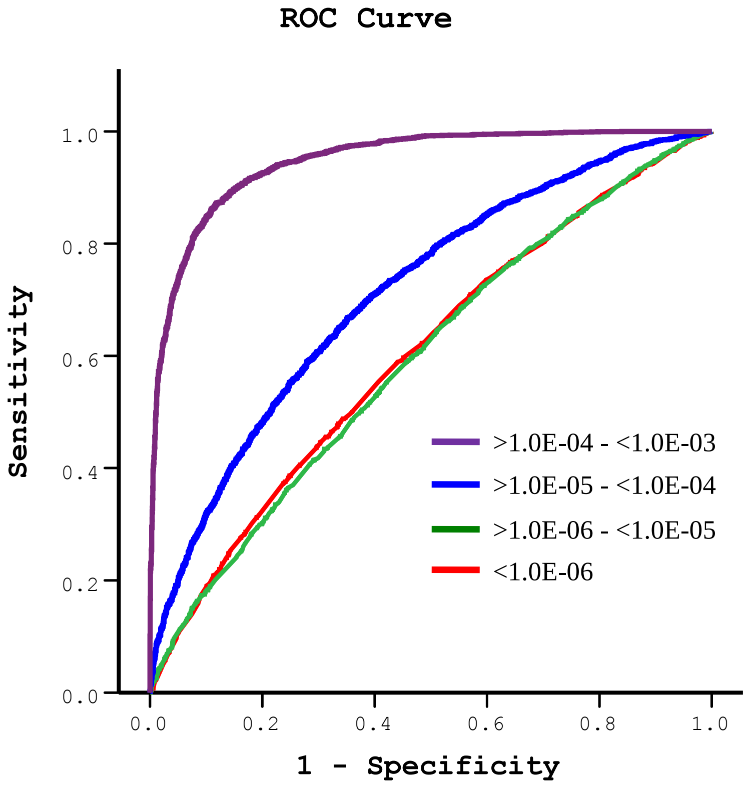


Figure 3.18 ROC Curve of SNP groups depending on P values in spearate mode.

Table 3.15 Additive (incremental) binary logistic regression analysis of SNPs grouped according to their P values

SNP groups according to their P values	Number of SNP (n)	NPV (Percentage correct for control)	PPV (Percentage correct for diabetes)	Overall %	AUC
<1.0E-06	10	75.0	38.7	58.3	0.602
<1.0E-05	27 (10+17)	72.8	45.3	60.2	0.636
<1.0E-04	118 (91+27)	74.3	59.3	67.4	0.735
<1.0E-03	798 (680+118)	90.7	89.1	90.0	0.965

NPV: negative predictive value, PPV: positive predictive value, AUC: area under curve

The summary of the analysis of classification depending on P value is given in Table 3.16. NPV, PPV, overall prediction, and AUC values are shown below. These parameters were analyzed separately for each P value group.

Table 3.16 Individual binary logistic regression analysis of SNPs that grouped according to P values.

SNP groups according to their P values	Number of SNP (n)	NPV (Percentage correct for control)	PPV (Percentage correct for diabetes)	Overall %	AUC
<1.0E-06	10	75.0	38.7	58.3	0.602
>1.0E-06 - <1.0E-05	17	76.0	35.6	57.4	0.595
>1.0E-05 - <1.0E-04	91	73.0	57.2	65.7	0.713
>1.0E-04 - <1.0E-03	680	88.9	86.2	87.7	0.947
All SNPs <1.0E-03	798	90.7	89.1	90.0	0.965

NPV: negative predictive value, PPV: positive predictive value, AUC: area under curve

SNPs that have lower P value are limited, i.e. lower than 1.0E-6 only 10 SNPs exist. However, their overall correctness percentage was 58.1 and AUC was 0.601. On the contrary, there is 604 SNPs which their P value between 1.0E-04 and 1.0E-03 and their correctness percentage was 85.4 and AUC was 0.933. The most important inference from these results is that the SNPs with lower P value than that 5 × 10^{-8} might be important. However, it has been generally accepted that P value lower than 5 × 10^{-8} is important in GWAS studies. Our finding is the contrary of this accepted criterion. Therefore, we should use more SNPs with P value from near the detaching point of line in QQ Plot to obtain more accurate prediction.

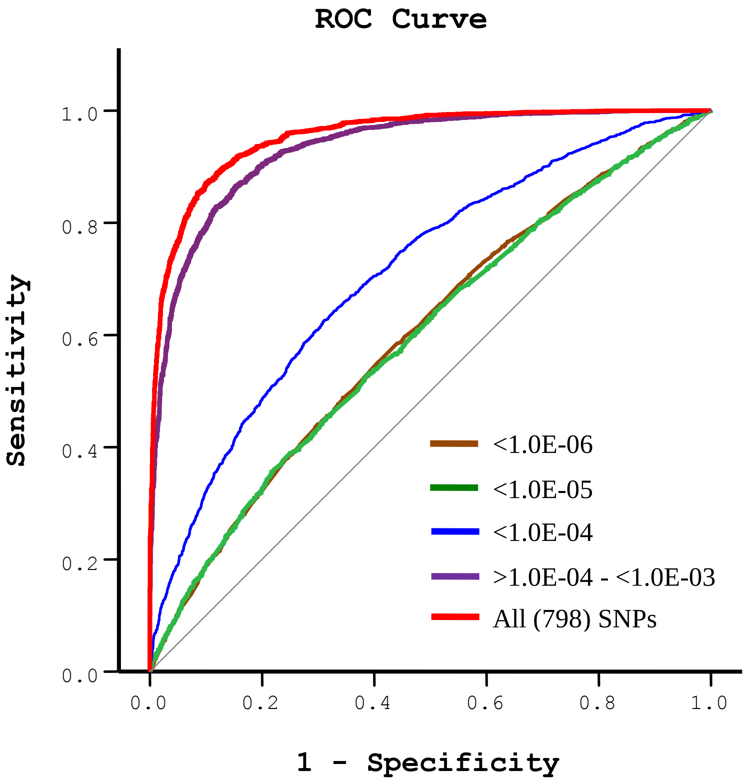


Figure 3.19 ROC Curve of SNP groups depending on P values in additive mode.

3.8.6 Determination of the Most Significant SNPs for the Prediction of Diabetes

3.8.6.1 Modeling with ENTER Method

We used ENTER method and used all SNPs (798 SNP). We chose 193 SNPs with P values less than 0.05 depending on the results of ENTER methods of 798 SNPs. When we analyzed these 193 SNPs only, they yielded the overall 76.6% prediction correctness and an AUC 0.852±0.005 (Table 3.13, Figure 3.21). When we compared to all SNP results (798 SNPs), overall prediction was reduced 13.4%, and AUC was reduced 0.113. Although, less number of SNP might make calculation easy and fast, but we might lose prediction accuracy.

Table 3.17 Classification table of BLR analysis of 193 SNPs [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2422	624	79.5
		Diab	697	1896	73.1
Overall percentage					76.6

a. The cut value is 0.5

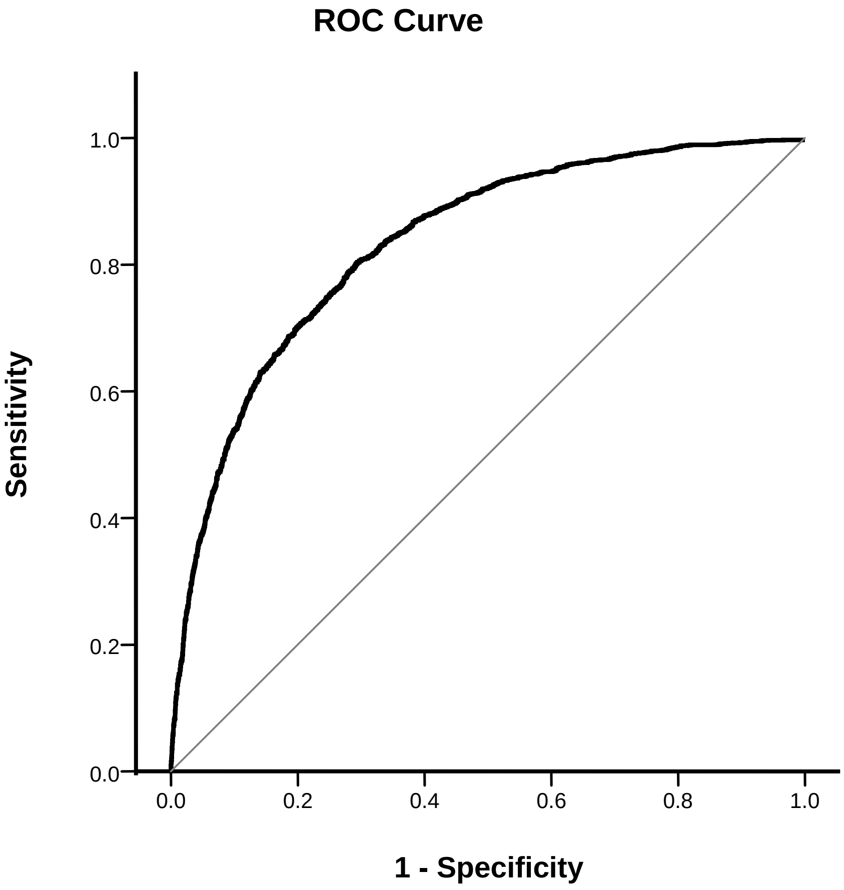


Figure 3.20 ROC Curve of 193 SNPs with P values <0.5 after BLR analysis of 798 SNPs.

3.8.6.2 Modelling with Forward Likelihood Ratio (LR) Method with SNPs Selected from Divided Set of SNPs for BLR Analysis

In another attempt to determine the most important SNPs, which contribute to the prediction accuracy, we chose Forward LR method for BLR. However, forward LR is very time consuming method when variable increased; so, we performed it for each 100 SNP. After elimination of SNPs by forward LR method, 333 SNPs were remained. Then, we analyzed 333 SNPs by using "ENTER" method. The result of this analysis was given below. AUC was 0.917±0.004.

Table 3.18 Classification table of 333 SNPs that are chosen with Forward LR method.[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2596	450	85.2
		Diab	473	2120	81.8
Overall percentage					83.6

a. The cut value is 0.5

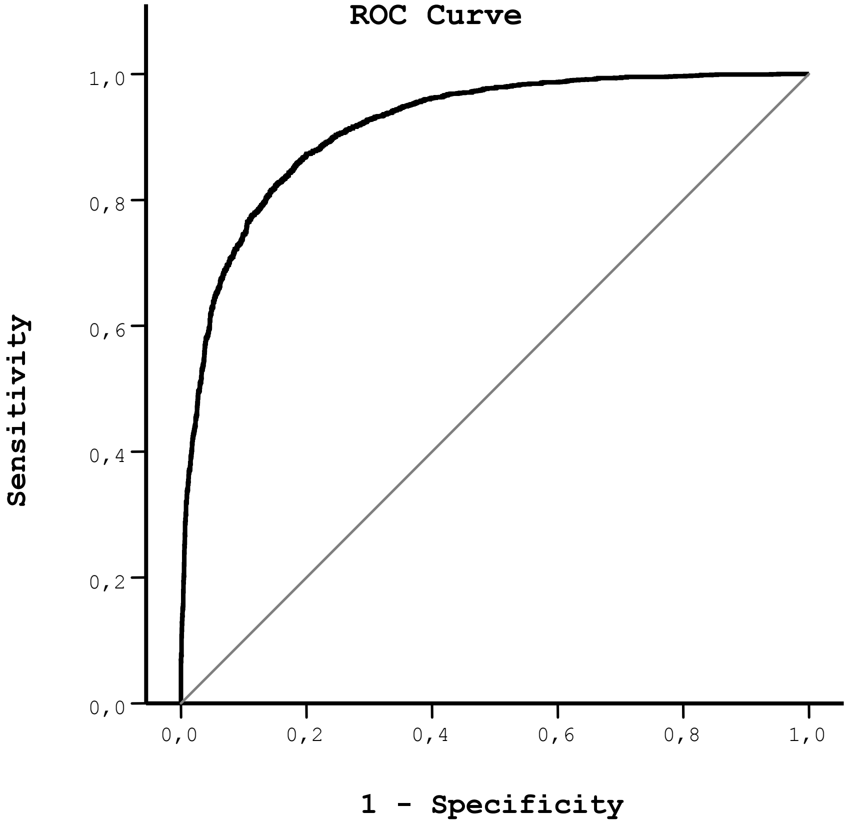


Figure 3.21 ROC curve for 333 SNPs that chosen with Forward LR method.

3.8.6.3 Forward Likelihood Ratio (LR) Method with All SNPs, for BLR Analysis

When we choose Forward LR method, it takes nearly ten days to complete the analysis and 114 SNPs is filtered. AUC was 0.825±0.005 and overall percentage was 74.4% for 114 SNPs. Therefore, both AUC and overall percentage significantly reduced in Forward LR when compared to ENTER method. If we want to estimate more precisely the risk prediction of diabetes ENTER method seems preferable. The other advantage of this method is calculation speed. If we construct SNP database ready for calculation, it takes nearly 20-30 minute to complete analysis. However, it takes nearly ten days with Forward LR with the same dataset. Forward LR method could be preferable if dataset is small and if yields similar results with the ENTER method. However, in our example we should choose the latter. AUC was 0.825±0.005.

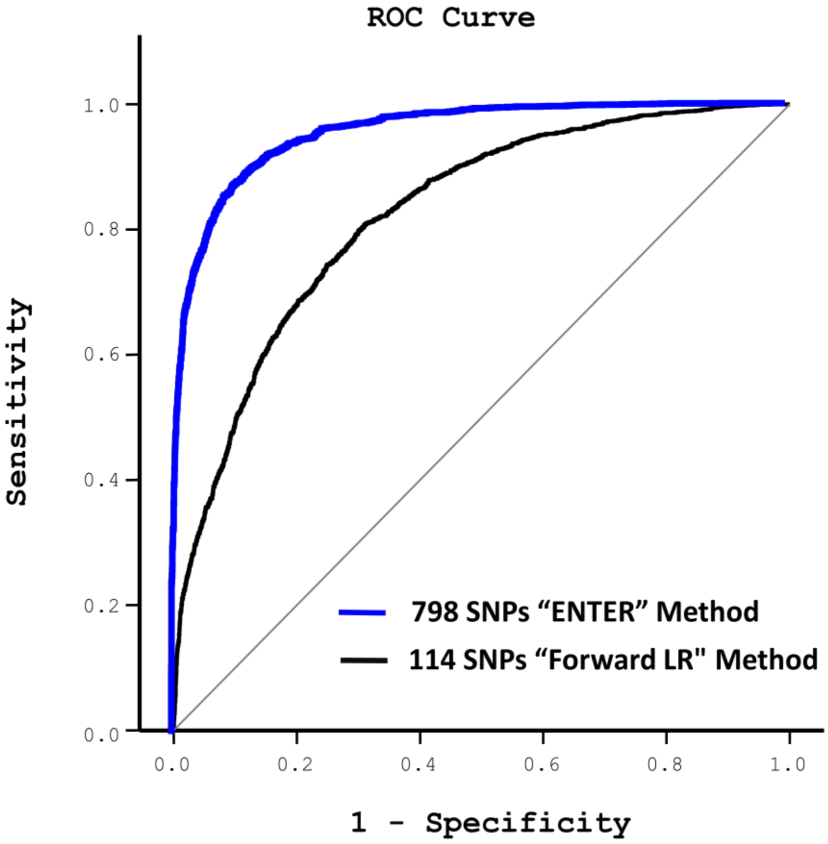


Figure 3.22 ROC curve for 114 SNPs that chosen Forward LR method in a single step, comparison with 798 SNPs.

Table 3.19 Classification table of 114 SNPs that chosen Forward LR method at one step [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2384	662	78.3
		Diab	782	1811	69.8
Overall percentage					74.4

a. The cut value is 0.5

3.8.6.4 SNP Selection Using Population Attributable Risk (PAR)

We also used a different approach by using "population attributable risk (PAR)" method for the selection of the best SNPs for better prediction of diabetes using genotypic data. PAR is the portion of the incidence of a disease in the population (exposed and nonexposed) that is due to exposure. It is the incidence of a disease in the population that would be eliminated if exposure were eliminated. PAR was calculated as described in method section. The summary of binary logistic regression analysis of SNPs depending on their PAR values is given in Table 3.16. The details of binary logistic regression analysis of the population attributable risk is given in Appendix J.

Table 3.20 The results of classification depending on PAR score.

SNP groups according to their PAR values	# SNPs (n)	NPV	PPV	Overall %	AUC
PAR high negative group	179	74.8	62.9	69.3	0.766
PAR lower negative group	179	74.8	67.1	71.3	0.782
PAR higher positive group	181	75.4	64.0	70.2	0.767
PAR low positive group	181	76.0	62.9	70.0	0.772
PAR negative total	358	77.6	71.6	74.8	0.832
PAR positive total	358	81.2	72.6	77.2	0.854
PAR high negative + high positive	360 (179+181)	80.5	74.7	77.8	0.856
PAR low negative + low positive	360 (179+181)	81.6	75.3	78.7	0.869
PAR high negative plus low positive	360 (179+181)	81.2	73.2	77.5	0.860
PAR low negative plus high positive	360 (179+181)	80.5	76.1	78.5	0.865
All SNPs	798	90.7	89.1	90.0	0.965

3.8.7 Effects of Cut-off Value on Prediction Percentage and AUC

We tested how various threshold levels in BLR analysis affect the prediction score and AUC (Table 3.19). Threshold level is chosen as 0.5 by default in BLR analysis. When the threshold level increases, negative predictive value (NPV) increases, positive predictive value (PPV) decreases, and AUC does not change. The details of binary logistic regression analysis of cut-off value is given in Appendix L.

Table 3.21 Summary table of the effects of cut-off value on prediction rate and AUC.

ROC cutoff value	# SNPs (n)	NPV	PPV	Overall %	AUC
0.5	798	90.7	89.1	90.0	0.965 ± 0.002
0.6	798	94.0	83.7	89.3	0.965 ± 0.002
0.7	798	96.8	76.8	87.6	0.965 ± 0.002
0.8	798	98.4	67.5	84.2	0.965 ± 0.002
0.9	798	99.3	52.6	77.8	0.965 ± 0.002

NPV: negative predictive value, PPV: positive predictive value, AUC: area under curve

CHAPTER 4

DISCUSSION

Several studies have investigated the use of risk-SNP markers as a mean of directly improving the accuracy of prognosis. Some have found that the accuracy of prognosis improves [48], while others report only minor benefits from this use [49]. A problem with this direct approach is the small magnitudes of the effects observed. A small effect of individual SNPs ultimately translates into a poor separation of cases and controls and thus reflects only a small improvement to the prognosis accuracy. On the otherhand GWA studies can identify hundreds of SNPs among a million studied, therefore have the potential to reveal SNP profiles associated with diseases for prediction and to elucidate pathophysiology [50].

GWAS has facilitated understanding the genetic basis of complex traits. It is a powerful method to detect genetic variations that predispose to a disease. GWAS provided us many useful insights into the pathophysiology of T2D by identifing novel susceptibility loci that had not been captured by classical approaches. However, for most of the identified T2D susceptibility loci, the causal variants and molecular mechanisms for diabetes risk are unknown. **Our findings do not reject the importance susceptibility loci for causal variants but also provides us the candidate SNP profile for more accurate risk prediction.** It is also important to remember that the effect size found for SNPs thus far could not be a reflection of their biological or clinical significance. Even though their individual predictive values are small, SNPs might point to important biological pathways, which could be targeted for therapeutic intervention.

In this study, we have confirmed several SNPs which were previously found associated with type 2 diabetes. In addition, we have also found several new candidate genes that are potential risk factors for T2D. In addition, we have identified several new candidate SNPs for previously reported and also novel genes associated with T2D.

The prediction of an individual’s risk of developing T2D is the most anticipated clinical use of genetic information. Prediction values of phenotypic and genotypic characters have been investigated in the Malmö Preventive Project (MPP), the Botnia Study [23], the Framingham Offspring Study I [24], Whitehall II study [25] and UK Type 2 Diabetes Genetics Consortium Study [51]. These studies examined loci ranging in number from 11 to 20 that were associated with T2D. The results of these analyses showed no clear improvement in predictive power on adding the genetic risk score to established risk prediction models using phenotypic variables such as age, sex, family history, body mass index, fasting glucose level, systolic blood pressure, and lipid profile. Basic demographic, clinical, and laboratory predictors have C statistics (AUC) ranging from 0.66 in the Rotterdam Study [26] to 0.90 in the Framingham Offspring Study I [24]. The C statistic improves from 0.903 to 0.906 with the addition of a 40-SNP score to the clinical model in the Framingham Offspring Study II [22], and from 0.74 to 0.75 in the larger Malmö Preventive Project [23]. In other studies, adding genetic information to phenotype-based risk models did not improve discrimination and showed a maximum increase of only 2% over phenotype in ROC curves [20, 25, 51]. AUC values were equal to or lower than 0.60 for genetic variants alone in these studies [24-26, 51]. Therefore, phenotype scores were found to be superior to the scores achieved thus far by using genotype alone. On the other hand, the reason for the substantial difference with AUC of phenotype variables amongst the studies, between 0.66 and 0.903, could be attributed to difference in age, case number, familial diabetes history, hypertension rate, BMI level and other variables as indicated in Appendix M.

The lack of clinical impact to date was not surprising of GWAS research since it is in their earlier phase. In order to translate GWAS findings into improved care for patients with diabetes, ongoing research efforts should focus on detailed functional characterization of the identified T2D susceptibility variants and the search for missing heritability. In the Framingham Offspring II study, the addition of a 40-SNP score to a full clinical model achieved better net reclassification improvement (NRI) among those younger than 50 years [22]. However, the degree of prediction scores obtained from genotype is still below the widely accepted clinical prevention target. A higher contribution of genotype over the prediction value of phenotype at a younger age is expected since phenotype variables are more overt only at middle age or older. The most desirable risk prediction method is that with a higher prediction value at an early age, even in childhood. **For the first time in this study, genotype based prediction has shown to yield as performance score as phenotype based for T2D. Here, we showed that genetic risk prediction alone using 798 SNPs yield 90.0% prediction correctness and AUC was 0.965 with only genotype (SNP) variables. This is highest score achieved in the literature for risk prediction of T2D.**

Also another limitation of the use of phenotypic variables is the limited range of ages and follow-up durations for T2D genetic prediction. In previous studies, participants with baseline ages were generally in middle adulthood and the follow-up period was around ten years. However, we need a model that can estimate the risk earlier, which should be validated at a young age with a longer prediction time horizon to help achieve early prevention. As noted above, in the Framingham Offspring Study II, the 40-SNP genotype risk score significantly improved NRI in younger participants but not in older ones. Fortunately, the incidence of T2D can be delayed or prevented by maintaining healthy lifestyle behaviors at early adulthood [28]. The identification of population subgroups at particularly high risk for T2D earlier might facilitate the targeting of prevention efforts to those who might benefit most. **Until this study, the genetic associations identified was not able to improve the T2D risk prediction, the clinical which has already achieved with clinical risk predictors alone. Therefore, our gentoype prediction model also provides an opportunity for risk prediction of T2D with high accuracy at an early stage. Genotype-based risk prediction proposed in this study can be beneficial at early adulthood to determine individuals with higher risk of T2D and to direct them to healthy life-style choices.**

Since the first GWAS data were published in 2007 by WTCCC [52], significant progress has been made and much information has been obtained from GWAS. However, GWAS-based studies to improve clinical decisions are still in their initial stages [53]. Studies have been focused mostly on the causation loci rather than entire risk prediction approach. In addition, the results of the risk prediction are not satisfactory for T2D. Nearly 40 susceptible loci has been identified in European and Asian populations but the entire heritability of T2D remains largely unexplained [54]. Only ~10% of the known T2D heritability could be explained based on the results of a European twin study [55]. This evidence suggests that large portion of heritability is missing. Since a statistical P value of 5×10^{-8} is generally accepted for genome-wide significance [56], previous studies did not use SNPs which has higher P value than that. Several limitations of the current approach for GWAS in revealing the missing heritability information have been proposed. One limitation is the accepted importance threshold level for GWAS ($P< 5 \times 10^{-8}$) which may produce type 2 errors (false-negative results). Therefore, many important loci could be obscured among loci having only borderline associations. In addition, Imamura et al. suggested that the other reason for low the percentage of genetic contribution might be omission of susceptibility variants that have an MAF value of less than 1%. **However, our findings do not agree with these suggestions. In this study, we used SNPs that had p-values greater**

than 5 × 10^{-8} and accepted 5% as the threshold for MAF, and thereby obtained a higher risk prediction score. The most important reason for the low genetic contribution reported so far is likely the use of a small number of SNPs for analysis to yield a sufficient composite risk score. We proposed that SNPs that have p-values less than the detaching point of a distribution (in QQ plot), 1.0E-3 in our study, could contribute to risk prediction. Furthermore, Imamura et al. suggested that genome-wide exon (exome) sequencing by next-generation sequencers might help explain the missing heritability. Our findings suggest that this might not be necessary to obtain a high risk-prediction score. However, next-generation sequencing technology may help find the exact causative loci near or encompassing the newly discovered SNPs.

Because individual SNPs do not yield adequate prediction scores, combining SNPs to yield composite genotype risk scores has also been tested. In such a simulation study by Janssens et al., in which they have studied only 40 SNPs, risk alleles were weighted according to the T2D effect size from the original GWAS; this might not substantially improve the C statistic for alleles with small effects sizes (odds ratio, 1.10–1.25) [57]. **However, we found that 680 SNPs with P values between 1.0E-04 and 1.0E-03 yielded an overall prediction score of 87.7% and AUC of 0.947, while 118 SNPs, with P values less than 1.0E-04, yielded an overall prediction score of 67.4% and AUC of 0.735. This shows that high SNP number is required for higher composite genotype risk scores.** The composite risk score is not equal to the sum of individual SNP scores. Probably, due to the overlapping effect of the risk alleles, we were able to obtain a higher composite risk score when a higher number of SNPs were considered. However, phenotype risk scores are higher than those of individual SNP scores, i.e. OR is 3.86 for BMI in our study; thus, low number of phenotype variables yields higher scores.

Small ratio of events per variable (EPV) can affect the accuracy and precision of regression coefficients. Bigger samples and high number of events are usually preferred. It is usually recommended to study at least ten events per predictor variable for multivariate logistic regression. These rules of thumb for the number of events per variable have primarily been established based on simulation studies for the logistic regression model [58]. Although recent simulation studies suggest as few as five events per predictor variable is sufficient. Vittinghoff et al (2007) found that minimum of ten outcome events per predictor variable (EPV) for logistic model may be too conservative [59]. They indicated that this rule can be relaxed to some extent especially when large populations are being studied. In a small study population, EPV should be higher than 10, but in a large population study it could be relaxed down to five event per variable. They showed that in a large simulation study, EPV a range of circumstances in which coverage and bias were within acceptable levels despite less than 10 EPV. When sample size increases (i.e. >1024), confidence interval coverage increases and five events per variable seems satisfactory. They also found that results for EPV between 5–9 were comparable to those with EPV count of 10–16. **We divided our dataset into two control and validation datasets, 80% and 20% respectively. Our validation set was bigger than 1024 data sets, which is the highest number of groups in Vittinghoffs' study. We have also confirmed that the binary logistic regression analysis of control and validation groups were comparable, as the was not any difference between the results of three sampling of control and validation groups. Therefore, we concluded that five events per predictor variable in our study would be sufficient and would not cause overfitting, and this allowed us to study up to 225 SNP variables at once.**

Due to the low predictive value of the genetic susceptibility loci of T2D so far, alternative GWAS strategies, such as enrichment of genetic effects for improving power (i.e., selecting more severe cases, early onset of disease, and family history of T2D), and original GWAS study designs (such as response to an anti-diabetic treatment or T2D in the presence of extreme obesity) [14, 60] have been proposed. Complementary epigenomic approaches such as DNA methylation studies have also been proposed in addition to GWAS [60]. **However, our strategy of using more SNPs may provide higher risk prediction for T2D; therefore, the need for a sophisticated approach to risk prediction could be reviewed. Our approach might be combined with epigenomic, environmental or other enrichment methods for further insight into T2D etiology.**

CHAPTER 5

CONCLUSION AND FUTURE STUDIES

In conclusion, we have found that genotype-based risk prediction could yield higher risk prediction values when a sufficient number of SNPs are used. This could enable early risk prediction for T2D. The threshold p-value in GWAS analysis to gain importance should be reviewed depending on the investigation field. Our findings open up new horizons for translating GWAS findings into improved care for patients with diabetes. The value of genotype-based risk prediction alone or in combination with phenotypic variables should be further investigated in follow-up studies for validation. Therefore, predictive value of our approach will be the most important usage area for GWAS studies.

Our results bring a new perspective to all GWAS studies. Since the results of GWAS studies for prediction were poor so far, scientists and media were questioning the methods used.

In the future, follow-up studies for a reasonable time period should be designed to evaluate the development of T2D using the genotype-based risk prediction value from our study. We were able to calculate individual risk scores using the constants of the present study obtained with the analysis. Our findings should be validated by comparing cumulative T2D incidence in low- and high-risk groups in a follow-up study. In addition, interethnic differences should be reviewed from the perspective of our results since some GWAS studies did not mention the gender of the participants [61, 62].

Pharmacogenetics is another promising clinical application of the genetic findings for T2D which could allow personalized medicine by facilitating optimal treatment choices that maximize clinical efficacy and minimize toxicity. Our prediction strategy could also be tested for treatment success of T2D via establishing pharmacogenetic investigation of a genome wide approach. In a previous study, it has been found that a SNP rs11212617 at a locus containing the ataxia telangiectasia mutated (ATM) gene could explain 2.5% of variance in metformin response [63]. Genetic background alone is insufficient to predict treatment response at an individual level at that time, accumulation of these pharmacogenetic data is necessary for the future development of personalized medicine. Variance greater than this can probably be explained by the composite SNP score approach. Translation of the findings of the present study will provide a gateway into personalized preventive and therapeutic medicine.

Prenatal screening risk prediction for diabetes and for other studies will be possible with results that are more accurate.

In conclusion, hope with the expected benefits above, we should take care that the value of genotype based risk prediction using our approach should be further investigated in follow-up studies for validation.

REFERENCES

[1] http://diabetes.niddk.nih.gov/index.aspx

[2] Alberti G, Zimmet P, Shaw J, et al. (2004) Type 2 diabetes in the young: the evolving epidemic: the international diabetes federation consensus workshop. Diabetes care 27: 1798-1811

[3] Shaw JE, Sicree RA, Zimmet PZ (2010) Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes research and clinical practice 87: 4-14

[4] Punnose J, Agarwal MM, Bin-Uthman S (2005) Type 2 diabetes mellitus among children and adolescents in Al-Ain: a case series. Eastern Mediterranean health journal = La revue de sante de la Mediterranee orientale = al-Majallah al-sihhiyah li-sharq al-mutawassit 11: 788-797

[5] Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H (1999) Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia 42: 139-145

[6] Florez JC (2008) Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. The Journal of clinical endocrinology and metabolism 93: 4633-4642

[7] Vaxillaire M, Veslot J, Dina C, et al. (2008) Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. Diabetes 57: 244-254

[8] Fajans SS, Bell GI, Polonsky KS (2001) Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. The New England journal of medicine 345: 971-980

[9] Magee MJ, Narayan KM (2013) Global confluence of infectious and non-communicable diseases - The case of type 2 diabetes. Preventive medicine 57: 149-151

[10] http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000091.v2.p1

[11] Vaxillaire M, Froguel P (2008) Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. Endocrine reviews 29: 254-264

[12] International HapMap C (2005) A haplotype map of the human genome. Nature 437: 1299-1320

[13] International HapMap C, Frazer KA, Ballinger DG, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851-861

[14] Wheeler E, Barroso I (2011) Genome-wide association studies and type 2 diabetes. Briefings in functional genomics 10: 52-60

[15] Qi L, Cornelis MC, Kraft P, et al. (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Human molecular genetics 19: 2706-2715

[16] Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM, Kliewer SA (1995) An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). The Journal of biological chemistry 270: 12953-12956

[17] Sagen JV, Raeder H, Hathout E, et al. (2004) Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. Diabetes 53: 2713-2718

[18] Riserus U, Arnlov J, Berglund L (2007) Long-term predictors of insulin resistance: role of lifestyle and metabolic factors in middle-aged men. Diabetes care 30: 2928-2933

[19] Riserus U, Willett WC, Hu FB (2009) Dietary fats and prevention of type 2 diabetes. Progress in lipid research 48: 44-51

[20] Balkau B, Lange C, Fezeu L, et al. (2008) Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). Diabetes care 31: 2056-2061

[21] Muhlenbruch K, Jeppesen C, Joost HG, Boeing H, Schulze MB (2013) The value of genetic information for diabetes risk prediction - differences according to sex, age, family history and obesity. PloS one 8: e64307

[22] de Miguel-Yanes JM, Shrader P, Pencina MJ, et al. (2011) Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. Diabetes care 34: 121-125

[23] Lyssenko V, Jonsson A, Almgren P, et al. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. The New England journal of medicine 359: 2220-2232

[24] Meigs JB, Shrader P, Sullivan LM, et al. (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. The New England journal of medicine 359: 2208-2219

[25] Talmud PJ, Hingorani AD, Cooper JA, et al. (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. Bmj 340: b4838

[26] van Hoek M, Dehghan A, Witteman JC, et al. (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 57: 3122-3128

[27] Vassy JL, Meigs JB (2012) Is genetic testing useful to predict type 2 diabetes? Best practice & research Clinical endocrinology & metabolism 26: 189-201

[28] Tuomilehto J, Lindstrom J, Eriksson JG, et al. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. The New England journal of medicine 344: 1343-1350

[29] Billings LK, Florez JC (2010) The genetics of type 2 diabetes: what have we learned from GWAS? Annals of the New York Academy of Sciences 1212: 59-77

[30] Jing C, Xueyao H, Linong J (2012) Meta-analysis of association studies between five candidate genes and type 2 diabetes in Chinese Han population. Endocrine 42: 307-320

[31] Kang HP, Yang X, Chen R, et al. (2012) Integration of disease-specific single nucleotide polymorphisms, expression quantitative trait loci and coexpression networks reveal novel candidate genes for type 2 diabetes. Diabetologia 55: 2205-2213

[32] Shai I, Jiang R, Manson JE, et al. (2006) Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. Diabetes care 29: 1585-1590

[33] Tang Y, Han X, Sun X, et al. (2013) Association study of a common variant near IRS1 with type 2 diabetes mellitus in Chinese Han population. Endocrine 43: 84-91

[34] Ustunkar G, Aydin Son Y (2011) METU-SNP: an integrated software system for SNP-complex disease association analysis. Journal of integrative bioinformatics 8: 187

[35] Peng G, Luo L, Siu H, et al. (2010) Gene and pathway-based second-wave analysis of genome-wide association studies. European journal of human genetics : EJHG 18: 111-117

[36] Purcell S, Neale B, Todd-Brown K, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 81: 559-575

[37] Blankers M, Koeter MW, Schippers GM (2010) Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. Journal of medical Internet research 12: e54

[38] Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29-36

[39] Bewick V, Cheek L, Ball J (2003) Statistics review 7: Correlation and regression. Critical care 7: 451-459

[40] Altshuler D, Hirschhorn JN, Klannemark M, et al. (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nature genetics 26: 76-80

[41] Wang H, Hays NP, Das SK, et al. (2009) Phenotypic and molecular evaluation of a chromosome 1q region with linkage and association to type 2 diabetes in humans. The Journal of clinical endocrinology and metabolism 94: 1401-1408

[42] Prokopenko I, Zeggini E, Hanson RL, et al. (2009) Linkage disequilibrium mapping of the replicated type 2 diabetes linkage signal on chromosome 1q. Diabetes 58: 1704-1709

[43] Cornelis MC, Qi L, Kraft P, Hu FB (2009) TCF7L2, dietary carbohydrate, and risk of type 2 diabetes in US women. The American journal of clinical nutrition 89: 1256-1262

[44] Rees SD, Bellary S, Britten AC, et al. (2008) Common variants of the TCF7L2 gene are associated with increased risk of type 2 diabetes mellitus in a UK-resident South Asian population. BMC medical genetics 9: 8

[45] Chandak GR, Janipalli CS, Bhaskar S, et al. (2007) Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. Diabetologia 50: 63-67

[46] Sanghera DK, Ortega L, Han S, et al. (2008) Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. BMC medical genetics 9: 59

[47] Sanghera DK, Nath SK, Ortega L, et al. (2008) TCF7L2 polymorphisms are associated with type 2 diabetes in Khatri Sikhs from North India: genetic variation affects lipid levels. Annals of human genetics 72: 499-509

[48] Muehlschlegel JD, Liu KY, Perry TE, et al. (2010) Chromosome 9p21 variant predicts mortality after coronary artery bypass graft surgery. Circulation 122: S60-65

[49] Paynter NP, Chasman DI, Pare G, et al. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. JAMA : the journal of the American Medical Association 303: 631-637

[50] Couzin-Frankel J (2010) Major heart disease genes prove elusive. Science 328: 1220-1221

[51] Lango H, Consortium UKTDG, Palmer CN, et al. (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 57: 3129-3135

[52] Wellcome Trust Case Control C (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678

[53] Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470: 187-197

[54] Imamura M, Maeda S (2011) Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]. Endocrine journal 58: 723-739

[55] Voight BF, Scott LJ, Steinthorsdottir V, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nature genetics 42: 579-589

[56] McCarthy MI, Abecasis GR, Cardon LR, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature reviews Genetics 9: 356-369

[57] Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ (2007) The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. Genetics in medicine : official journal of the American College of Medical Genetics 9: 528-535

[58] Concato J, Peduzzi P, Holford TR, Feinstein AR (1995) Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. Journal of clinical epidemiology 48: 1495-1501

[59] Vittinghoff E, McCulloch CE (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. American journal of epidemiology 165: 710-718

[60] Bell CG, Finer S, Lindgren CM, et al. (2010) Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PloS one 5: e14040

[61] Zeggini E, Weedon MN, Lindgren CM, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316: 1336-1341

[62] Yamauchi T, Hara K, Maeda S, et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nature genetics 42: 864-868

[63] GoDarts, Group UDPS, Wellcome Trust Case Control C, et al. (2011) Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nature genetics 43: 117-120

APPENDICES

APPENDIX A: THE DETAILS OF THE IMPUTATION WITH AMELIA

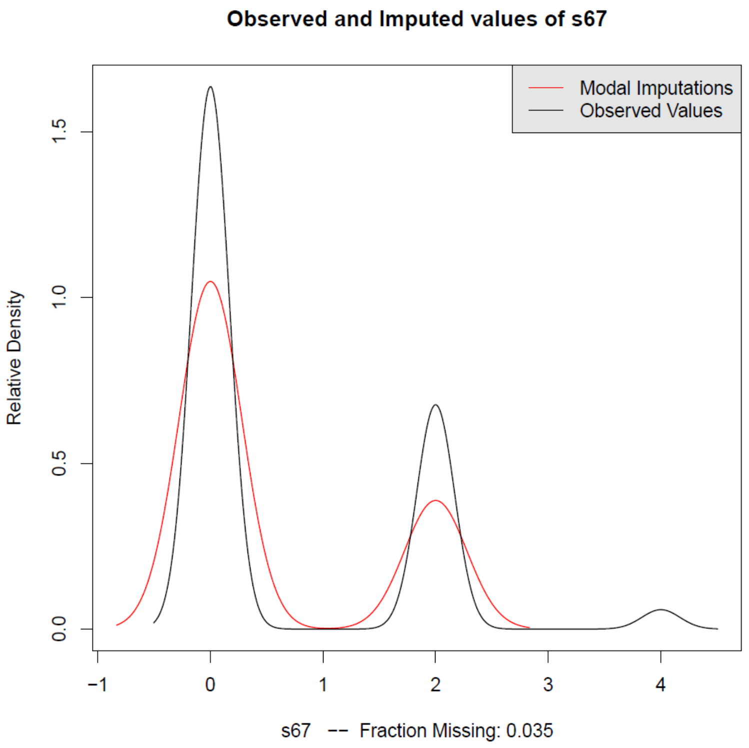


Figure 1 Comparison of the relative density distribution of filling alleles with the original using Amelia Toolbox. The imputed alleles are similar to originals in a proportional level. Here we transformed allele information as nominal.

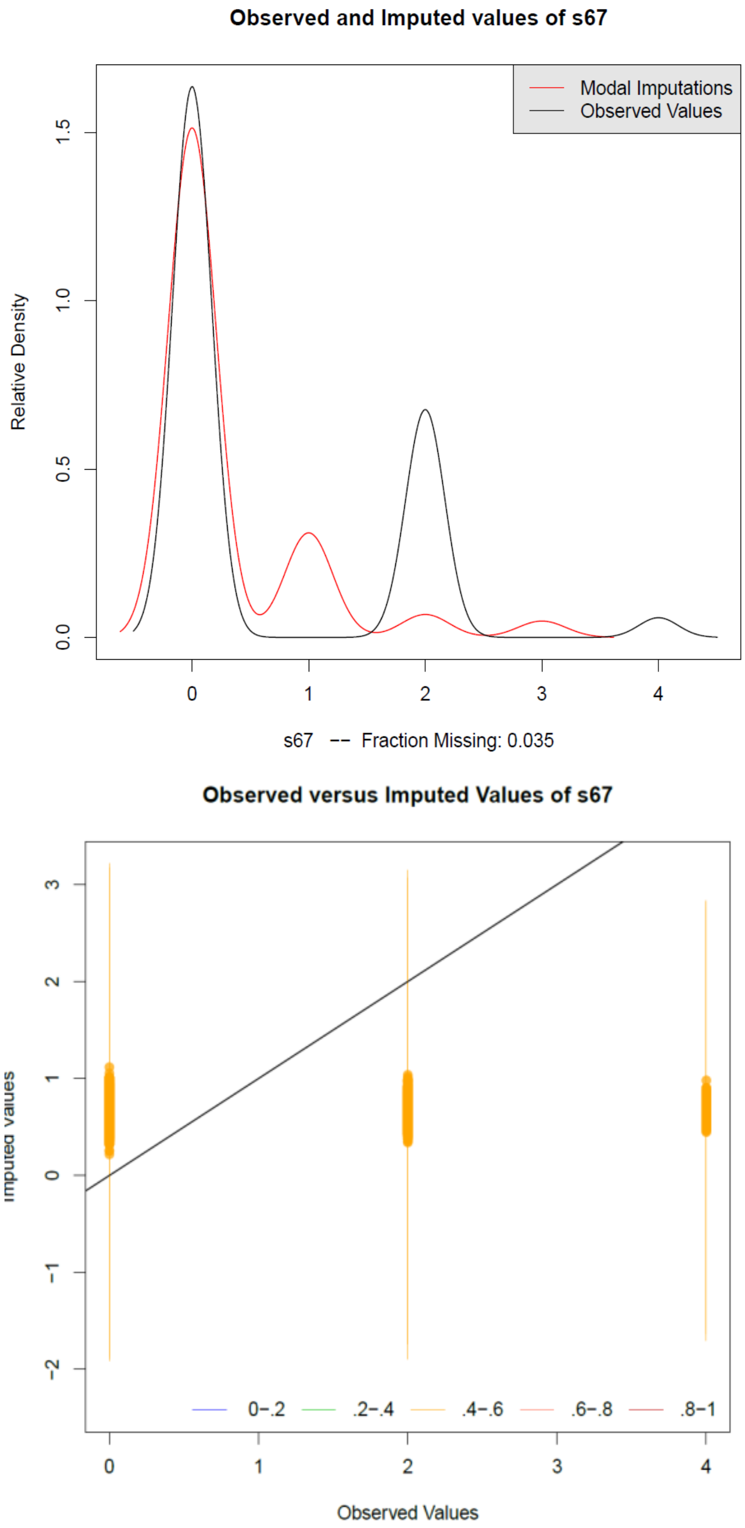


Figure 2.a,b Relative density distribution of imputed alleles when alleles coded as ordinal value. The imputed alleles are not similar to originals. Here we transformed allele information as ordinal and Amelia handled it as a numerical value, so distribution density is not similar with the original. In addition, some of the imputed data is not in the range of confidence interval. Therefore, we used nominal transformation for allele.

APPENDIX B: CHROMOSOMES, P VALUES, ODDS RATIOS, START BASE PAIR, MAJOR/MINOR ALLELE, MAF VALUES, AND MAPPED GENES OF 798 SNPS

#	rsid	OR	P value	MAF	CHR	BP	A1	A2	Entrez Gene	Gene Symbol	Gene Name
1	rs4654582	0,85	5,28E-04	0,213	1	4630143	T	A	55966	AJAP1	adherens junctions associated protein 1
2	rs11121467	0,79	2,34E-04	0,105	1	9620920	A	T			
3	rs2336381	0,76	9,28E-04	0,055	1	11931611	G	A	5351	PLOD1	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1
4	rs11580525	1,23	5,35E-04	0,114	1	14119518	C	T			
5	rs149562	1,15	8,72E-04	0,259	1	16667788	T	C	114819	CROCCP3	ciliary rootlet coiled-coil, rootletin pseudogene 3
6	rs6660946	0,86	7,63E-04	0,239	1	18606142	G	A			
7	rs7529705	1,14	8,30E-04	0,380	1	19592679	A	G	832	CAPZB	capping protein (actin filament) muscle Z-line, beta
8	rs10492998	0,86	8,99E-04	0,219	1	19645434	T	C	832	CAPZB	capping protein (actin filament) muscle Z-line, beta
10	rs6701048	1,24	7,70E-04	0,097	1	29676041	G	C			
11	rs6704040	1,29	7,84E-04	0,065	1	30417261	C	T			
12	rs215770	1,17	3,04E-04	0,268	1	37358560	A	C			
13	rs215773	0,87	2,84E-04	0,369	1	37368827	T	G			
14	rs215792	0,87	6,10E-04	0,376	1	37378028	C	T			
15	rs215791	0,88	8,13E-04	0,375	1	37378878	C	T			
16	rs12131641	0,84	2,90E-04	0,198	1	37384100	A	G			
18	rs1587578	0,85	2,23E-04	0,254	1	37401328	C	A			
19	rs11579242	0,85	9,45E-04	0,197	1	47987966	G	A			
20	rs11584807	0,83	1,28E-04	0,187	1	47993041	T	C			
21	rs783323	0,87	3,10E-04	0,468	1	66713368	A	G			
22	rs699253	0,85	2,96E-05	0,476	1	66713736	A	G			
23	rs12739235	1,23	3,14E-04	0,119	1	66728087	C	T			
25	rs7537440	1,14	5,02E-04	0,426	1	66804495	G	T	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
26	rs1373909	1,14	5,08E-04	0,426	1	66813463	G	A	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
27	rs6697088	0,86	8,66E-05	0,404	1	66817312	C	G	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
28	rs10889634	0,87	3,05E-04	0,414	1	66838499	G	A	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
29	rs6696927	0,87	2,89E-04	0,414	1	66842969	T	C	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
30	rs1562217	0,87	2,24E-04	0,414	1	66846154	T	C	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
32	rs4655648	0,87	3,51E-04	0,415	1	66886897	C	T	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
33	rs9662943	0,88	5,82E-04	0,407	1	66893720	C	T	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
34	rs6681460	0,87	4,57E-04	0,415	1	66895645	A	G	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
35	rs6694782	1,14	8,31E-04	0,457	1	66899350	G	A	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
36	rs6588215	0,87	2,56E-04	0,414	1	66914537	A	G	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
38	rs7542924	0,87	2,78E-04	0,414	1	66915643	G	A	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1

40	rs10789215	0,88	5,22E-04	0,413	1	66923773	T	C	84251	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
41	rs344935	0,87	4,06E-04	0,321	1	67910451	G	A			
42	rs1780731	1,15	9,22E-04	0,274	1	79108612	C	T			
43	rs1434431	1,14	5,85E-04	0,477	1	87960918	A	G			
44	rs2143992	1,14	5,19E-04	0,437	1	94109636	C	T	30836	DNTTIP2	deoxynucleotidyltransferase, terminal, interacting protein 2
45	rs3789439	0,86	8,37E-04	0,220	1	94352014	C	T	24	ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4
46	rs3789442	0,85	4,95E-04	0,222	1	94354044	C	G	24	ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4
47	rs2220760	1,14	9,66E-04	0,372	1	94977931	A	G			
48	rs3767273	1,15	2,96E-04	0,417	1	103173621	G	C	1301	COL11A1	collagen, type XI, alpha 1
49	rs12046389	1,15	3,45E-04	0,416	1	103181688	A	C	1301	COL11A1	collagen, type XI, alpha 1
51	rs7550118	1,14	8,33E-04	0,357	1	103335690	T	C	1301	COL11A1	collagen, type XI, alpha 1
52	rs1415359	1,14	9,58E-04	0,359	1	103337029	C	T	1301	COL11A1	collagen, type XI, alpha 1
53	rs10493988	1,14	8,36E-04	0,356	1	103338744	G	A	1301	COL11A1	collagen, type XI, alpha 1
54	rs2761441	0,88	6,23E-04	0,493	1	110538385	G	A	388662	SLC6A17	solute carrier family 6, member 17
55	rs1942216	0,86	7,95E-04	0,262	1	115721322	A	C			
56	rs1543594	0,84	2,22E-04	0,204	1	115845877	A	C			
57	rs11579824	0,80	5,29E-06	0,190	1	145469224	C	T			
58	rs12133943	1,24	4,74E-04	0,106	1	145561705	G	C	607	BCL9	B-cell CLL/lymphoma 9
59	rs1208517	0,84	9,76E-04	0,155	1	183539106	T	C	10625	IVNS1ABP	influenza virus NS1A binding protein
60	rs7539680	1,23	2,53E-04	0,123	1	186584179	G	C			
61	rs10753046	1,25	8,65E-05	0,124	1	186631148	G	C			
62	rs6425178	1,25	1,18E-05	0,164	1	186632905	C	G			
64	rs10753049	1,25	1,12E-05	0,165	1	186639485	A	G			
65	rs7516670	1,24	1,78E-05	0,164	1	186642303	T	C			
66	rs6667131	1,25	1,04E-05	0,164	1	186649073	T	A			
68	rs172235	1,17	1,22E-04	0,296	1	186726999	C	A			
69	rs4313401	1,14	7,69E-04	0,499	1	187650996	A	G			
70	rs11800563	0,88	8,25E-04	0,496	1	187698667	G	C			
71	rs4428892	0,88	7,86E-04	0,495	1	187719770	T	A			
72	rs10922227	0,88	8,96E-04	0,486	1	187787128	A	G			
73	rs1119030	0,88	9,94E-04	0,485	1	187787822	A	G			
75	rs2250509	0,82	7,20E-04	0,129	1	201405593	A	G	4608	MYBPH	myosin binding protein H
76	rs340835	1,14	6,34E-04	0,473	1	212230298	A	G	5629	PROX1	prospero homeobox 1
77	rs2820444	0,87	6,80E-04	0,276	1	217808443	A	G			
78	rs3002142	0,83	8,38E-04	0,129	1	220854685	C	T			
79	rs2133189	0,86	2,78E-04	0,283	1	220881065	C	T	375056	MIA3	melanoma inhibitory activity family, member 3
81	rs17465637	0,86	5,02E-04	0,281	1	220890152	A	C	375056	MIA3	melanoma inhibitory activity family, member 3
82	rs1053316	0,81	4,82E-04	0,117	1	220906461	A	G	375056	MIA3	melanoma inhibitory activity family, member 3
83	rs2378607	0,86	1,50E-04	0,315	1	220986518	T	G	400823	FAM177B	family with sequence similarity 177, member B
84	rs6429366	0,87	3,87E-04	0,426	1	240833628	T	C			
85	rs2362255	0,81	3,66E-04	0,126	1	244130482	G	A	64754	SMYD3	SET and MYND domain containing 3
86	rs7520116	0,87	7,19E-04	0,329	1	244271398	G	C	64754	SMYD3	SET and MYND domain containing 3
87	rs3893111	0,87	4,40E-04	0,302	2	8692795	G	A			
88	rs1550105	0,88	7,11E-04	0,444	2	20613584	T	C			
89	rs11897611	0,84	6,05E-04	0,161	2	20638798	C	T			

90	rs4666430	0,83	2,33E-04	0,173	2	20641940	G	A			
91	rs930760	0,86	8,79E-05	0,355	2	20669817	C	T			
92	rs4666438	1,16	7,02E-04	0,264	2	20674067	A	G			
93	rs11096680	1,15	9,87E-04	0,270	2	20675712	A	T			
94	rs3796064	1,16	5,39E-04	0,260	2	20701799	A	G	64342	HS1BP3	HCLS1 binding protein 3
95	rs10166174	0,87	3,97E-04	0,349	2	20702484	A	G	64342	HS1BP3	HCLS1 binding protein 3
96	rs17803553	0,87	6,86E-04	0,356	2	25678607	T	C	1838	DTNB	dystrobrevin, beta
97	rs12613835	0,87	6,66E-04	0,356	2	25682705	A	G	1838	DTNB	dystrobrevin, beta
98	rs7562790	1,15	3,94E-04	0,399	2	36527059	G	T	51232	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)
99	rs2160367	1,15	3,09E-04	0,429	2	36535123	G	C	51232	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)
100	rs3821153	1,14	6,86E-04	0,417	2	36606626	G	T	51232	CRIM1	cysteine rich transmembrane BMP regulator 1 (chordin-like)
101	rs2727880	1,14	5,83E-04	0,429	2	52408156	C	T			
102	rs17730780	0,86	4,53E-04	0,236	2	52416883	G	A			
103	rs6545274	0,85	3,46E-04	0,233	2	52497718	C	T			
104	rs2552356	0,87	3,56E-04	0,459	2	52508248	G	A			
105	rs12622811	0,86	2,44E-04	0,303	2	52641453	T	C			
106	rs6720390	1,14	6,06E-04	0,467	2	52654578	C	T			
107	rs13430296	0,85	8,30E-05	0,313	2	52672168	G	C			
108	rs17043120	0,86	1,70E-04	0,313	2	52679905	G	A			
109	rs1843032	1,14	8,27E-04	0,396	2	52694816	A	G			
110	rs1446441	0,83	2,42E-04	0,170	2	53155170	T	C			
111	rs7575107	1,23	1,94E-04	0,133	2	55159490	G	T			
112	rs4672367	0,83	2,06E-04	0,176	2	60251920	T	C			
113	rs17329726	1,23	2,03E-04	0,129	2	60338590	A	G			
114	rs359274	1,18	9,55E-04	0,178	2	60360385	C	G			
115	rs17662176	0,74	1,65E-04	0,059	2	64950508	G	C			
116	rs12470994	1,29	3,02E-04	0,082	2	67528010	A	C			
117	rs1159766	1,15	8,91E-04	0,273	2	72317749	T	C	23233	EXOC6B	exocyst complex component 6B
118	rs1159764	1,15	9,47E-04	0,273	2	72317874	A	T	23233	EXOC6B	exocyst complex component 6B
119	rs10221769	1,16	5,05E-04	0,276	2	72332562	T	A	23233	EXOC6B	exocyst complex component 6B
120	rs2118836	1,17	2,46E-04	0,292	2	96526699	C	T			
121	rs11123406	1,15	4,66E-04	0,365	2	111667012	T	C			
122	rs17715688	0,80	2,28E-04	0,113	2	115089550	G	T	57628	DPP10	dipeptidyl-peptidase 10 (non-functional)
123	rs17715867	0,77	2,45E-04	0,078	2	115096853	C	A	57628	DPP10	dipeptidyl-peptidase 10 (non-functional)
125	rs17010780	0,81	4,87E-04	0,113	2	124531274	G	T	129684	CNTNAP5	contactin associated protein-like 5
127	rs4954045	0,88	9,06E-04	0,390	2	133695340	A	C	344148	NCKAP5	NCK-associated protein 5
128	rs17786300	1,19	9,41E-04	0,149	2	140253872	C	A			
129	rs1355421	0,79	4,95E-05	0,118	2	160621464	A	G	22925	PLA2R1	phospholipase A2 receptor 1, 180kDa
130	rs1355420	0,80	1,12E-04	0,119	2	160621517	T	C	22925	PLA2R1	phospholipase A2 receptor 1, 180kDa
131	rs4665146	0,80	5,34E-05	0,147	2	160624329	A	C	22925	PLA2R1	phospholipase A2 receptor 1, 180kDa
132	rs16844742	0,79	1,94E-05	0,148	2	160639530	T	A			
133	rs7573469	0,79	1,27E-05	0,149	2	160653973	G	A			
134	rs3111397	0,82	2,58E-05	0,204	2	160759609	C	T	3694	ITGB6	integrin, beta 6
135	rs12692585	1,16	4,91E-04	0,254	2	160789087	G	A			
136	rs10181181	0,81	4,03E-07	0,290	2	160795657	T	C			
137	rs2925757	0,79	1,71E-06	0,183	2	160809415	G	A			
139	rs12692588	0,85	2,43E-05	0,435	2	160832428	C	T			

140	rs7572970	0,83	5,97E-06	0,281	2	160844902	A	G	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
141	rs1020731	0,81	2,45E-07	0,293	2	160852301	G	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
142	rs1020732	0,86	4,42E-05	0,422	2	160852485	G	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
143	rs12692590	0,86	9,21E-05	0,419	2	160861443	C	G	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
144	rs12692592	0,81	5,95E-06	0,221	2	160871627	G	T	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
145	rs9917155	0,86	5,20E-05	0,454	2	160871805	C	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
146	rs4077463	0,81	3,16E-06	0,218	2	160874480	A	G	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
147	rs7593730	0,81	2,55E-06	0,218	2	160879700	T	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
148	rs4589705	0,81	2,75E-06	0,219	2	160884382	T	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
149	rs4386280	0,86	7,99E-05	0,449	2	160891041	A	G	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
150	rs4664013	0,83	6,49E-06	0,331	2	160892410	G	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
151	rs10165319	0,86	1,41E-04	0,337	2	160901051	T	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
152	rs4538150	0,85	2,18E-05	0,451	2	160917573	G	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
153	rs9287795	0,81	2,66E-06	0,218	2	160918034	C	G	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
154	rs6718526	0,78	2,74E-07	0,197	2	160922421	T	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
155	rs11693602	0,80	2,29E-06	0,219	2	160932904	C	T	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
156	rs10929982	0,80	4,55E-06	0,195	2	160944523	C	T	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
157	rs12998587	0,83	1,19E-05	0,307	2	160950541	T	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
158	rs7587102	0,84	1,99E-05	0,306	2	160967528	T	C	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
159	rs4664323	0,87	3,11E-04	0,428	2	160967931	C	T	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
160	rs13009374	0,85	5,84E-05	0,305	2	160973345	C	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
161	rs6742799	0,84	2,39E-04	0,198	2	161025706	C	A	5937	RBMS1	RNA binding motif, single stranded interacting protein 1
162	rs6752569	1,15	4,89E-04	0,327	2	161182219	C	T			
163	rs13390172	1,17	1,69E-04	0,287	2	161233847	C	T			
164	rs12473293	1,18	6,70E-05	0,287	2	161237591	C	A			
165	rs4383351	1,17	1,35E-04	0,286	2	161242414	A	G			
166	rs4368343	1,20	4,39E-06	0,353	2	161242897	C	G			
167	rs16851382	1,21	1,55E-04	0,169	2	166621721	A	G	6323	SCN1A	sodium channel, voltage-gated, type I, alpha subunit
168	rs1402108	0,86	6,78E-04	0,257	2	176957972	G	T			
169	rs12185628	1,21	4,01E-05	0,219	2	179389216	C	T			
170	rs10190741	0,88	5,27E-04	0,443	2	179396117	T	C			
172	rs10176147	1,14	6,24E-04	0,466	2	184378789	G	C			
173	rs826186	0,88	7,66E-04	0,397	2	184403897	G	A			
174	rs2369202	1,13	9,11E-04	0,468	2	184694310	T	C			
175	rs12232884	1,14	8,70E-04	0,454	2	184709630	G	C			
176	rs1526212	1,14	8,54E-04	0,460	2	184719102	A	G			
177	rs10497643	1,14	6,98E-04	0,462	2	184761027	T	C			

178	rs13010985	1,15	2,49E-04	0,458	2	184812923	A	G			
179	rs719736	1,14	7,00E-04	0,487	2	184895389	G	A			
180	rs4241279	1,24	3,06E-04	0,117	2	192317911	T	C			
181	rs6739080	1,22	8,61E-04	0,119	2	192322352	T	G			
182	rs4675425	0,82	6,52E-04	0,124	2	204734173	A	G			
183	rs7583852	0,85	4,74E-04	0,214	2	204766132	T	G			
184	rs10198084	0,84	4,67E-05	0,297	2	204855576	A	G			
185	rs6435252	1,23	4,89E-04	0,116	2	205366261	A	G	117583	PARD3B	par-3 partitioning defective 3 homolog B (C. elegans)
187	rs2663891	1,27	9,40E-04	0,078	2	208281566	A	G			
188	rs16840004	1,30	8,31E-04	0,064	2	208325193	A	G	151195	CCNYL1	cyclin Y-like 1
189	rs7585736	1,17	7,85E-04	0,213	2	214300694	T	G	79582	SPAG16	sperm associated antigen 16
190	rs4673054	0,86	1,25E-04	0,487	2	223796106	A	T			
191	rs2203733	0,86	8,41E-05	0,484	2	223801345	A	G			
192	rs10933000	0,87	1,37E-04	0,488	2	223801654	G	A			
193	rs969494	0,86	1,07E-04	0,484	2	223803302	G	A			
194	rs970816	0,86	7,28E-05	0,481	2	223805584	G	A			
195	rs7595029	1,22	7,62E-05	0,168	2	236056702	C	T			
196	rs4663596	1,20	2,29E-04	0,167	2	236065943	A	G			
197	rs4685598	1,18	7,98E-04	0,191	3	348693	A	C	10752	CHL1	cell adhesion molecule with homology to L1CAM (close homolog of L1)
198	rs7630509	1,17	8,62E-04	0,195	3	349168	G	A	10752	CHL1	cell adhesion molecule with homology to L1CAM (close homolog of L1)
199	rs7649544	1,24	8,03E-04	0,092	3	353069	C	A	10752	CHL1	cell adhesion molecule with homology to L1CAM (close homolog of L1)
200	rs6442929	1,15	7,55E-04	0,308	3	5072993	T	C			
201	rs6773179	1,15	6,47E-04	0,327	3	5073759	A	T			
202	rs1161171	0,83	6,51E-05	0,225	3	8417494	C	T	100288428	LOC100288428	uncharacterized LOC100288428
203	rs359025	0,83	8,15E-05	0,221	3	8420729	T	C	100288428	LOC100288428	uncharacterized LOC100288428
204	rs359024	0,84	2,13E-04	0,224	3	8421265	G	A	100288428	LOC100288428	uncharacterized LOC100288428
205	rs359033	0,86	9,99E-04	0,227	3	8431789	A	G	100288428	LOC100288428	uncharacterized LOC100288428
206	rs359032	0,85	5,74E-04	0,232	3	8432379	C	T	100288428	LOC100288428	uncharacterized LOC100288428
207	rs2088620	0,85	4,58E-04	0,223	3	8435932	G	T	100288428	LOC100288428	uncharacterized LOC100288428
208	rs11712016	1,21	7,15E-04	0,134	3	9174613	G	C	9901	SRGAP3	SLIT-ROBO Rho GTPase activating protein 3
209	rs12185978	0,86	3,98E-04	0,272	3	11061367	C	G			
210	rs2130505	0,84	1,43E-05	0,368	3	21727970	G	A	79750	ZNF385D	zinc finger protein 385D
211	rs4858348	0,86	7,59E-05	0,358	3	21730685	G	A	79750	ZNF385D	zinc finger protein 385D
212	rs4858352	0,85	4,51E-05	0,374	3	21743250	G	A	79750	ZNF385D	zinc finger protein 385D
213	rs9830825	1,28	3,53E-04	0,083	3	31431027	A	C			
214	rs12485914	1,21	7,85E-04	0,133	3	31437904	C	T			
215	rs11917010	0,86	7,17E-04	0,259	3	54181107	A	G	55799	CACNA2D3	calcium channel, voltage-dependent, alpha 2/delta subunit 3
216	rs6794229	0,85	2,15E-04	0,260	3	54189989	T	G	55799	CACNA2D3	calcium channel, voltage-dependent, alpha 2/delta subunit 3
217	rs13061634	0,87	7,75E-04	0,307	3	56029117	C	T	26059	ERC2	ELKS/RAB6-interacting/CAST family member 2
218	rs1021734	0,82	7,23E-04	0,129	3	56938384	T	C	50650	ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3
219	rs17288993	0,81	3,12E-04	0,131	3	56940107	G	A	50650	ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3

222	rs17400084	1,25	1,46E-04	0,117	3	60261426	T	C	2272	FHIT	fragile histidine triad
223	rs11707184	1,18	2,12E-04	0,215	3	62316084	T	C			
224	rs831080	0,86	3,13E-04	0,273	3	71515191	C	G	27086	FOXP1	forkhead box P1
225	rs831081	0,85	1,91E-04	0,250	3	71515298	A	G	27086	FOXP1	forkhead box P1
226	rs6766190	1,24	9,16E-04	0,102	3	73871082	A	T			
227	rs291475	1,26	6,23E-04	0,091	3	73883578	C	G			
228	rs524431	0,87	5,60E-04	0,296	3	74383584	A	G			
229	rs471800	0,87	9,77E-04	0,312	3	74392334	T	C			
230	rs6551483	0,88	9,58E-04	0,364	3	87568689	C	T			
231	rs9815149	0,87	4,77E-04	0,366	3	87569165	G	C			
232	rs9816344	1,14	5,71E-04	0,408	3	115162780	C	T	254887	ZDHHC23	zinc finger, DHHC-type containing 23
233	rs9840925	1,34	7,41E-04	0,054	3	116582546	G	A			
234	rs16823934	1,16	9,09E-04	0,228	3	116818374	A	G			
235	rs17281612	1,21	8,85E-04	0,122	3	120606689	C	T	57514	ARHGAP31	Rho GTPase activating protein 31
236	rs1132202	1,22	5,93E-04	0,122	3	120633181	C	G	55254	TMEM39A	transmembrane protein 39A
237	rs4314124	0,87	6,53E-04	0,278	3	127270322	A	G	54946	SLC41A3	solute carrier family 41, member 3
238	rs6796610	0,87	7,97E-04	0,278	3	127280603	A	G	54946	SLC41A3	solute carrier family 41, member 3
239	rs2365012	0,85	6,97E-05	0,348	3	127299894	T	A	54946	SLC41A3	solute carrier family 41, member 3
240	rs11715474	1,17	6,46E-05	0,350	3	150284605	T	G	6596	HLTF	helicase-like transcription factor
241	rs7646166	1,15	4,62E-04	0,345	3	150307102	A	G			
242	rs6792168	1,15	6,29E-04	0,307	3	150319263	C	T			
243	rs12695943	1,16	9,81E-04	0,225	3	150988107	A	T	389161	ANKUB1	ankyrin repeat and ubiquitin domain containing 1
244	rs877439	0,88	8,89E-04	0,497	3	169282596	C	T	27333	GOLIM4	golgi integral membrane protein 4
245	rs1522378	0,88	5,13E-04	0,496	3	169283231	G	A	27333	GOLIM4	golgi integral membrane protein 4
246	rs10490809	0,84	1,64E-04	0,227	3	172699449	G	A			
248	rs1565567	0,85	4,30E-04	0,224	3	172706855	A	T			
249	rs1402002	1,14	8,86E-04	0,451	3	185125488	A	G	10057	ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5
250	rs939338	1,14	4,01E-04	0,446	3	185186762	G	A	10057	ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5
251	rs10937330	1,14	3,64E-04	0,479	3	189221460	G	A			
252	rs7613340	0,85	7,90E-04	0,179	3	189233423	C	T			
254	rs10938681	0,78	8,69E-05	0,099	4	8066769	A	G	84448	ABLIM2	actin binding LIM protein family, member 2
255	rs7662477	1,14	6,44E-04	0,482	4	23847568	A	G			
256	rs11726723	1,17	8,07E-04	0,190	4	26065365	T	G			
257	rs10034033	1,21	1,21E-04	0,176	4	26071049	A	C			
258	rs17219704	1,17	9,18E-04	0,200	4	61735051	A	G			
259	rs13150883	0,81	4,64E-04	0,108	4	65828632	C	T			
260	rs17750311	0,83	3,36E-04	0,157	4	65866784	G	A			
261	rs6849315	1,17	9,60E-04	0,191	4	83795901	A	T	79966	SCD5	stearoyl-CoA desaturase 5
262	rs7377204	0,85	2,29E-04	0,261	4	88727430	C	T			
263	rs7377225	0,86	3,10E-04	0,262	4	88727547	C	T			
264	rs4693846	0,85	1,52E-04	0,262	4	88728693	A	C			
265	rs10006978	1,26	3,07E-05	0,136	4	96898971	G	A			
266	rs7657124	1,24	3,38E-05	0,155	4	96914610	C	A			
267	rs11931752	1,27	1,64E-05	0,135	4	96938876	A	T			
268	rs11946552	1,24	4,24E-05	0,151	4	96940053	A	C			
269	rs17024571	1,23	1,09E-04	0,147	4	96942220	G	A			
270	rs1836900	1,17	6,12E-04	0,204	4	96958725	G	A			
271	rs10433975	1,18	6,58E-04	0,195	4	96960072	G	A			
272	rs1836899	1,17	8,21E-04	0,195	4	96966126	A	G			

274	rs17473405	1,24	1,54E-04	0,126	4	96970645	A	T			
275	rs13107501	1,17	9,34E-04	0,194	4	96971822	C	T			
276	rs17024826	1,22	5,54E-04	0,126	4	96983626	C	T			
277	rs17475948	1,28	3,09E-05	0,109	4	97002780	C	G			
278	rs12501586	1,17	8,91E-05	0,307	4	102861035	T	C			
279	rs12505043	1,18	2,15E-04	0,244	4	102874385	T	C			
282	rs13136521	0,86	9,10E-05	0,392	4	144425014	T	C			
284	rs7679856	0,88	7,03E-04	0,401	4	160315988	G	C			
285	rs7683671	0,88	8,79E-04	0,401	4	160334667	A	G			
286	rs11939106	0,88	9,63E-04	0,402	4	160335911	T	C			
287	rs10050099	0,88	8,45E-04	0,358	4	160343857	T	G			
288	rs1434621	1,15	8,80E-04	0,274	4	162869105	G	C	56884	FSTL5	follistatin-like 5
289	rs7660373	1,37	9,21E-06	0,077	4	162915033	T	C	56884	FSTL5	follistatin-like 5
290	rs13117869	1,19	3,19E-05	0,269	4	189923244	G	C			
291	rs4863069	1,20	1,64E-05	0,263	4	189928060	A	C			
292	rs6553232	1,19	2,00E-04	0,206	4	189947109	G	A			
293	rs11942138	1,18	2,58E-04	0,238	4	189969188	G	C			
295	rs10491223	0,87	4,47E-04	0,395	5	8843528	C	G			
296	rs10491222	0,87	4,18E-04	0,395	5	8870497	A	G			
297	rs396	0,88	8,05E-04	0,428	5	9668339	C	G			
298	rs2530913	0,78	1,32E-04	0,094	5	11638455	T	C	1501	CTNND2	catenin (cadherin-associated protein), delta 2
299	rs4866046	0,88	8,30E-04	0,357	5	20270802	A	G			
300	rs4866047	0,87	7,19E-04	0,354	5	20270828	C	A			
301	rs10037115	0,88	8,77E-04	0,355	5	20272670	G	A			
302	rs8180522	0,87	5,26E-04	0,355	5	20274979	C	G			
303	rs2974602	0,88	8,66E-04	0,431	5	20286581	C	T			
304	rs13164886	0,88	7,06E-04	0,484	5	20302871	T	G			
305	rs2974591	0,88	6,79E-04	0,437	5	20325791	C	T			
306	rs4429812	0,86	3,48E-04	0,277	5	27209030	C	T			
307	rs4518345	0,86	2,68E-04	0,277	5	27221661	A	G			
308	rs4510545	0,86	2,17E-04	0,277	5	27225107	C	A			
309	rs6880526	0,86	2,22E-04	0,280	5	27227465	T	C			
310	rs6890310	0,86	2,95E-04	0,278	5	27229330	A	G			
311	rs2199214	0,84	8,33E-04	0,166	5	27338180	C	T			
312	rs1428256	1,18	4,18E-04	0,198	5	38309217	T	G	133584	EGFLAM	EGF-like, fibronectin type III and laminin G domains
313	rs1834967	1,23	2,95E-04	0,123	5	38401890	A	G	133584	EGFLAM	EGF-like, fibronectin type III and laminin G domains
314	rs4336383	1,15	5,91E-04	0,323	5	38831091	A	T			
315	rs6886001	0,88	6,63E-04	0,483	5	52222194	C	T	3672	ITGA1	integrin, alpha 1
316	rs6866823	0,88	5,37E-04	0,484	5	52222328	A	G	3672	ITGA1	integrin, alpha 1
317	rs6871286	0,88	5,92E-04	0,479	5	52222513	T	C	3672	ITGA1	integrin, alpha 1
318	rs1979398	0,88	7,33E-04	0,473	5	52230084	A	G	3672	ITGA1	integrin, alpha 1
319	rs16886034	0,76	3,06E-04	0,067	5	56019613	C	T			
320	rs16886364	0,77	5,91E-04	0,068	5	56158101	G	A	4214	MAP3K1	mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase
321	rs16886448	0,77	5,91E-04	0,068	5	56206570	G	C	4214	MAP3K1	mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase
322	rs16886496	0,78	1,20E-04	0,093	5	56253286	C	T			
323	rs7726354	0,75	3,94E-04	0,056	5	56292240	T	C			
324	rs7725377	0,81	6,65E-04	0,103	5	56292353	A	G			

325	rs786699	1,32	8,76E-04	0,056	5	64711237	A	C	11174	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif, 6
326	rs12514992	1,15	6,33E-04	0,282	5	75554502	G	T	22987	SV2C	synaptic vesicle glycoprotein 2C
327	rs12516836	1,15	8,41E-04	0,278	5	75554524	A	G	22987	SV2C	synaptic vesicle glycoprotein 2C
328	rs4704438	0,86	1,83E-04	0,339	5	76980795	G	A			
329	rs1422406	0,88	5,85E-04	0,433	5	76981162	C	A			
331	rs3846620	1,23	3,40E-04	0,120	5	103014552	C	G			
332	rs6892259	1,22	5,65E-04	0,121	5	110113641	C	A	91137	SLC25A46	solute carrier family 25, member 46
333	rs456236	0,88	8,55E-04	0,413	5	110115057	G	T	91137	SLC25A46	solute carrier family 25, member 46
334	rs7723767	1,17	9,17E-04	0,216	5	110182685	C	T			
335	rs12517265	1,17	6,16E-04	0,224	5	110189680	T	C			
337	rs1350294	1,17	4,86E-04	0,222	5	110205180	A	C			
338	rs2416248	1,17	8,16E-04	0,224	5	110206705	G	A			
339	rs11745646	1,14	8,64E-04	0,323	5	110521442	G	A			
341	rs9327027	1,29	7,69E-04	0,067	5	116418496	A	T			
342	rs9327165	1,14	9,62E-04	0,418	5	120168056	C	T			
344	rs6878559	1,14	4,69E-04	0,445	5	120236091	G	A			
346	rs31330	0,85	3,07E-04	0,225	5	132889400	C	G	23105	FSTL4	follistatin-like 4
347	rs2160505	0,87	3,34E-04	0,431	5	157292346	A	C			
348	rs7709212	1,16	3,08E-04	0,335	5	158696755	C	T			
350	rs6887695	1,16	3,08E-04	0,321	5	158755223	C	G			
351	rs454036	1,14	9,82E-04	0,326	5	172486267	C	G	153222	CREBRF	CREB3 regulatory factor
352	rs255318	1,35	5,60E-05	0,068	5	172548635	A	G			
353	rs10456781	0,87	4,91E-04	0,396	6	16125021	G	A			
354	rs1150644	1,17	2,13E-04	0,278	6	16922283	A	C			
356	rs9396712	1,16	5,37E-04	0,276	6	16926604	T	C			
359	rs7767391	1,18	5,69E-04	0,198	6	20833219	C	T	54901	CDKAL1	CDK5 regulatory subunit associated protein 1-like 1
361	rs2516478	1,20	3,64E-04	0,168	6	31606716	A	G			
362	rs2523503	1,19	7,91E-04	0,152	6	31621538	A	C	534	ATP6V1G2	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G2
363	rs3117108	0,87	5,92E-04	0,305	6	32450800	C	G			
365	rs9269202	0,86	3,91E-04	0,289	6	32557501	T	C			
366	rs12202197	0,86	1,32E-04	0,363	6	39200945	C	T			
367	rs12195232	0,87	2,60E-04	0,360	6	39201111	T	C			
369	rs6910476	1,18	9,97E-04	0,164	6	48633649	G	A			
370	rs6458620	1,16	8,94E-04	0,261	6	48678807	C	G			
371	rs3010529	1,16	7,94E-04	0,262	6	48701049	C	T			
372	rs761167	1,14	7,52E-04	0,471	6	52219767	T	C			
373	rs1266825	1,14	7,75E-04	0,466	6	52221625	T	C			
374	rs3765446	1,14	7,63E-04	0,428	6	52249629	T	A	4172	MCM3	minichromosome maintenance complex component 3
375	rs12204627	0,82	2,42E-05	0,204	6	71778351	A	T			
376	rs9342803	0,84	6,73E-04	0,172	6	71781245	C	T			
377	rs1996679	0,84	6,36E-05	0,241	6	71783440	G	C			
378	rs9446323	0,84	1,81E-04	0,216	6	71789723	G	A			
379	rs7739908	0,78	9,65E-04	0,067	6	72090767	G	T			
380	rs16885102	0,82	2,30E-04	0,157	6	75341319	T	C			
381	rs9343877	1,28	5,26E-04	0,076	6	79922723	T	A			
382	rs6454097	1,28	5,05E-04	0,077	6	79934936	T	G			
383	rs1343232	1,14	7,49E-04	0,415	6	82187051	G	A			
384	rs17438648	1,14	7,37E-04	0,418	6	82216223	A	G			

385	rs11966310	1,14	7,49E-04	0,418	6	82217732	G	A			
386	rs11964002	1,14	9,02E-04	0,418	6	82217840	A	T			
387	rs4642522	0,88	5,66E-04	0,407	6	82249727	T	G			
388	rs1341230	0,88	9,00E-04	0,482	6	82436294	C	T			
389	rs9373855	1,14	7,74E-04	0,363	6	106922248	T	G			
390	rs488282	1,14	8,44E-04	0,363	6	106923806	A	G			
391	rs10457307	0,76	1,56E-04	0,071	6	116927364	A	G	100128327	BET3L	BET3 like (S. cerevisiae)
392	rs1338980	1,17	4,64E-04	0,232	6	118325563	A	G			
393	rs1998458	1,16	7,43E-04	0,239	6	118367258	G	T	222553	SLC35F1	solute carrier family 35, member F1
394	rs2789010	1,16	6,40E-04	0,239	6	118368173	T	G	222553	SLC35F1	solute carrier family 35, member F1
395	rs1416419	1,16	7,72E-04	0,239	6	118369087	T	A	222553	SLC35F1	solute carrier family 35, member F1
397	rs9321916	0,83	7,90E-04	0,139	6	143878225	A	T			
398	rs6570562	0,84	4,51E-04	0,181	6	143879508	A	G			
399	rs6908896	1,22	2,20E-04	0,151	6	156869074	C	A			
400	rs317801	0,86	8,16E-04	0,240	6	159010196	T	C	94120	SYTL3	synaptotagmin-like 3
401	rs6902491	1,21	4,71E-04	0,142	6	166381836	G	T			
402	rs4722483	0,83	5,78E-04	0,143	7	3159273	C	G			
403	rs17789894	0,82	4,66E-04	0,131	7	6709622	T	G	7559	ZNF12	zinc finger protein 12
405	rs7782529	0,81	3,11E-04	0,132	7	27264316	A	G			
406	rs11769156	1,24	8,68E-04	0,091	7	28545359	C	T	9586	CREB5	cAMP responsive element binding protein 5
407	rs10228072	1,15	2,98E-04	0,428	7	29542212	C	T			
409	rs12700969	1,14	7,82E-04	0,425	7	29552772	A	C			
411	rs17159921	1,36	3,09E-04	0,051	7	31123725	T	C			
413	rs2113643	1,14	5,56E-04	0,482	7	52104228	G	T			
414	rs7787769	0,80	1,25E-04	0,126	7	52963068	G	C			
415	rs11763192	0,81	6,06E-04	0,110	7	53002660	T	C			
417	rs1404198	0,81	1,46E-04	0,130	7	54052372	A	G			
418	rs10225389	0,84	4,45E-04	0,171	7	62973655	C	A			
420	rs4416776	0,85	2,66E-05	0,439	7	82814002	G	A			
421	rs2618989	1,15	8,54E-04	0,299	7	95193842	C	A			
422	rs450854	0,88	8,10E-04	0,392	7	101485453	T	C	1523	CUX1	cut-like homeobox 1
423	rs12538286	0,86	3,15E-04	0,285	7	101536040	A	G	1523	CUX1	cut-like homeobox 1
424	rs10270614	0,88	4,83E-04	0,459	7	101624464	A	G	1523	CUX1	cut-like homeobox 1
425	rs7341475	1,18	9,79E-04	0,169	7	103192051	A	G	5649	RELN	reelin
426	rs4730052	1,17	8,69E-04	0,213	7	104269557	C	T	375612	LHFPL3	lipoma HMGIC fusion partner-like 3
427	rs4730053	1,17	6,34E-04	0,213	7	104269619	A	G	375612	LHFPL3	lipoma HMGIC fusion partner-like 3
428	rs10245031	0,88	8,89E-04	0,346	7	117285697	C	T	83992	CTTNBP2	cortactin binding protein 2
429	rs7801931	0,86	2,49E-04	0,329	7	117294094	G	C	83992	CTTNBP2	cortactin binding protein 2
430	rs10270960	0,86	2,77E-04	0,328	7	117312875	C	G			
431	rs1357674	1,17	8,70E-04	0,191	7	119236456	G	A			
432	rs11764046	1,17	9,94E-04	0,188	7	119324606	G	A			
433	rs12707008	1,14	8,86E-04	0,404	7	131282522	T	C			
434	rs6467643	0,86	5,23E-04	0,274	7	135614381	T	G			
435	rs2701016	0,87	9,73E-04	0,279	7	135622254	A	C			
436	rs2555048	1,14	8,64E-04	0,348	7	135622266	C	T			
437	rs361445	1,21	7,34E-04	0,127	7	141838625	T	C	28601	TRBV6-6	T cell receptor beta variable 6-6
438	rs855733	0,86	2,38E-04	0,333	7	148993580	A	G			
439	rs1731847	0,88	5,00E-04	0,467	7	155348283	C	T			

440	rs1968853	1,14	8,65E-04	0,457	8	9083722	C	A			
441	rs2929301	1,15	5,16E-04	0,363	8	9085514	G	A			
442	rs2705042	1,37	2,66E-04	0,053	8	17366632	T	C			
443	rs11989798	0,74	3,97E-04	0,055	8	22326597	A	C	23516	SLC39A14	solute carrier family 39 (zinc transporter), member 14
444	rs2976405	0,87	4,33E-04	0,358	8	24911831	A	G			
445	rs12681837	0,86	9,95E-04	0,249	8	27191888	T	G			
446	rs6997728	0,86	9,68E-04	0,250	8	27196000	T	A			
447	rs4733453	0,88	7,26E-04	0,443	8	33770287	G	A			
448	rs4733456	0,88	7,79E-04	0,443	8	33775901	A	G			
449	rs4389890	0,88	8,23E-04	0,442	8	33777560	A	G			
450	rs7825337	0,88	8,20E-04	0,431	8	41626394	C	T			
451	rs12549902	0,88	6,57E-04	0,405	8	41628416	G	A			
452	rs4317621	0,88	9,46E-04	0,410	8	41635738	A	G	286	ANK1	ankyrin 1, erythrocytic
453	rs10504242	0,77	2,66E-04	0,079	8	59148749	G	A	90362	FAM110B	family with sequence similarity 110, member B
454	rs12678728	0,86	6,74E-04	0,218	8	62909278	G	A			
456	rs4268118	0,81	2,08E-04	0,125	8	63217632	G	A			
457	rs4256587	0,80	1,06E-04	0,132	8	63218545	T	C			
458	rs7832144	0,80	1,35E-04	0,126	8	63225135	A	G			
459	rs10504344	0,81	2,11E-04	0,128	8	63229338	G	T			
460	rs16928545	0,75	4,81E-06	0,105	8	63256978	G	A			
461	rs7833958	0,82	1,75E-04	0,152	8	63273320	A	G			
462	rs16928602	0,82	9,57E-05	0,156	8	63309109	T	C			
463	rs10957216	0,81	1,14E-04	0,143	8	63319367	T	A			
464	rs13278423	0,88	5,08E-04	0,488	8	87789535	A	C	54714	CNGB3	cyclic nucleotide gated channel beta 3
465	rs2436860	1,25	5,49E-04	0,092	8	103811225	A	G			
466	rs2514756	1,16	8,04E-04	0,247	8	119151124	A	G	2131	EXT1	exostosin 1
468	rs10960363	1,20	5,55E-04	0,162	9	1190703	C	T			
470	rs10811330	1,23	1,65E-05	0,200	9	20197095	C	T			
471	rs10964477	1,37	9,75E-05	0,060	9	20206063	C	T			
473	rs4977395	1,47	7,64E-06	0,053	9	20216358	G	A			
474	rs10964493	1,34	2,87E-04	0,061	9	20229840	C	T			
475	rs10964495	1,37	6,32E-05	0,064	9	20235283	C	T			
476	rs16923521	1,44	1,07E-05	0,058	9	20251635	C	T			
478	rs7041951	1,46	6,41E-06	0,057	9	20265354	G	C			
479	rs4977251	1,37	6,35E-05	0,063	9	20269793	G	A			
480	rs13300741	0,83	1,41E-04	0,178	9	20953339	C	T	54914	FOCAD	focadhesin
481	rs10966484	0,83	1,74E-04	0,169	9	24802191	G	A			
482	rs676484	1,14	7,52E-04	0,367	9	25953989	C	A			
483	rs17559639	0,87	5,63E-04	0,334	9	26011612	A	C			
484	rs10738743	1,15	4,94E-04	0,371	9	26027974	C	T			
486	rs506086	0,84	1,07E-04	0,257	9	78516428	C	G	158471	PRUNE2	prune homolog 2 (Drosophila)
488	rs2209882	1,26	5,06E-04	0,090	9	81127236	A	G			
490	rs6479067	0,86	3,92E-04	0,257	9	103635386	A	T			
491	rs2786716	0,86	6,66E-04	0,257	9	103636342	C	T			
492	rs1415647	0,86	8,06E-04	0,256	9	103636455	A	T			
493	rs10739816	0,86	6,53E-04	0,251	9	103656291	C	T			
494	rs10739592	1,34	2,08E-14	0,485	9	123011433	G	A			
495	rs10760182	1,14	6,75E-04	0,481	9	123452782	A	G	153090	DAB2IP	DAB2 interacting protein
496	rs7468351	1,14	8,33E-04	0,369	9	138114710	T	C	138151	NACC2	NACC family member 2, BEN and BTB (POZ) domain containing
497	rs3802577	0,87	5,18E-04	0,298	10	13361864	C	T	5264	PHYH	phytanoyl-CoA 2-hydroxylase

498	rs956007	1,18	1,35E-04	0,276	10	23761418	G	T			
499	rs7920535	1,15	8,72E-04	0,309	10	23774744	G	A			
500	rs12246098	1,16	4,55E-04	0,312	10	23786957	G	A			
501	rs11013514	1,17	1,12E-04	0,291	10	23799607	A	G			
502	rs7085999	1,14	9,35E-04	0,347	10	23800758	G	C			
503	rs7900252	1,16	1,62E-04	0,340	10	23802398	G	A			
504	rs6482285	1,14	7,67E-04	0,348	10	23808719	T	C			
505	rs4333914	1,16	3,31E-04	0,328	10	23810664	A	G			
506	rs6482289	1,14	7,67E-04	0,347	10	23816469	T	C			
508	rs7913401	1,15	3,30E-04	0,341	10	23844221	A	C			
509	rs1856113	1,16	2,23E-04	0,341	10	23844775	T	C			
510	rs983990	1,16	2,63E-04	0,342	10	23846388	G	A			
511	rs11013555	1,15	9,25E-04	0,284	10	23858933	A	G			
512	rs10763790	0,86	9,06E-04	0,244	10	30831361	C	G			
513	rs11593943	0,87	4,44E-04	0,377	10	33585087	T	C	8829	NRP1	neuropilin 1
514	rs10430541	0,88	7,30E-04	0,395	10	56494253	A	G			
516	rs2658630	0,85	3,09E-04	0,208	10	59409250	A	G			
517	rs1930450	0,84	1,74E-04	0,221	10	59410701	T	G			
518	rs2939583	0,85	2,19E-04	0,224	10	59412336	T	C			
519	rs2393400	0,85	2,32E-04	0,225	10	59414510	T	G			
520	rs1930455	0,85	3,87E-04	0,223	10	59414530	A	G			
521	rs1930456	0,84	2,03E-04	0,224	10	59414551	A	G			
522	rs10740725	0,88	9,56E-04	0,371	10	59460061	G	A			
523	rs11006021	0,87	4,72E-04	0,377	10	59460712	C	T			
524	rs1759365	0,85	3,68E-04	0,226	10	59490502	A	G			
525	rs3915932	0,85	4,70E-05	0,409	10	80611942	C	G	57178	ZMIZ1	zinc finger, MIZ-type containing 1
526	rs810517	0,85	3,08E-05	0,452	10	80612626	T	C	57178	ZMIZ1	zinc finger, MIZ-type containing 1
527	rs12571751	0,85	3,05E-05	0,452	10	80612637	G	A	57178	ZMIZ1	zinc finger, MIZ-type containing 1
528	rs703982	0,86	6,72E-05	0,395	10	80612727	G	A	57178	ZMIZ1	zinc finger, MIZ-type containing 1
529	rs11553840	1,32	7,27E-04	0,054	10	82268160	C	T	81619	TSPAN14	tetraspanin 14
530	rs17415112	0,84	7,52E-04	0,156	10	99194781	A	G	51013	EXOSC1	exosome component 1
531	rs11191841	0,87	3,97E-04	0,500	10	105629601	C	T			
532	rs7100920	0,87	3,81E-04	0,488	10	105630968	T	C			
533	rs10883942	0,87	2,30E-04	0,486	10	105641376	C	T	79991	OBFC1	oligonucleotide/oligosaccharide-binding fold containing 1
534	rs12765878	0,87	2,30E-04	0,486	10	105659612	C	T	79991	OBFC1	oligonucleotide/oligosaccharide-binding fold containing 1
535	rs1421503	1,16	7,37E-04	0,227	10	107485090	G	A			
536	rs2111995	1,16	9,82E-04	0,226	10	107497352	G	A			
537	rs10787019	1,15	9,79E-04	0,306	10	109050808	T	G			
538	rs2804611	1,25	4,64E-04	0,107	10	113837462	C	T			
539	rs2804614	1,24	7,72E-04	0,107	10	113841831	C	T			
540	rs4074720	1,16	1,30E-04	0,476	10	114738487	T	C	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
541	rs7901695	1,28	8,18E-10	0,328	10	114744078	C	T	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
542	rs4506565	1,28	9,48E-10	0,331	10	114746031	T	A	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
543	rs4132670	1,28	6,53E-10	0,331	10	114757761	A	G	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
544	rs6585201	1,14	3,96E-04	0,459	10	114758773	A	G	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
545	rs10787472	1,16	1,38E-04	0,474	10	114771287	C	A	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
546	rs12243326	1,29	6,12E-10	0,295	10	114778805	C	T	6934	TCF7L2	transcription factor 7-like 2 (T-cell

											specific, HMG-box)
548	rs11196205	1,17	3,49E-05	0,473	10	114797037	C	G	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
549	rs10885409	1,17	2,71E-05	0,472	10	114798062	C	T	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
550	rs12255372	1,29	4,37E-10	0,302	10	114798892	T	G	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
551	rs11196208	1,17	2,40E-05	0,472	10	114801306	C	T	6934	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
552	rs10510004	0,86	1,97E-04	0,373	10	116214569	A	G	3983	ABLIM1	actin binding LIM protein 1
555	rs2420928	0,87	4,36E-04	0,408	10	123143462	G	A			
556	rs1322328	0,88	9,04E-04	0,466	10	123911094	C	G	10579	TACC2	transforming, acidic coiled-coil containing protein 2
557	rs12412485	1,16	6,93E-04	0,234	10	131731590	T	G			
558	rs7075825	0,76	1,83E-04	0,075	10	133720979	T	C			
559	rs11827296	1,19	3,63E-04	0,186	11	3334236	C	T			
560	rs7104128	1,23	9,03E-04	0,106	11	4697321	T	C			
561	rs935951	0,83	2,12E-04	0,166	11	5918145	T	G			
562	rs2723663	0,88	4,90E-04	0,466	11	6440086	C	A	10612	TRIM3	tripartite motif containing 3
565	rs1881820	0,85	1,09E-04	0,292	11	13757134	G	C			
566	rs2351044	1,17	6,83E-05	0,363	11	15535033	A	G			
567	rs7117077	0,84	4,65E-04	0,182	11	19510993	C	T	89797	NAV2	neuron navigator 2
568	rs329526	1,14	7,56E-04	0,438	11	29458729	T	G			
569	rs2926461	0,87	7,80E-04	0,303	11	34208169	C	T	25841	ABTB2	ankyrin repeat and BTB (POZ) domain containing 2
570	rs2957523	0,87	5,25E-04	0,303	11	34208431	G	A	25841	ABTB2	ankyrin repeat and BTB (POZ) domain containing 2
571	rs2926463	0,87	7,23E-04	0,302	11	34208964	G	T	25841	ABTB2	ankyrin repeat and BTB (POZ) domain containing 2
572	rs2955949	0,86	3,64E-04	0,304	11	34210651	A	T	25841	ABTB2	ankyrin repeat and BTB (POZ) domain containing 2
573	rs7115702	1,18	9,69E-05	0,291	11	61787955	T	A			
574	rs11603383	1,17	9,99E-05	0,291	11	61794159	A	G	4250	SCGB2A2	secretoglobin, family 2A, member 2
575	rs17709552	1,25	1,59E-04	0,117	11	61797095	G	A	4250	SCGB2A2	secretoglobin, family 2A, member 2
576	rs11228506	0,88	8,60E-04	0,458	11	68645758	A	G			
577	rs644961	0,88	9,73E-04	0,473	11	78370468	T	C	26011	ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
578	rs10793350	0,86	1,17E-04	0,483	11	78372163	T	C	26011	ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
579	rs10751301	0,86	9,09E-05	0,485	11	78372286	G	C	26011	ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
580	rs11237675	0,87	2,56E-04	0,490	11	78375191	C	T	26011	ODZ4	odz, odd Oz/ten-m homolog 4 (Drosophila)
581	rs17310875	1,21	7,14E-04	0,130	11	79832113	C	G			
582	rs11232429	1,35	3,79E-04	0,052	11	80397567	T	A			
583	rs11235302	1,21	5,02E-04	0,135	11	87132574	A	T			
584	rs17150852	1,26	3,74E-04	0,089	11	87202808	A	G			
585	rs17833579	1,26	5,03E-04	0,087	11	87203798	C	T			
586	rs17150882	1,27	2,10E-04	0,095	11	87219070	C	T			
587	rs9666479	1,20	3,35E-04	0,160	11	87250138	G	A			
588	rs7121252	1,21	2,52E-04	0,163	11	87256116	C	T			
589	rs1939168	1,17	7,65E-04	0,212	11	87288340	A	G			
590	rs7101865	1,17	3,13E-04	0,231	11	87577209	A	G			
592	rs7937882	1,23	8,26E-05	0,157	11	87579997	G	A			
594	rs11020093	0,86	4,36E-04	0,247	11	92267291	T	C	120114	FAT3	FAT tumor suppressor homolog 3

											(Drosophila)
595	rs17134278	1,25	6,85E-04	0,094	11	99106275	G	C	53942	CNTN5	contactin 5
596	rs4559717	1,25	7,39E-04	0,087	11	112656309	A	G			
597	rs1600223	0,84	4,84E-04	0,165	11	126798259	T	C			
598	rs3935794	0,72	4,41E-05	0,063	11	127895887	G	A	2113	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
599	rs3935795	0,72	3,45E-05	0,064	11	127896001	C	T	2113	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
600	rs3935796	0,75	2,56E-04	0,061	11	127896137	A	T	2113	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
601	rs4937342	0,75	2,31E-04	0,063	11	127903519	G	T	2113	ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
602	rs433443	0,87	6,98E-04	0,334	11	130876412	A	G	50863	NTM	neurotrimin
603	rs1870199	0,87	3,73E-04	0,444	12	656499	A	G			
604	rs10849464	0,86	6,59E-05	0,391	12	659413	A	C			
607	rs10849040	1,15	3,01E-04	0,498	12	4312167	C	T	57103	C12orf5	chromosome 12 open reading frame 5
608	rs17700406	0,86	1,81E-04	0,378	12	4332859	C	T	57103	C12orf5	chromosome 12 open reading frame 5
609	rs10849045	0,86	1,57E-04	0,368	12	4337744	A	G	57103	C12orf5	chromosome 12 open reading frame 5
610	rs7135390	0,87	3,18E-04	0,444	12	21489968	T	C	79912	PYROXD1	pyridine nucleotide-disulphide oxidoreductase domain 1
611	rs11610942	0,87	4,42E-04	0,443	12	21492898	G	A	79912	PYROXD1	pyridine nucleotide-disulphide oxidoreductase domain 1
612	rs10841843	1,18	2,20E-04	0,249	12	21583158	T	C	2998	GYS2	glycogen synthase 2 (liver)
613	rs10492118	1,17	9,48E-04	0,215	12	21583225	T	C	2998	GYS2	glycogen synthase 2 (liver)
614	rs6487236	1,17	5,34E-04	0,227	12	21591183	G	A	2998	GYS2	glycogen synthase 2 (liver)
615	rs10841848	1,16	8,61E-04	0,226	12	21600821	A	G	2998	GYS2	glycogen synthase 2 (liver)
616	rs11046116	1,16	8,22E-04	0,226	12	21600886	G	C	2998	GYS2	glycogen synthase 2 (liver)
617	rs10770836	1,16	6,93E-04	0,248	12	21608008	A	G	2998	GYS2	glycogen synthase 2 (liver)
618	rs10841850	1,16	6,69E-04	0,248	12	21608123	G	A	2998	GYS2	glycogen synthase 2 (liver)
619	rs11046122	1,16	7,69E-04	0,250	12	21608288	T	C	2998	GYS2	glycogen synthase 2 (liver)
620	rs10783760	0,88	9,15E-04	0,358	12	54262230	A	G			
621	rs4759173	0,88	8,81E-04	0,357	12	54287453	A	G			
622	rs10747758	0,88	9,48E-04	0,369	12	54287594	T	C			
623	rs4759186	0,84	5,66E-04	0,176	12	54350346	A	G			
625	rs3916529	0,83	3,60E-04	0,153	12	62721863	G	A	57522	SRGAP1	SLIT-ROBO Rho GTPase activating protein 1
626	rs7132617	1,14	9,37E-04	0,392	12	63482244	A	G			
627	rs10878211	1,14	5,35E-04	0,396	12	63486189	C	T			
628	rs3851608	1,14	9,35E-04	0,381	12	63495765	G	A			
629	rs998314	1,14	5,49E-04	0,392	12	63506634	G	A	23329	TBC1D30	TBC1 domain family, member 30
632	rs12582634	1,29	9,45E-04	0,066	12	80385922	T	C	8499	PPFIA2	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2
633	rs12815988	0,77	1,70E-04	0,083	12	82183441	T	C			
634	rs11115663	0,76	9,08E-05	0,084	12	82184765	G	A			
635	rs12578418	0,79	4,98E-04	0,083	12	95081078	A	G			
636	rs7300815	1,30	2,15E-04	0,076	12	100486144	C	A			
637	rs12580632	1,21	9,75E-04	0,131	12	100486792	C	T			
638	rs855287	0,83	6,57E-04	0,146	12	101470239	A	T			
639	rs753479	0,83	3,25E-04	0,163	12	101482692	G	A			
640	rs10860877	0,84	6,74E-04	0,173	12	101483695	A	G			
641	rs4964671	1,15	6,62E-04	0,349	12	107227824	G	C	1240	CMKLR1	chemokine-like receptor 1

642	rs10400410	1,23	9,70E-04	0,100	12	109677882	A	G			
643	rs11067587	0,86	7,45E-04	0,259	12	114338107	C	T			
644	rs12313339	0,78	9,90E-04	0,070	12	119870876	A	G			
646	rs10773182	0,88	6,21E-04	0,393	12	124686312	G	T	114795	TMEM132B	transmembrane protein 132B
647	rs2058012	0,88	7,02E-04	0,462	12	124693362	G	A	114795	TMEM132B	transmembrane protein 132B
648	rs979589	0,87	4,55E-04	0,337	12	124693655	T	C	114795	TMEM132B	transmembrane protein 132B
649	rs3803152	0,87	4,00E-04	0,454	12	124701148	G	A	114795	TMEM132B	transmembrane protein 132B
650	rs3825381	0,86	6,46E-04	0,253	12	124702816	T	C	114795	TMEM132B	transmembrane protein 132B
651	rs10846941	1,14	4,56E-04	0,483	12	124720392	T	C			
652	rs10773187	1,14	7,97E-04	0,484	12	124724088	G	A			
653	rs10846955	1,13	8,82E-04	0,487	12	124762711	T	C			
654	rs10846980	1,14	4,13E-04	0,486	12	124857508	T	G			
655	rs7313371	0,88	9,86E-04	0,319	12	124861036	A	G			
656	rs7954415	0,86	2,46E-04	0,343	12	124862134	T	C			
657	rs917334	0,87	2,41E-04	0,395	12	124864111	G	A			
658	rs6489019	0,86	1,47E-04	0,393	12	124866835	A	G			
659	rs6489020	0,86	5,91E-05	0,445	12	124866956	C	T			
661	rs7978045	1,19	5,76E-04	0,159	12	124873743	T	C			
662	rs11058369	1,19	7,00E-04	0,162	12	124891017	T	A			
663	rs11610391	0,85	1,61E-05	0,441	12	124894658	T	G			
664	rs11058371	1,20	3,66E-05	0,248	12	124894762	A	G			
665	rs917337	0,86	8,28E-05	0,393	12	124903032	T	C			
666	rs11058574	0,87	1,44E-04	0,477	12	125256569	T	C			
667	rs10847114	0,86	1,24E-04	0,486	12	125256831	G	A			
668	rs10773245	0,87	1,55E-04	0,486	12	125257058	A	C			
669	rs10773247	0,86	9,24E-05	0,484	12	125260779	C	T			
670	rs10744243	0,86	9,12E-05	0,484	12	125260860	A	G			
671	rs2346669	0,87	1,78E-04	0,483	12	125264114	G	A			
672	rs10773257	0,87	5,11E-04	0,378	12	125292483	G	A			
673	rs2010484	1,17	1,36E-04	0,309	12	126298378	C	G			
674	rs10847919	1,19	5,42E-04	0,164	12	128706939	T	C	121256	TMEM132D	transmembrane protein 132D
675	rs452876	0,88	9,55E-04	0,416	12	129326198	G	T			
676	rs17357143	0,74	1,37E-04	0,063	13	22504391	C	T			
677	rs549305	1,18	8,55E-04	0,175	13	26037876	T	G	10810	WASF3	WAS protein family, member 3
679	rs2026960	0,83	8,28E-04	0,139	13	30458544	C	T			
680	rs4258502	1,14	5,87E-04	0,456	13	48461701	A	G	22862	FNDC3A	fibronectin type III domain containing 3A
681	rs9568143	1,15	2,88E-04	0,456	13	48468631	A	T	22862	FNDC3A	fibronectin type III domain containing 3A
682	rs4942796	0,87	4,10E-04	0,311	13	48484019	T	C	22862	FNDC3A	fibronectin type III domain containing 3A
683	rs9316428	0,87	9,53E-04	0,311	13	48526552	A	G	22862	FNDC3A	fibronectin type III domain containing 3A
684	rs1407827	0,87	6,17E-04	0,308	13	48608987	C	T	22862	FNDC3A	fibronectin type III domain containing 3A
685	rs1983805	0,87	7,21E-04	0,310	13	48609971	C	T	22862	FNDC3A	fibronectin type III domain containing 3A
687	rs1013347	0,86	2,95E-04	0,284	13	48708882	T	G			
688	rs9571208	1,22	2,03E-04	0,149	13	63876687	C	T			
689	rs7991210	1,19	9,62E-06	0,395	13	99549906	G	A	5095	PCCA	propionyl CoA carboxylase, alpha polypeptide
690	rs916048	1,14	9,35E-04	0,379	14	21860760	A	C			
691	rs3751488	1,16	9,91E-04	0,224	14	22373934	A	G	122704	MRPL52	mitochondrial ribosomal protein L52
692	rs424964	1,15	3,33E-04	0,425	14	30055554	A	G			

693	rs10135562	1,19	5,23E-04	0,182	14	32198399	T	C	9472	AKAP6	A kinase (PRKA) anchor protein 6
694	rs6571647	0,79	2,51E-04	0,093	14	33836835	G	A			
695	rs1998193	1,14	5,78E-04	0,456	14	38760507	T	G			
696	rs28502509	1,14	4,79E-04	0,457	14	38761768	C	T			
697	rs1387754	1,14	4,85E-04	0,421	14	62341315	C	T	27133	KCNH5	potassium voltage-gated channel, subfamily H (eag-related), member 5
698	rs4899384	0,88	7,61E-04	0,455	14	70695709	T	A			
699	rs10483837	1,21	1,07E-04	0,193	14	71519244	G	A	9628	RGS6	regulator of G-protein signaling 6
700	rs7156200	1,18	2,48E-04	0,241	14	71527435	C	A	9628	RGS6	regulator of G-protein signaling 6
701	rs12884777	1,18	2,35E-04	0,240	14	71530913	T	C	9628	RGS6	regulator of G-protein signaling 6
702	rs12885258	1,20	1,58E-04	0,197	14	71531051	A	G	9628	RGS6	regulator of G-protein signaling 6
703	rs2283422	1,18	1,44E-04	0,239	14	71531955	C	T	9628	RGS6	regulator of G-protein signaling 6
704	rs2283381	0,81	5,34E-07	0,255	14	71978830	G	A	9628	RGS6	regulator of G-protein signaling 6
706	rs1548687	0,84	5,46E-05	0,248	14	72028222	A	G	9628	RGS6	regulator of G-protein signaling 6
707	rs17119980	0,86	2,01E-04	0,333	14	72253994	A	T	8110	DPF3	D4, zinc and double PHD fingers, family 3
708	rs740974	0,86	2,76E-04	0,338	14	72257827	G	A	8110	DPF3	D4, zinc and double PHD fingers, family 3
709	rs4243642	0,86	2,33E-04	0,335	14	72258454	C	G	8110	DPF3	D4, zinc and double PHD fingers, family 3
710	rs17808467	0,82	7,61E-04	0,121	14	76239145	A	G			
711	rs11159227	0,85	8,86E-04	0,186	14	76269385	A	T			
712	rs17109221	1,16	9,72E-04	0,214	14	78979872	T	C	9369	NRXN3	neurexin 3
713	rs7144011	1,17	8,36E-04	0,216	14	79010136	T	G	9369	NRXN3	neurexin 3
714	rs7153625	0,81	3,27E-04	0,123	14	79119015	A	G	9369	NRXN3	neurexin 3
715	rs7154599	0,82	4,68E-04	0,126	14	79119562	C	G	9369	NRXN3	neurexin 3
716	rs17764096	0,81	3,01E-04	0,125	14	79120259	T	G	9369	NRXN3	neurexin 3
717	rs190092	0,85	1,40E-04	0,299	14	79121236	C	A	9369	NRXN3	neurexin 3
718	rs327465	0,86	1,07E-04	0,494	14	80299793	C	T	145508	CEP128	centrosomal protein 128kDa
719	rs2556611	0,86	1,19E-04	0,495	14	80415919	A	G	145508	CEP128	centrosomal protein 128kDa
720	rs12050342	0,87	2,07E-04	0,492	14	80438617	T	C	145508	CEP128	centrosomal protein 128kDa
721	rs2888032	0,87	1,77E-04	0,500	14	80439264	C	T	145508	CEP128	centrosomal protein 128kDa
722	rs11625199	0,87	3,73E-04	0,498	14	80442498	A	G	145508	CEP128	centrosomal protein 128kDa
723	rs6574608	0,87	3,77E-04	0,500	14	80444575	A	C	145508	CEP128	centrosomal protein 128kDa
724	rs10444745	0,86	1,00E-04	0,470	14	87891057	G	T			
725	rs11848957	1,24	7,53E-04	0,096	14	94731600	C	G	79789	CLMN	calmin (calponin-like, transmembrane)
726	rs12907278	1,15	2,60E-04	0,451	15	31760469	A	G	6263	RYR3	ryanodine receptor 3
727	rs12592542	1,14	5,00E-04	0,456	15	31773110	A	G	6263	RYR3	ryanodine receptor 3
728	rs16962542	0,72	4,52E-05	0,057	15	34189070	A	T			
729	rs7170955	1,19	1,32E-04	0,209	15	44444884	C	A			
730	rs7180600	1,20	5,53E-04	0,151	15	50857433	A	G	3175	ONECUT1	one cut homeobox 1
731	rs10518694	1,23	1,58E-04	0,142	15	50859965	A	C	3175	ONECUT1	one cut homeobox 1
732	rs2456526	1,22	2,77E-04	0,145	15	50876734	C	T			
733	rs10519107	0,86	3,93E-05	0,481	15	59114168	G	C	6095	RORA	RAR-related orphan receptor A
734	rs6494307	0,88	6,36E-04	0,413	15	60181982	G	C			
735	rs10083587	0,88	5,60E-04	0,413	15	60185825	T	C			
736	rs8030240	0,86	7,29E-04	0,259	15	60186856	T	C			
737	rs1436955	0,87	8,85E-04	0,263	15	60191674	T	C			
738	rs749555	#N/A	#N/A	#N/A	####	#N/A	##	##	#N/A	#N/A	#N/A
739	rs10083639	0,84	1,78E-04	0,193	15	68368811	A	G			
740	rs11072156	0,84	1,82E-04	0,191	15	68369472	A	T			
741	rs2059322	1,25	5,21E-04	0,106	15	68792114	C	A	55075	UACA	uveal autoantigen with coiled-coil

											domains and ankyrin repeats
742	rs10518921	1,24	6,24E-04	0,105	15	68793963	T	C	55075	UACA	uveal autoantigen with coiled-coil domains and ankyrin repeats
743	rs7177970	1,24	6,51E-04	0,104	15	68832308	G	A	55075	UACA	uveal autoantigen with coiled-coil domains and ankyrin repeats
744	rs6495081	1,14	6,53E-04	0,435	15	71903355	G	T			
745	rs2290271	1,15	3,43E-04	0,358	15	83248639	C	A	9154	SLC28A1	solute carrier family 28 (sodium-coupled nucleoside transporter), member 1
747	rs11636210	0,88	8,33E-04	0,434	15	89415484	C	T			
748	rs2131659	0,87	2,84E-04	0,475	15	98925810	T	G			
749	rs11247226	0,87	3,46E-04	0,468	15	98938486	C	T	55180	LINS	lines homolog (Drosophila)
750	rs8033689	0,87	4,08E-04	0,434	15	98951820	G	C	55180	LINS	lines homolog (Drosophila)
751	rs7180844	0,87	3,11E-04	0,473	15	98953582	T	C	55180	LINS	lines homolog (Drosophila)
753	rs8043935	1,18	6,16E-04	0,183	16	6306727	G	A			
754	rs12597219	1,18	3,83E-04	0,195	16	6310627	A	C			
755	rs8062975	1,19	1,86E-04	0,200	16	6311714	T	A			
756	rs809684	1,20	3,94E-05	0,227	16	6322435	A	G			
757	rs249301	1,21	1,61E-04	0,165	16	9377408	T	A			
759	rs216944	0,87	5,42E-04	0,356	16	58798850	A	G			
760	rs8063424	1,18	5,47E-04	0,208	16	78424987	T	C			
761	rs3924889	1,20	2,83E-04	0,172	16	78426000	C	A			
762	rs8062047	1,37	2,18E-04	0,052	16	78457228	G	T			
766	rs228768	1,14	9,15E-04	0,333	17	39547419	G	T	10014	HDAC5	histone deacetylase 5
767	rs1968393	1,15	7,52E-04	0,320	17	49919431	A	G			
768	rs4968816	0,88	6,27E-04	0,398	17	64200570	G	A			
769	rs11656969	1,20	5,35E-04	0,150	17	69822772	T	C	64446	DNAI2	dynein, axonemal, intermediate chain 2
770	rs8076794	1,14	4,76E-04	0,461	17	74172864	A	C			
771	rs6501238	1,14	5,87E-04	0,460	17	74181334	T	C			
772	rs10512617	1,13	9,67E-04	0,487	17	74205146	C	G	9267	CYTH1	cytohesin 1
773	rs1531797	1,15	2,27E-04	0,488	17	74333763	T	C	57602	USP36	ubiquitin specific peptidase 36
774	rs767300	1,14	9,45E-04	0,328	18	9830230	A	G	11031	RAB31	RAB31, member RAS oncogene family
775	rs471999	1,16	2,01E-04	0,329	18	9834729	G	A	11031	RAB31	RAB31, member RAS oncogene family
776	rs555935	1,15	2,39E-04	0,387	18	9835307	T	C	11031	RAB31	RAB31, member RAS oncogene family
777	rs575420	1,14	8,21E-04	0,421	18	9838371	A	C	11031	RAB31	RAB31, member RAS oncogene family
778	rs688248	1,16	2,85E-04	0,342	18	9838613	C	T	11031	RAB31	RAB31, member RAS oncogene family
779	rs508816	1,15	2,54E-04	0,393	18	9839620	T	C	11031	RAB31	RAB31, member RAS oncogene family
780	rs2299836	1,17	1,90E-04	0,287	18	9840212	A	G	11031	RAB31	RAB31, member RAS oncogene family
781	rs559655	1,20	2,26E-04	0,173	18	10055123	T	C			
782	rs3737361	1,16	3,64E-04	0,295	18	12821324	C	T	5771	PTPN2	protein tyrosine phosphatase, non-receptor type 2
783	rs9947011	0,86	4,90E-04	0,249	18	18674407	A	G			
784	rs6507323	0,85	2,93E-04	0,244	18	18680676	G	C			
785	rs3911557	0,85	2,06E-04	0,246	18	18726561	T	C			
786	rs4800138	0,85	2,15E-04	0,245	18	18768295	G	A	5932	RBBP8	retinoblastoma binding protein 8
787	rs9304261	0,86	8,68E-04	0,225	18	18860594	T	C	5932	RBBP8	retinoblastoma binding protein 8
789	rs2056015	1,13	8,99E-04	0,492	18	32164600	G	T	80206	FHOD3	formin homology 2 domain containing 3
790	rs16973756	0,75	6,68E-04	0,055	18	36617062	G	A			

791	rs7234864	1,17	1,57E-04	0,281	18	55885837	T	C			
792	rs1942867	1,18	8,71E-05	0,285	18	55887250	A	G			
793	rs11664327	1,17	6,36E-05	0,348	18	55890603	C	T			
794	rs8091524	1,19	5,56E-05	0,267	18	55902940	C	T			
795	rs1539952	1,18	1,73E-04	0,266	18	55917492	G	A			
796	rs9966951	1,15	6,16E-04	0,335	18	55926275	A	G			
797	rs6567157	1,16	2,35E-04	0,340	18	55941205	G	T			
798	rs1942880	1,17	1,02E-04	0,339	18	55944189	T	C			
799	rs7235626	1,16	1,92E-04	0,338	18	55949677	T	G			
800	rs17782313	1,16	7,07E-04	0,253	18	56002077	C	T			
801	rs476828	1,17	3,76E-04	0,258	18	56003567	C	T			
802	rs9947403	1,15	4,18E-04	0,349	18	56020730	T	C			
803	rs639407	1,15	3,44E-04	0,352	18	56021159	G	A			
804	rs619662	1,15	3,05E-04	0,398	18	56035531	A	G			
805	rs607104	1,14	8,97E-04	0,355	18	56042573	G	C			
806	rs557416	1,15	5,53E-04	0,347	18	56046039	G	A			
808	rs1421521	1,14	9,78E-04	0,348	18	60236486	A	G			
809	rs470443	1,16	5,80E-04	0,267	18	72832968	A	G	4155	MBP	myelin basic protein
810	rs4805258	1,17	8,10E-04	0,211	19	32763882	A	G			
811	rs7252689	1,17	6,56E-04	0,202	19	33080647	T	C			
812	rs1017207	1,18	8,08E-04	0,183	19	39057327	A	G			
813	rs7251215	1,17	2,09E-04	0,259	19	39099587	G	A			
814	rs10409299	1,17	9,10E-04	0,208	19	41016164	G	A	4868	NPHS1	nephrosis 1, congenital, Finnish type (nephrin)
815	rs41332947	0,82	4,90E-04	0,127	19	55391757	C	T			
816	rs2876409	1,14	7,11E-04	0,367	20	15415075	A	G	140733	MACROD2	MACRO domain containing 2
817	rs3746476	0,82	9,10E-04	0,109	20	36373583	G	A	671	BPI	bactericidal/permeability-increasing protein
818	rs6103249	0,83	7,24E-04	0,147	20	41399350	C	T			
819	rs6073055	0,87	7,12E-04	0,338	20	41406604	G	A			
820	rs16985285	0,84	7,84E-04	0,153	20	41448057	T	C			
821	rs6103716	1,16	3,05E-04	0,327	20	42433044	C	A	3172	HNF4A	hepatocyte nuclear factor 4, alpha
822	rs6063438	1,18	3,04E-04	0,204	20	47874575	T	C	23315	SLC9A8	solute carrier family 9, subfamily A (NHE8, cation proton antiporter 8), member 8
823	rs676035	1,19	2,26E-04	0,205	20	47916399	G	A	23315	SLC9A8	solute carrier family 9, subfamily A (NHE8, cation proton antiporter 8), member 8
825	rs1883553	1,19	6,38E-05	0,240	20	48007523	T	C			
826	rs6020178	1,17	9,91E-04	0,200	20	48037347	C	T	6615	SNAI1	snail homolog 1 (Drosophila)
827	rs2257	1,16	3,47E-04	0,306	20	51245861	G	C	128553	TSHZ2	teashirt zinc finger homeobox 2
830	rs6061921	0,86	6,35E-05	0,432	20	59966907	C	T			
831	rs6089568	0,85	2,69E-05	0,425	20	59967110	A	G			
832	rs2037994	1,23	8,93E-04	0,105	21	15880541	A	C			
833	rs2823759	1,25	2,57E-04	0,111	21	16667957	C	G	388815	LINC00478	long intergenic non-protein coding RNA 478
834	rs915856	1,24	3,59E-04	0,111	21	16668120	A	G	388815	LINC00478	long intergenic non-protein coding RNA 478
835	rs1667570	1,25	1,42E-04	0,112	21	16668591	G	A	388815	LINC00478	long intergenic non-protein coding RNA 478
836	rs380220	1,25	2,02E-04	0,109	21	16668953	A	G	388815	LINC00478	long intergenic non-protein coding RNA 478
837	rs369347	1,25	1,44E-04	0,115	21	16669662	G	A	388815	LINC00478	long intergenic non-protein coding RNA 478
838	rs158046	1,22	4,29E-04	0,132	21	18316684	C	T	140578	CHODL	chondrolectin
839	rs2826239	0,87	5,02E-04	0,326	21	20709622	T	G			

840	rs9980427	0,87	6,81E-04	0,346	21	20709933	A	G			
841	rs2826242	0,87	5,09E-04	0,326	21	20713322	T	C			
842	rs1985053	0,86	2,27E-04	0,328	21	20713786	G	A			
843	rs2826244	0,87	3,54E-04	0,340	21	20716906	G	C			
844	rs1029258	0,82	5,69E-04	0,136	21	26710675	C	A			
845	rs2831054	1,17	5,52E-04	0,239	21	27954865	A	G			
846	rs1888433	1,14	8,63E-04	0,386	21	27954964	T	C			
847	rs2831854	1,16	3,31E-04	0,283	21	28782159	T	C			
849	rs1999318	1,15	9,33E-04	0,284	21	28817820	C	A			
850	rs9975371	1,22	9,45E-04	0,114	21	28817851	T	C			
852	rs8132538	1,23	2,21E-05	0,183	21	37197902	A	G	3141	HLCS	holocarboxylase synthetase (biotin-(proprionyl-CoA-carboxylase (ATP-hydrolysing)) ligase)
853	rs2835530	1,24	1,20E-05	0,182	21	37199189	C	T	3141	HLCS	holocarboxylase synthetase (biotin-(proprionyl-CoA-carboxylase (ATP-hydrolysing)) ligase)
854	rs2845812	1,20	1,09E-04	0,189	21	37220194	T	C	3141	HLCS	holocarboxylase synthetase (biotin-(proprionyl-CoA-carboxylase (ATP-hydrolysing)) ligase)
856	rs220161	0,81	9,16E-04	0,109	21	42422362	C	G	89766	UMODL1	uromodulin-like 1
857	rs9981459	0,82	2,92E-04	0,148	21	42681878	G	C	64699	TMPRSS3	transmembrane protease, serine 3
858	rs2401163	0,83	4,02E-04	0,154	22	16511078	C	T	23786	BCL2L13	BCL2-like 13 (apoptosis facilitator)
859	rs2587103	0,84	8,87E-04	0,156	22	16528454	T	C	23786	BCL2L13	BCL2-like 13 (apoptosis facilitator)
860	rs713999	0,84	1,07E-05	0,376	22	46210776	A	G			
861	rs6008226	1,19	3,71E-04	0,184	22	46243314	C	T			
862	rs11090806	1,29	9,95E-04	0,065	22	46777160	A	C			
863	rs12009434	1,18	2,51E-04	0,324	23	12875922	A	G			
864	rs5979784	1,17	5,82E-04	0,329	23	12876296	C	A			
865	rs17277503	1,18	8,38E-04	0,239	23	56833086	G	A	550643	LOC550643	uncharacterized LOC550643
866	rs5914799	1,19	5,27E-04	0,239	23	56840879	C	T	550643	LOC550643	uncharacterized LOC550643
867	rs5914807	1,19	4,32E-04	0,240	23	56867944	G	T			
868	rs5960811	1,19	4,95E-04	0,239	23	56870079	G	A			
869	rs1930978	1,20	3,48E-04	0,241	23	56927132	T	C			
870	rs11091598	1,19	6,06E-04	0,240	23	56927696	G	T			
871	rs5914852	1,19	9,99E-04	0,214	23	56948766	C	T			
872	rs4379572	1,18	6,85E-04	0,243	23	56968844	G	A			
875	rs5942729	1,22	9,40E-04	0,155	23	108181279	A	G			
876	rs5942752	1,22	7,71E-04	0,155	23	108209528	G	A			
877	rs6642958	1,22	9,84E-04	0,154	23	108249271	G	A			
878	rs4825603	1,21	8,05E-04	0,177	23	117727895	C	G			
879	rs2495622	1,22	5,13E-04	0,176	23	117737020	G	T			
880	rs2495626	1,22	4,27E-04	0,174	23	117742946	T	C			
881	rs2256173	1,20	9,46E-04	0,174	23	117773617	C	T	3597	IL13RA1	interleukin 13 receptor, alpha 1
884	rs5919623	0,86	5,09E-04	0,411	23	144779048	C	G			
885	rs12862591	0,86	7,72E-04	0,406	23	144801782	G	T			
886	rs12861185	0,86	3,44E-04	0,403	23	144801952	G	C			
887	rs5965955	0,86	4,87E-04	0,406	23	144807588	A	T			

APPENDIX C: THE DETAILS OF TCF7L2 AND RBMS1 GENE ANALYSIS

Table 1 Strong signals on chromosome 2 for RBMS1 gene (RNA binding motif, single stranded interacting protein 1); Comparison of NHS and HPFS p values

Name of SNP (rsids)	Chr	Start position of SNP on Chromosome	Minor Allele	Major Allele	P value			
					Total	NHS	HPFS	Ratio (NHS/HPFS)
rs1020731	2	160852301	G	A	2.45E-07	1.97E-06	0.01397	7,091
rs6718526	2	160922421	T	C	2.74E-07	6.44E-06	0.006857	1,065
rs11693602	2	160932904	C	T	2.29E-06	1.14E-06		
rs7593730	2	160879700	T	C	2.55E-06	1.27E-06		
rs9287795	2	160918034	C	G	2.66E-06	1.64E-06		
rs4589705	2	160884382	T	A	2.75E-06	1.44E-06		
rs4077463	2	160874480	A	G	3.16E-06	1.43E-06		
rs10929982	2	160944523	C	T	4.55E-06	8.9E-06		
rs12692592	2	160871627	G	T	5.95E-06	4.87E-06		
rs7572970	2	160844902	A	G	5.97E-06	0.00015	0.009491	63
rs4664013	2	160892410	G	C	6.49E-06	0.000115	0.01334	116
rs12998587	2	160950541	T	C	1.19E-05	0.000302	0.01057	35
rs7587102	2	160967528	T	C	1.99E-05	0.000405	0.01354	33
rs4538150	2	160917573	G	A	2.18E-05	0.000794	0.008541	11
rs1020732	2	160852485	G	A	4.42E-05	0.00179	0.007844	4
rs9917155	2	160871805	C	A	0.000052	0.002043	0.008286	4
rs13009374	2	160973345	C	A	5.84E-05	0.00051	0.03073	60
rs4386280	2	160891041	A	G	7.99E-05	0.002389	0.01123	5
rs12692590	2	160861443	C	G	9.21E-05	0.001188	0.02429	20
rs10165319	2	160901051	T	C	0.000141	0.000334		
rs6742799	2	161025706	C	A	0.000239	5.14E-05		
rs4664323	2	160967931	C	T	0.000311	0.006433	0.01716	3
rs4664327	2	161002594	G	A	0.00174	0.005031		
rs10210349	2	160994684	C	T	0.001857	0.005916		
rs13008416	2	160925781	A	G	0.004055	0.01322		
rs11889328	2	160867938	A	G	0.007604	0.02807		
rs11694165	2	160903741	A	G	0.007669	0.0308		
rs12997772	2	160936449	T	C	0.008033	0.006992		
rs12692593	2	160905114	A	C	0.01672			

rs12692605	2	161023622	G	A	0.03111			
rs13397529	2	160944227	C	G	0.03386			
rs10176456	2	161026250	G	A	0.03476			

Table 2 P value for TCF7L2 gene in GWAS analysis.

SNP for TCF7L2 gene	P value		
	Total	NHS (Female)	HPFS (Male)
rs12255372	4.37E-10	9.72E-05	5.52E-07
rs12243326	6.12E-10	0.00018	3.47E-07
rs4132670	6.53E-10	0.000256	1.94E-07
rs7901695	8.18E-10	0.000153	5.83E-07
rs4506565	9.48E-10	0.000167	5.92E-07
rs11196208	0.000024	0.003665	0.002077
rs10885409	2.71E-05	0.005333	0.0015
rs11196205	3.49E-05	0.005594	0.001914
rs4074720	0.00013	0.01216	0.003286
rs7077039	0.000135	0.01502	0.002767
rs10787472	0.000138	0.01358	0.003086
rs6585201	0.000396	0.03572	0.002953
rs4073288	0.003882	>0.05	0.01396
rs7901275	0.004712	>0.05	0.002251
rs11196212	0.007301	0.04391	>0.05
rs7917983	0.007762	>0.05	0.002616
rs11196181	0.02668	>0.05	>0.05
rs12266632	0.03284	0.04173	>0.05
rs11196203	0.03392	>0.05	0.01642
Average	0.0062	0.012	0.003

APPENDIX D: THE DETAIL OF HAPLOVIEW ANALYSIS AND DISTRIBUTION DENSITY OF rs10739592

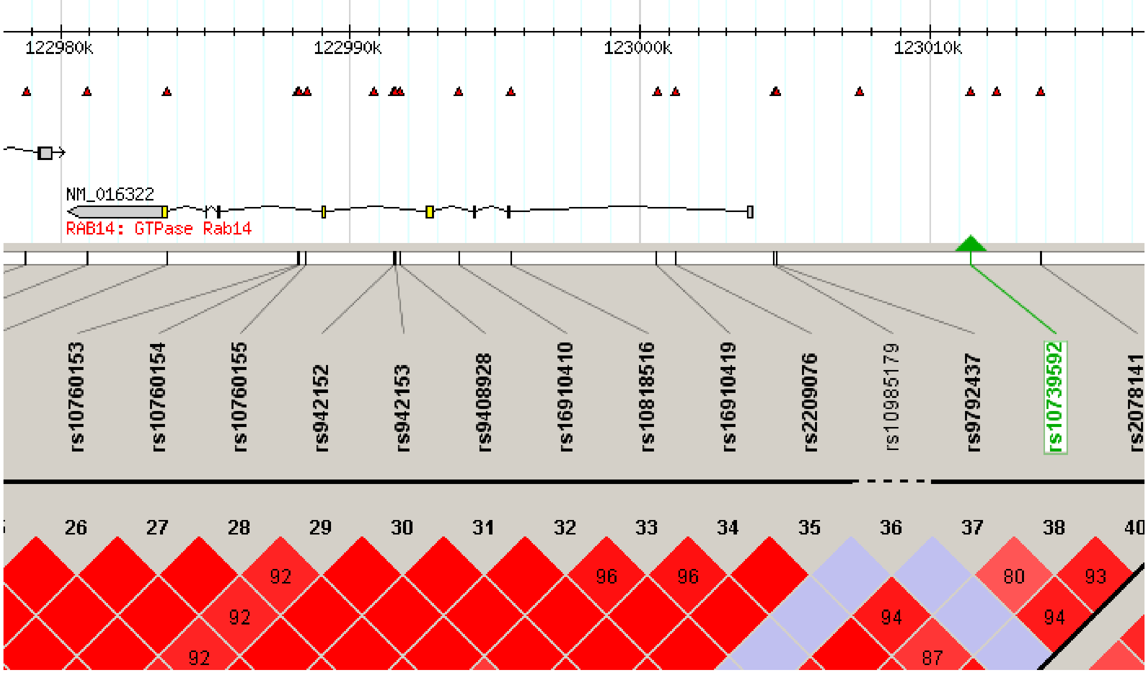


Figure 1.a Proximal region of rs10739592.

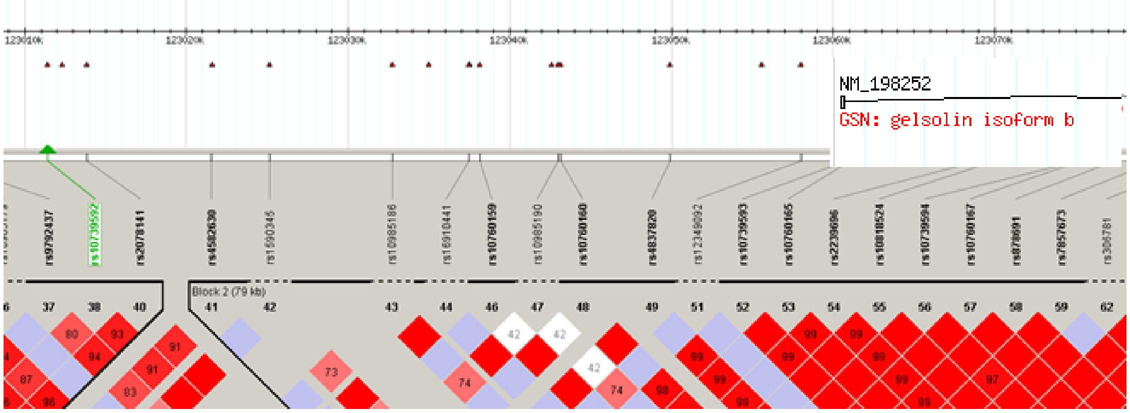


Figure 1.b Distal region of rs10739592.

Furthermore, we wanted to show the difference between male and female for rs10739592, so we plotted the distribution density of alleles for total population, male-male, female-female, and female-male comparison of control and case participants.

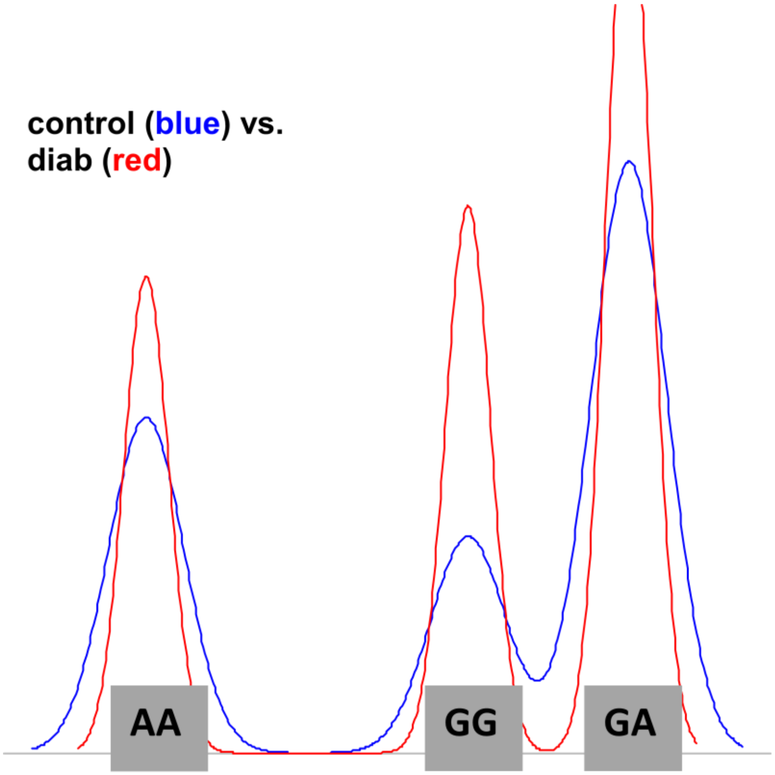


Figure 2.a Comparison of distribution density of control and case alleles for rs10739592.

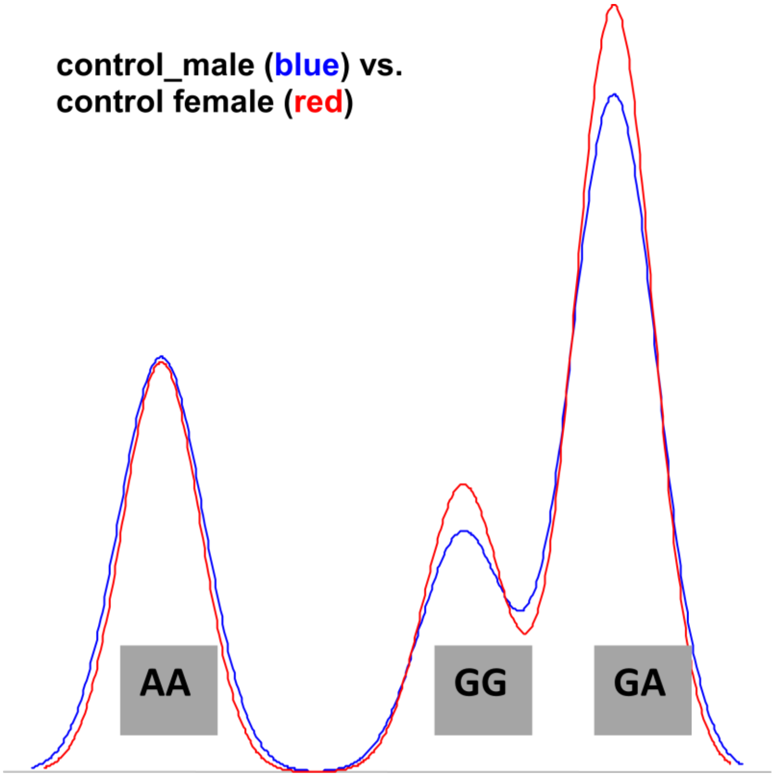


Figure 2.b Comparison of distribution density of control male and control female alleles for rs10739592.

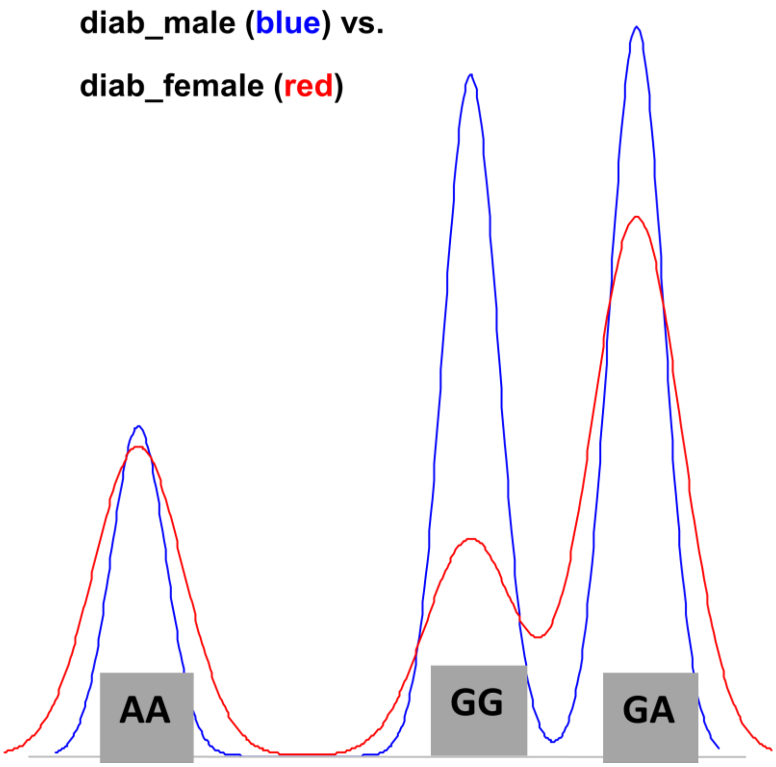


Figure 2.c Comparison of distribution density of diabetic male and diabetic female alleles for rs10739592.

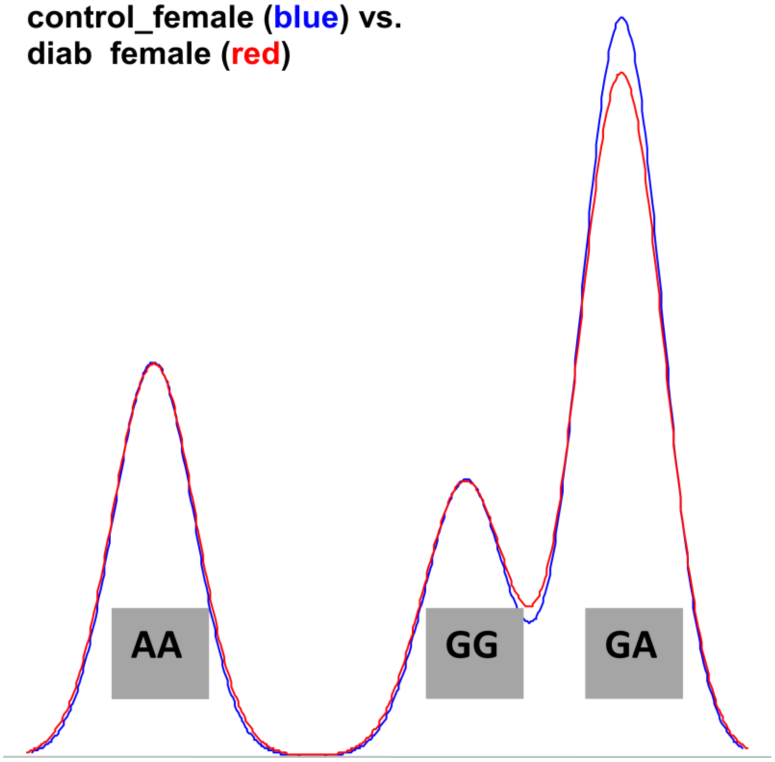


Figure 2.d Comparison of distribution density of control female and diabetic female alleles for rs10739592.

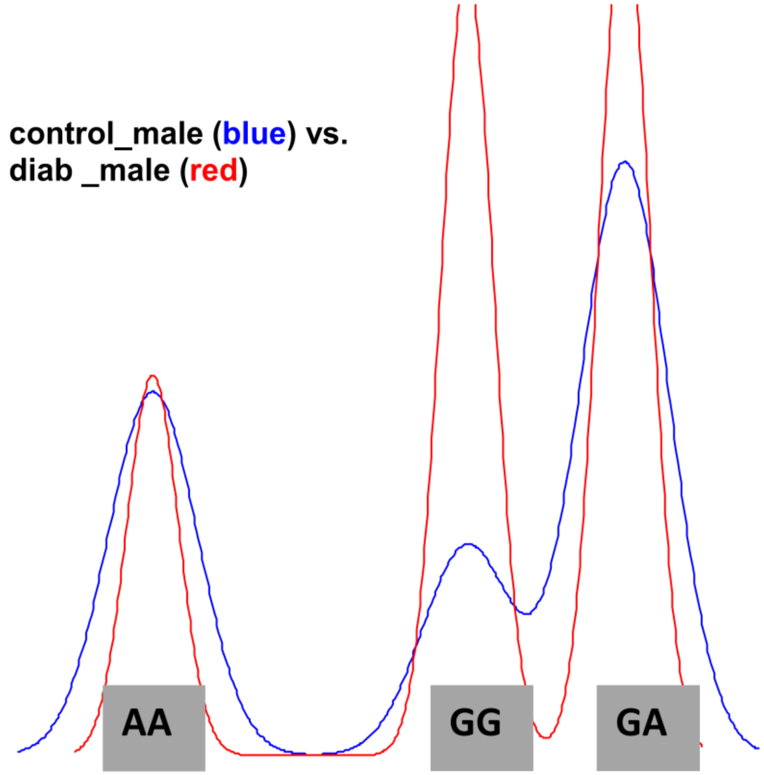


Figure 2.e Comparison of distribution density of control male and diabetic male alleles for rs10739592.

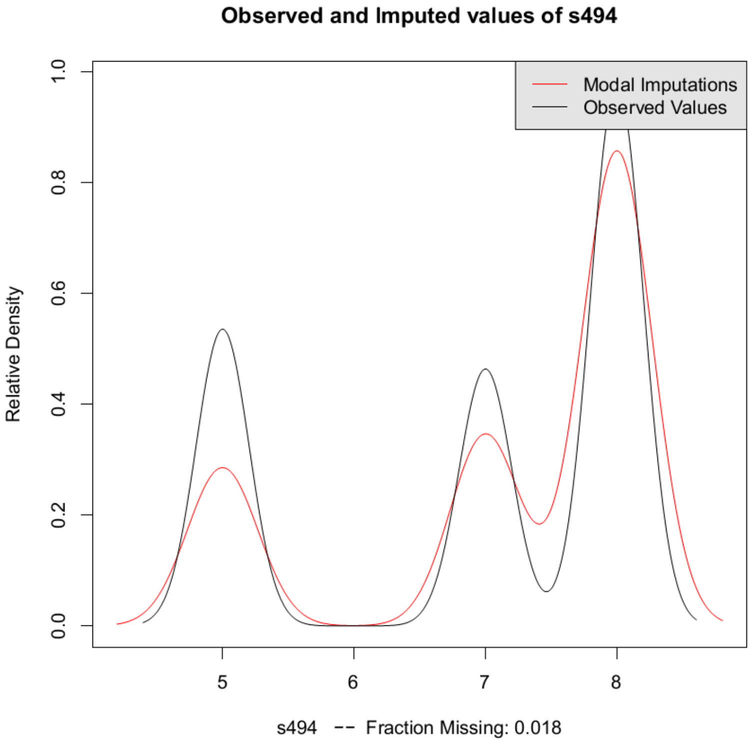


Figure 3 Relative density distribution of rs10739592 before and after imputation. P value of rs10739592 was 2.08E-14 before filling missing allele while after filling it was 3.13E-14. Difference in P value level and density profile of alleles suggest that filling missing allele does not have significant impact on the significance of rs10739592.

Table 1 Classification table for rs10739592 obtained with BLR analysis.[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2469	577	81.1
		Diab	1856	737	28.4
Overall percentage					56.9

a. The cut value is 0.5

APPENDIX E: THE DETAILS OF BLR ANALYSIS OF PHENOTYPE VARIABLES

Table 1 Classification table of study population at start level (without addition of any phenotype variable).

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	3046	0	100
		Diab	2593	0	0
Overall percentage					54.0

a. The cut value is 0.5

Table 2 Classification table for BMI only obtained with BLR analysis[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2167	879	71.1
		Diab	925	1667	64.3
Overall percentage					68.0

a. The cut value is 0.5

Table 3 Classification table for "familial diabetes history" only obtained with BLR analysis[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2382	664	78.2
		Diab	1382	1211	46.7
Overall percentage					63.7

a. The cut value is 0.5

Table 4 Classification table for “high blood pressure” only obtained with BLR analysis[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2412	634	79.2
		Diab	1413	1180	45.5
Overall percentage					63.7

a. The cut value is 0.5

Table 5 Classification table for the phenotype of “cholesterol” only obtained with BLR analysis[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2494	552	81.9
		Diab	1793	800	30.9
Overall percentage					58.4

a. The cut value is 0.5

Table 6 Classification table for the four phenotype of (BMI+FAMDB+HBP+CHOL) obtained with BLR analysis[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2403	643	78.9
		Diab	1008	1585	61.1
Overall percentage					70.7

a. The cut value is 0.5

APPENDIX F: THE DETAILS OF YOUDEN INDEX (YI) ANALYSIS FOR BODY MASS INDEX

Table 1 Youden Index for male in case 1, training group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,57	0,62	0,64	0,68	0,69	0,73
Negative Predictive Value	0,71	0,67	0,67	0,66	0,65	0,63
Likelihood Ratio +	1,52	1,89	1,98	2,44	2,56	3,14
Likelihood Ratio -	0,46	0,55	0,57	0,59	0,60	0,66
Sensitivity	0,77	0,63	0,61	0,54	0,52	0,43
Specificity	0,49	0,66	0,69	0,78	0,80	0,86
YI index	0,264	0,297	0,301	0,321	**0,320**	0,290

Table 2 Youden Index for male in case 1, test group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,56	0,63	0,64	0,66	0,69	0,73
Negative Predictive Value	0,71	0,70	0,68	0,66	0,67	0,65
Likelihood Ratio +	1,54	2,01	2,14	2,37	2,67	3,20
Likelihood Ratio -	0,50	0,52	0,55	0,61	0,60	0,65
Sensitivity	0,74	0,65	0,60	0,52	0,52	0,44
Specificity	0,52	0,68	0,72	0,78	0,81	0,86
YI index	0,259	0,325	0,322	0,301	**0,322**	0,302

Table 3 Youden Index for female in case 1, training group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,60	0,64	0,64	0,65	0,66	0,67
Negative Predictive Value	0,76	0,75	0,74	0,71	0,71	0,68
Likelihood Ratio +	1,84	2,15	2,22	2,31	2,32	2,46
Likelihood Ratio -	0,38	0,42	0,42	0,49	0,49	0,56
Sensitivity	0,78	0,72	0,72	0,65	0,64	0,57
Specificity	0,57	0,66	0,68	0,72	0,72	0,77
YI index	0,357	0,387	**0,393**	0,368	0,367	0,337

Table 4 Youden Index for male in case 1, test group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,61	0,63	0,63	0,66	0,66	0,68
Negative Predictive Value	0,75	0,72	0,72	0,71	0,70	0,67
Likelihood Ratio +	1,72	1,83	1,88	2,16	2,14	2,33
Likelihood Ratio -	0,36	0,43	0,43	0,45	0,46	0,54
Sensitivity	0,81	0,75	0,74	0,69	0,69	0,60
Specificity	0,53	0,59	0,61	0,68	0,68	0,74
YI index	0,339	0,339	**0,345**	0,372	0,365	0,342

Table 5 Youden Index for male in case 2, training group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,56	0,62	0,63	0,67	0,68	0,73
Negative Predictive Value	0,70	0,68	0,67	0,66	0,66	0,63
Likelihood Ratio +	1,48	1,88	1,98	2,40	2,53	3,07
Likelihood Ratio -	0,49	0,55	0,57	0,60	0,61	0,67
Sensitivity	0,76	0,64	0,61	0,54	0,52	0,42
Specificity	0,48	0,66	0,69	0,78	0,80	0,86
YI index	0,246	0,299	0,301	0,312	**0,312**	0,282

Table 6 Youden Index for male in case 2, test group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,62	0,66	0,67	0,70	0,73	0,76
Negative Predictive Value	0,73	0,67	0,66	0,65	0,66	0,64
Likelihood Ratio +	1,74	2,06	2,14	2,57	2,85	3,47
Likelihood Ratio -	0,39	0,54	0,55	0,57	0,56	0,61
Sensitivity	0,78	0,62	0,61	0,55	0,55	0,47
Specificity	0,55	0,70	0,72	0,79	0,81	0,86
YI index	0,332	0,321	0,323	0,336	**0,356**	0,334

Table 7 Youden Index for female in case 2, training group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,61	0,64	0,64	0,66	0,66	0,67
Negative Predictive Value	0,77	0,74	0,74	0,72	0,71	0,69
Likelihood Ratio +	1,84	2,09	2,16	2,29	2,30	2,47
Likelihood Ratio -	0,36	0,41	0,41	0,47	0,48	0,54
Sensitivity	0,80	0,73	0,73	0,66	0,66	0,58
Specificity	0,57	0,65	**0,66**	0,71	0,71	0,76
YI index	0,363	0,383	**0,390**	0,375	0,373	0,348

Table 8 Youden Index for male in case 2, test group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,59	0,63	0,63	0,65	0,65	0,66
Negative Predictive Value	0,74	0,73	0,72	0,70	0,70	0,66
Likelihood Ratio +	1,73	2,03	2,09	2,23	2,23	2,31
Likelihood Ratio -	0,43	0,44	0,46	0,52	0,52	0,61
Sensitivity	0,76	0,71	0,69	0,63	0,63	0,53
Specificity	0,56	0,65	0,67	0,72	0,72	0,77
YI index	0,319	0,361	**0,361**	0,348	0,348	0,300

Table 9 Youden Index for male in case 3, training group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,57	0,63	0,65	0,68	0,69	0,73
Negative Predictive Value	0,71	0,68	0,68	0,66	0,66	0,63
Likelihood Ratio +	1,51	1,95	2,09	2,46	2,59	3,02
Likelihood Ratio -	0,47	0,53	0,54	0,58	0,59	0,66
Sensitivity	0,77	0,65	0,62	0,55	0,53	0,44
Specificity	0,49	0,67	0,70	0,78	0,79	0,86
YI index	0,260	0,315	0,323	0,328	**0,327**	0,291

Table 10 Youden Index for male in case 3, test group

Threshold	25,0	26,0	26,3	27,0	**27,1**	28,0
Positive Predictive Value	0,57	0,60	0,60	0,66	0,69	0,77
Negative Predictive Value	0,72	0,66	0,64	0,64	0,65	0,64
Likelihood Ratio +	1,56	1,75	1,73	2,29	2,58	3,84
Likelihood Ratio -	0,47	0,62	0,65	0,66	0,64	0,67
Sensitivity	0,76	0,59	0,56	0,48	0,48	0,40
Specificity	0,52	0,66	0,68	0,79	0,81	0,90
YI index	0,273	0,252	0,235	0,270	**0,293**	0,297

Table 11 Youden Index for female in case 3, training group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,61	0,65	0,65	0,66	0,66	0,68
Negative Predictive Value	0,76	0,74	0,73	0,71	0,70	0,68
Likelihood Ratio +	1,83	2,14	2,19	2,30	2,29	2,51
Likelihood Ratio -	0,37	0,42	0,42	0,49	0,49	0,56
Sensitivity	0,79	0,72	0,72	0,65	0,64	0,57
Specificity	0,57	0,66	0,67	0,72	0,72	0,77
YI index	0,359	0,385	**0,389**	0,367	0,363	0,343

Table 12 Youden Index for male in case 3, test group

Threshold	25,0	26,0	**26,3**	27,0	27,1	28,0
Positive Predictive Value	0,57	0,59	0,61	0,63	0,64	0,63
Negative Predictive Value	0,77	0,76	0,76	0,74	0,75	0,70
Likelihood Ratio +	1,74	1,90	1,99	2,24	2,28	2,22
Likelihood Ratio -	0,38	0,41	0,41	0,45	0,44	0,57
Sensitivity	0,79	0,75	0,74	0,69	0,69	0,58
Specificity	0,55	0,60	0,63	0,69	0,70	0,74
YI index	0,337	0,356	**0,370**	0,383	0,388	0,319

APPENDIX G: DETAILS OF BINARY LOGISTIC REGRESSION ANALYSIS OF PHENOTYPE VARIABLES ON PREDICTION RATE AND AUC

Table 1 Incremental BLR analysis of seven phenotype variables; BMI (step 1), FAMDB (step 2), CHOL (step 3), HBP (step 4), activity (step 5), smoking (step 6) and alcohol (step 7).

Observed			Predicted		
			case		Percentage Correct
			control	diabetes	
Step 1	case	control	2130	868	71.0
		diabetes	905	1652	64.6
	Overall Percentage				68.1
Step 2	case	control	2130	868	71.0
		diabetes	905	1652	64.6
	Overall Percentage				68.1
Step 3	case	control	247	501	83.3
		diabetes	1174	1383	54.1
	Overall Percentage				69.8
Step 4	case	control	2366	632	78.9
		diabetes	993	1564	61.2
	Overall Percentage				70.7
Step 5	case	control	2361	637	78.8
		diabetes	958	1599	62.5
	Overall Percentage				71.3
Step 6	case	control	2380	618	79.4
		diabetes	977	1580	61.8
	Overall Percentage				71.3
Step 7	case	control	2354	644	78.5
		diabetes	941	1616	63.2
	Overall Percentage				71.5

a. The cut value is 0.5

Table 2 Area under curve values for phenotype variables.

Test Result Variable(s)	Area	Std. Error (a)	Asymp-totic Sig.(b)	Asymptotic 95% Confidence Interval	
				Upper Bound	Lower Bound
BMI+FAMDB+HBP+CHOL	.770	.006	.000	.758	.782
BMI+FAMDB+HBP+CHOL+ Activity+smoking+alcohol	.776	.006	.000	.764	.788

The test results variable(s): Phenotype has at least one tie between the positive actual state and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

APPENDIX H: THE DETAILS OF BINARY LOGISTIC REGRESSION ANALYSIS OF EACH CHROMOSOME

Table 1 Summary of classification table for each chromosome shows NPV, PPV and AUC.

Chr	SNPs between	SNP number after excluding of high missing alleles	# SNPs	NPV (% correct for control)	PPV (% correct for diabetes)	Overall Correction percentage	AUC
1	1-86	75	86	71.3	51.1	62.1	0.667
2	87-196	105	110	71.7	53.1	63.1	0.675
3	197-253	53	57	72.1	50.0	61.9	0.652
4	254-293	36	40	74.8	40.0	58.8	0.612
5	294-352	52	59	72.2	46.6	60.4	0.632
6	353-401	42	49	72.5	43.1	58.9	0.624
7	402-439	32	38	71.7	44.9	59.4	0.631
8	440-467	26	28	72.0	41.9	58.1	0.613
9	468-496	23	29	74.4	40.3	58.7	0.65
10	497-558	57	62	72.2	45.4	59.9	0.624
11	559-602	40	44	72.4	45.0	59.8	0.631
12	603-675	66	73	71.8	46.9	60.4	0.637
13	676-689	12	14	80.0	25.8	55.1	0.562
14	690-725	35	36	73.6	43.0	59.5	0.61
15	726-751	25	26	74.0	40.3	58.5	0.595
16	752-763	9	12	80.1	25.0	54.8	0.562
17	764-773	8	10	85.6	19.7	55.3	0.57
18	774-809	34	36	75.7	35.9	57.4	0.596
19	810-815	6	6	82.6	23.4	55.4	0.552
20	816-831	13	16	77.7	30.9	56.2	0.577
21	832-857	23	26	76.1	35.9	57.6	0.594
22	858-862	5	5	92.6	11.1	55.1	0.561
23	863-886	21	25	84.0	22.7	55.8	0.55

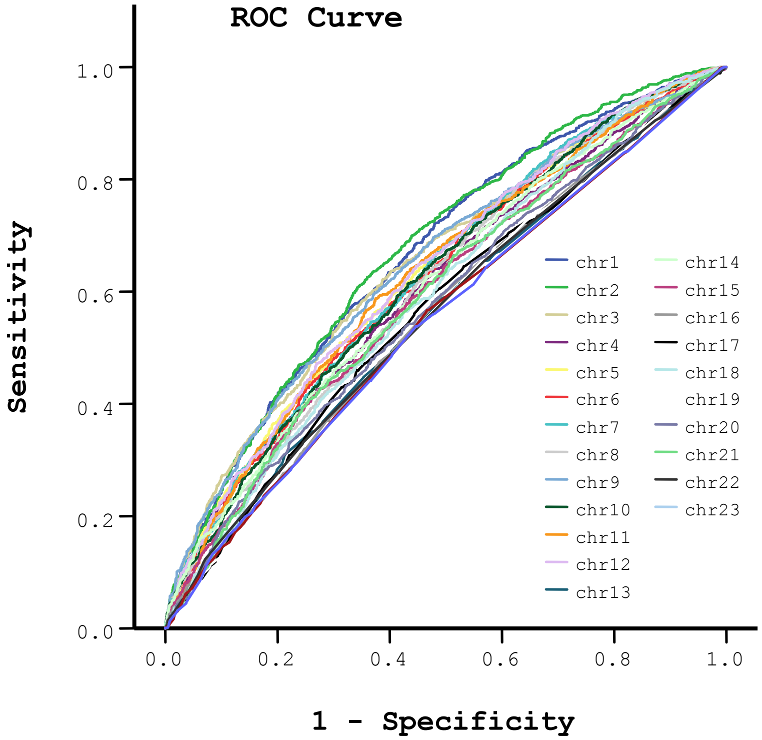


Figure 1 ROC Curve for 23 chromosomes.

APPENDIX I: THE DETAILS OF THE COMPARISON OF TRAINING AND TEST GROUPS

1. CASE 1, Statistical analysis (Chi square) of phenotype variables and binary logistic regression analysis of training and test groups.

Table 1 FAMDB comparison in case 1 between training and test groups.

Groups	FAMDB		Total	Chi Square P value
	Non exist	exist		
Training	721	393	1114	0.11
Test	3043	1482	4525	
Total	3764	1875	5639	

Table 2 HBP comparison in case 1 between training and test groups.

Groups	HBP		Total	Chi Square P value
	normal	high		
Training	749	365	1114	0.642
Test	3076	1449	4525	
Total	3825	1814	5639	

Table 3 CHOL comparison in case 1 between training and test groups.

Groups	CHOL		Total	Chi Square P value
	normal	high		
Training	854	260	1114	0.611
Test	3433	1092	4525	
Total	4287	1352	5639	

Table 4 BMI comparison in case 1 between training and test groups.

Groups	BMI		Total	Chi Square P value
	normal	obese		
Training	593	521	1114	0.24
Test	2499	2026	4525	
Total	3092	2547	5639	

Table 5 Gender comparison in case 1 between training and test groups.

Groups	GENDER		Total	Chi Square P value
	male	female		
Training	494	620	1114	0.146
Test	1897	2628	4525	
Total	2391	3248	5639	

Table 6 Case comparison in case 1 between training and test groups.

Groups	CASE		Total	Chi Square P value
	control	diabetes		
Training	593	521	1114	0.568
Test	2453	2072	4525	
Total	3046	2593	5639	

Table 7 Binary logistic regression analysis of training group in case 1.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	2307	146	94.05
Diabetes	162	1910	92.18
Overall Percentage			93.19

Table 8 Binary logistic regression analysis of test group in case 1.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	555	38	93.59
Diabetes	55	466	89.44
Overall Percentage			91.65

2. CASE 2, Statistical analysis (Chi square) of phenotype variables and binary logistic regression analysis of training and test groups.

Table 9 FAMDB comparison in case 2 between training and test groups.

Groups	FAMDB		Total	Chi Square P value
	Non exist	exist		
Training	3017	1497	4514	0.777
Test	747	378	1125	
Total	3764	1875	5639	

Table 10 HBP comparison in case 2 between training and test groups.

Groups	HBP		Total	Chi Square P value
	normal	high		
Training	3064	1450	4514	0.887
Test	761	364	1125	
Total	3825	1814	5639	

Table 11 CHOL comparison in case 2 between training and test groups.

Groups	CHOL		Total	Chi Square P value
	normal	high		
Training	3434	1080	4514	0.876
Test	853	272	1125	
Total	4287	1352	5639	

Table 12 BMI comparison in case 2 between training and test groups.

Groups	BMI		Total	Chi Square P value
	normal	obese		
Training	2470	2044	4514	0.738
Test	622	503	1125	
Total	3092	2547	5639	

Table 13 Gender comparison in case 2 between training and test groups.

Groups	GENDER		Total	Chi Square P value
	male	female		
Training	1921	2593	4514	0.661
Test	470	655	1125	
Total	2391	3248	5639	

Table 14 Case comparison in case 2 between training and test groups.

Groups	CASE		Total	Chi Square P value
	control	diabetes		
Training	2444	2070	4514	0.713
Test	602	573	1125	
Total	3046	2593	5639	

Table 15 Binary logistic regression analysis of training group in case 2.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	2323	121	95.05
Diabetes	155	1915	92.51
Overall Percentage			93.89

Table 16 Binary logistic regression analysis of test group in case 2.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	548	54	91.03
Diabetes	53	470	89.87
Overall Percentage			90.49

3. CASE 3, Statistical analysis (Chi square) of phenotype variables and binary logistic regression analysis of training and test groups.

Table 17 FAMDB comparison in case 3 between training and test groups.

Groups	FAMDB		Total	Chi Square P value
	Non exist	exist		
Training	3017	1497	4514	0.777
Test	747	378	1125	
Total	3764	1875	5639	

Table 18 HBP comparison in case 3 between training and test groups.

Groups	HBP		Total	Chi Square P value
	normal	high		
Training	3064	1450	4514	0.887
Test	761	364	1125	
Total	3825	1814	5639	

Table 19 CHOL comparison in case 3 between training and test groups.

Groups	CHOL		Total	Chi Square P value
	normal	high		
Training	3434	1080	4514	0.876
Test	853	272	1125	
Total	4287	1352	5639	

Table 20 BMI comparison in case 3 between training and test groups.

Groups	BMI		Total	Chi Square P value
	normal	obese		
Training	2470	2044	4514	0.738
Test	622	503	1125	
Total	3092	2547	5639	

Table 21 Gender comparison in case 3 between training and test groups.

Groups	GENDER		Total	Chi Square P value
	male	female		
Training	1921	2593	4514	0.661
Test	470	655	1125	
Total	2391	3248	5639	

Table 22 Case comparison in case 3 between training and test groups.

Groups	CASE		Total	Chi Square P value
	control	diabetes		
Training	2444	2070	4514	0.713
Test	602	523	1125	
Total	3046	2593	5639	

Table 23 Binary logistic regression analysis of training group in case 3.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	2294	128	94.72
Diabetes	145	1947	93.07
Overall Percentage			93.95

Table 24 Binary logistic regression analysis of test group in case 3.

Groups	Predicted		
	control	diabetes	Percentage correct
Control	572	52	91.67
Diabetes	66	435	86.83
Overall Percentage			89.51

APPENDIX J: THE DETAILS OF BINARY LOGISTIC REGRESSION ANALYSIS OF SNPS DEPENDING ON THE PAR VALUES

We used 235 SNPs at first with PAR values are equal, or greater than 10%.

Table 1 Classification table of 235 SNPs with PAR values are >= 10%. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2323	723	76.3
		Diab	861	1732	66.8
Overall percentage					71.9

a. The cut value is 0.5

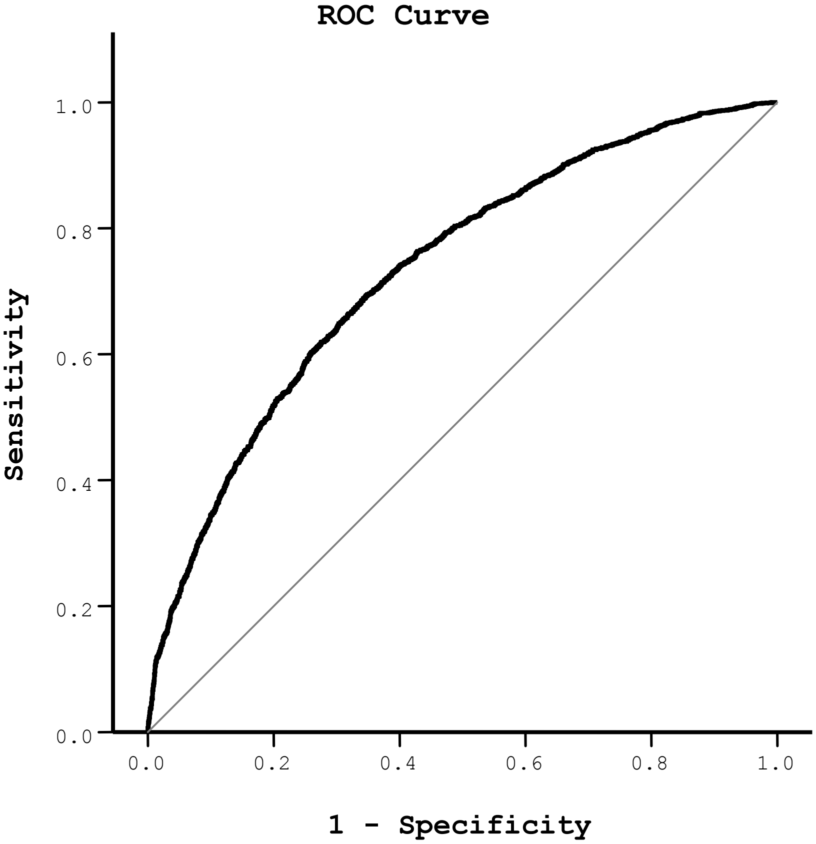


Figure 1 ROC curve for 235 SNPs with PAR values are equal, or greater than 10%.

Table 2 Area under curve for 235 SNPs with PAR values are >= 10%.

Area	Std. Error (a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Upper Bound	Lower Bound
.797	.006	.000	.786	.809

a Under the nonparametric assumption
b Null hypothesis: true area = 0.5

Then we used 485 SNPs with PAR values less than 10% to understand whether PAR is the best method for SNP selection for better prediction of risk SNPs for diabetes.

Table 3 Classification table of 485 SNPs with PAR values < 10%. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2563	483	84.1
		Diab	540	2053	79.2
Overall percentage					81.9

a. The cut value is 0.5

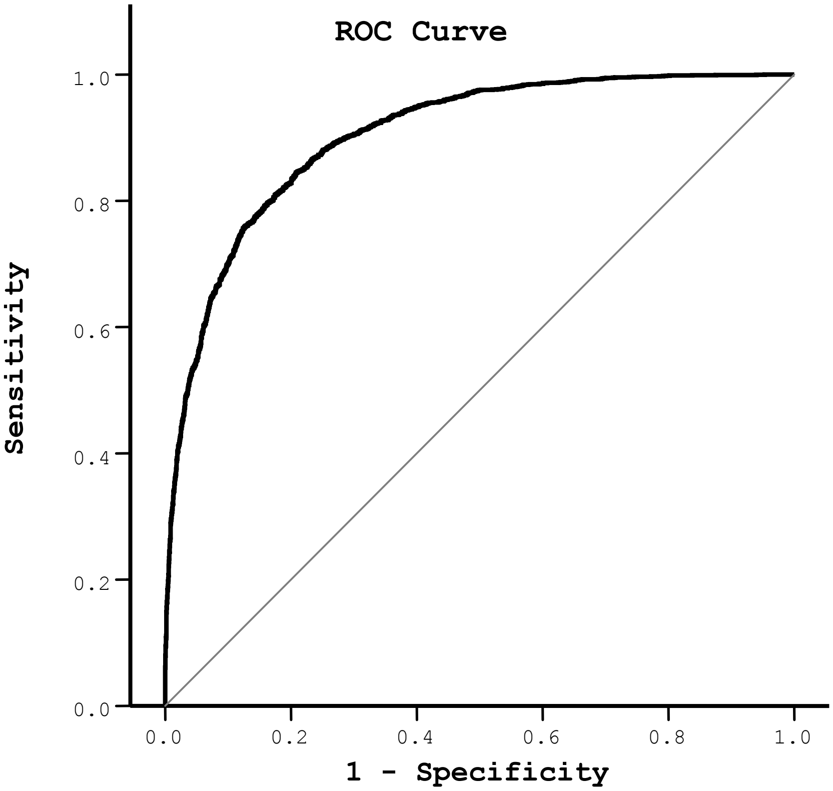


Figure 2 ROC curve for 485 SNPs with PAR values < 10%.

Table 4 Area under curve for 485 SNPs with PAR values are < 10%.

Area	Std. Error (a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Upper Bound	Lower Bound
.902	.004	.000	.895	.910

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

We want to investigate more deeply using PAR paradigm, so we separated SNPs according to their PAR values either negative (decreased risk of diabetes) or positively (increased risk of diabetes). SNPs which have negative PAR value were 358 ranging from -15.56 to -2.72 (average (-9.14). I divided set of SNPs into two group from middle (n=179) and analyzed separately and together.

Table 5 Classification table of PAR negative high group (n=179, ranging from -15.56 to -9.15). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2278	768	74.8
		Diab	961	1632	62.9
Overall percentage					69.3

a. The cut value is 0.5

Table 6 Classification table of PAR negative low group (n=179, ranging from -9.13 to -2.72). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2278	768	74.8
		Diab	852	1741	67.1
Overall percentage					71.3

a. The cut value is 0.5

Table 7 Classification table of PAR negative total (n=358).[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2363	683	77.6
		Diab	736	1857	71.6
Overall percentage					74.8

a. The cut value is 0.5

SNPs which have positive PAR value were 362 ranging from 3.41 to 26.31 (average (8.18). We divided set of SNPs into two groups each containing 181 SNPs and analyzed separately and in combination.

Table 8 Classification table of PAR positive high group (n=181, ranging from 26.31 to 7.82). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2296	750	75.4
		Diab	933	1660	64.0
Overall percentage					70.2

a. The cut value is 0.5

Table 9 Classification table of PAR positive low group (n=181, ranging from 7.80 to 3.41). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2316	730	76.0
		Diab	962	1631	62.9
Overall percentage					70.0

a. The cut value is 0.5

Table 10 Classification table of PAR positive total (n=358). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2473	573	81.2
		Diab	711	1882	72.6
Overall percentage					77.2

a. The cut value is 0.5

Table 11 Classification table of PAR positive high group (n=181) plus negative high group (n=179). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2451	595	80.5
		Diab	656	1937	74.7
Overall percentage					77.8

a. The cut value is 0.5

Table 12 Classification table of PAR low positive group (n=181) plus low negative group (n=179). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2485	561	81.6
		Diab	640	1953	75.3
Overall percentage					78.7

a. The cut value is 0.5

Table 13 Classification table of High negative plus low positive group (n=179 plus n=181). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2472	574	81.2
		Diab	695	1898	73.2
Overall percentage					77.5

a. The cut value is 0.5

Table 14 Classification table of High negative plus Low positive group (n=179 plus n=181). [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2452	594	80.5
		Diab	619	1974	76.1
Overall percentage					78.5

a. The cut value is 0.5

Table 15 Classification Table of the 798 SNPs by BLR analysis. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2762	284	90.7
		Diab	282	2311	89.1
Overall percentage					90.0

a. The cut value is 0.5

Table 16 Area under the curve for various PAR scenarios.

Test Result Variable(s)	Area	Std. Error (a)	Asymp-totic Sig.(b)	Asymptotic 95% Confidence Interval	
				Upper Bound	Lower Bound
PAR_neg_high	.766	.006	.000	.754	.779
PAR_neg_low	.782	.006	.000	.770	.794
PAR_neg_total	.832	.005	.000	.822	.843
PAR_Positive_low	.772	.006	.000	.760	.784
PAR_positive_high	.767	.006	.000	.755	.779
PAR_positive_total	.854	.005	.000	.844	.863
PAR_pos_high_plus_neg_high	.856	.005	.000	.846	.866
PAR_lowpos_low_neg	.869	.005	.000	.860	.879
All (798) SNPs	.965	.002	.000	.949	.959
PAR_high neg plus low pos	.860	.005	.000	.851	.870
PAR_low neg plus high pos	.865	.005	.000	.855	.874

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

APPENDIX K: INDIVIDUAL AND ADDITIVE EFFECTS ON BINARY LOGISTIC REGRESSION ANALYSIS OF SNP GROUPS DEPENDING ON THEIR P VALUES

A. Individual Analysis of Each P Value Group

Table 1 Classification Table of SNPs with P values lower than <1.0E-06 (n=10) [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2283	763	75.0
		Diab	1590	1003	38.7
Overall percentage					58.3

a. The cut value is 0.5

Table 2 Classification Table of SNPs with P values between >1.0E-06 - <1.0E-05 (n=17) [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2315	731	76.0
		Diab	1669	924	35.6
Overall percentage					57.4

a. The cut value is 0.5

Table 3 Classification Table of SNPs with P values between >1.0E-05 - <1.0E-04 (n=91) [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2223	823	73.0
		Diab	1109	1484	57.2
Overall percentage					65.7

a. The cut value is 0.5

Table 4 Classification Table of SNPs with P values between >1.0E-04 - <1.0E-03 (n=604) [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2708	338	88.9
		Diab	358	2235	86.2
Overall percentage					87.7

a. The cut value is 0.5

B. Incremental (Additive) Analysis of Groups

Table 1 Classification table of SNPs with P values lower than <1.0E-06 (n=10) in BLR analysis. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2283	763	74.7
		Diab	1590	1003	38.7
Overall percentage					58.3

a. The cut value is 0.5

Table 2 Classification Table of SNPs with P values lower than <1.0E-05 (n=27) in BLR analysis.[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2262	828	72.8
		Diab	1055	1175	45.3
Overall percentage					60.2

a. The cut value is 0.5

Table 3 Classification Table of SNPs with P values lower than <1.0E-04 (n=118) in BLR analysis.[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2262	784	74.3
		Diab	1055	1538	59.3
Overall percentage					67.4

a. The cut value is 0.5

Table 4 Classification Table of SNPs with P values lower than <1.0E-03 (n=798) in BLR analysis.[a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2762	284	90.7
		Diab	282	2311	89.1
Overall percentage					90.0

a. The cut value is 0.5

APPENDIX L: THE DETAILS OF THE EFFECT OF CUT-OFF VALUE ON THE CLASSIFICATION AND AUC IN BINARY LOGISTIC REGRESSION ANALYSIS

Table 1 Classification table of 798 SNP in BLR analysis for cut-off value of 0.5. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2721	325	
		Diab	324	2269	
Overall percentage					

a. The cut value is 0.5

Table 2 Classification table of 798 SNP in BLR analysis for cut-off value of 0.6. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2862	184	94.0
		Diab	422	2171	83.7
Overall percentage					89.3

a. The cut value is 0.6

Table 3 Classification table of 798 SNP in BLR analysis for cut-off value of 0.7. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2949	97	96.8
		Diab	602	1991	76.8
Overall percentage					87.6

a. The cut value is 0.7

Table 4 Classification table of 798 SNP in BLR analysis for cut-off value of 0.8. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	2997	49	98.4
		Diab	842	1751	67.5
Overall percentage					84.2

a. The cut value is 0.8

Table 5 Classification table of 798 SNP in BLR analysis for cut-off value of 0.9. [a]

Observed			Predicted		
			case		Percentage Correct
			control	diab	
Step 1	case	Control	3024	22	99.3
		Diab	1229	1364	52.6
Overall percentage					77.8

a. The cut value is 0.9

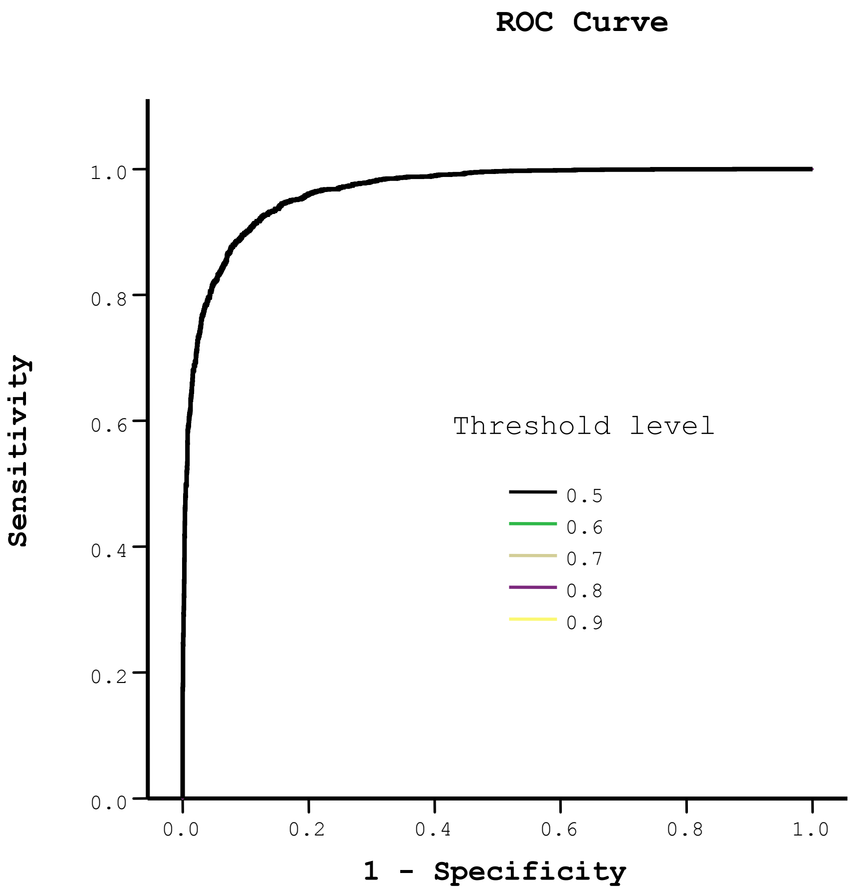


Figure 1 ROC curve of 798 SNPs depending on the various threshold levels. Whereas threshold level changes, but AUC does not change. Because, ROC curve lines overlap each other, only black line could be seen.

Table 6 Area Under the Curve of 798 SNPs depending on the various threshold levels

Threshold level	Area
0.5	0.965
0.6	0.965
0.7	0.965
0.8	0.965
0.9	0.965

APPENDIX M: DIFFERENCES OF PHENOTYPE VARIABLES AMONGST THE STUDIES IN THE LITERATURE

Abbrevition of the Study	Age (y)	Control	Number diabetic patients	Hypertension
Framingham I [27]	50±9.7	2377	255	
Framingham II [25]	144 patients < 50 (mean 49.30) 302 patients > 50 (mean 66.07)	3471	446	
Malmö Study [26]	Not known	12,210	2063	
Botnia Study [26]	Not known	2632	138	
Rotterdam Study [29]	69.5±0.11	5221	1287	Control 30.5% Diabetes 46.9-52.9%
DESIR 2 [23]	Men diabetes 50±9 Men no diabetes 47±10 Woman diabetes 52±8 Woman no diabetes 47±10	3614	203	Men diabetes 62% Men no diabetes 39% Woman diabetes 62% Woman no diabetes 28%
Whitehall II [28]	49	5233	302	
Our study	Control subjects 57.1±7.7 Diabetic subjects 57.4±7.7	3046	2593	Men diabetes 41% Men no diabetes 21.8% Woman diabetes 49.2% Woman no diabetes 20.1%

Abbrevition of the Study	Body Mass Index	Familial Diabetes History
Framingham I [27]		
Framingham II [25]		
Malmö Study [26]		
Botnia Study [26]		
Rotterdam Study [29]		
DESIR 2 [23]	Men diabetes 27.5±4 Men no diabetes 25.1±3 Woman diabetes 29.2±5.1 Woman no diabetes 23.7±3.8	Men diabetes 28 of 140 (20%) Men no diabetes 312 of 1723 (18%) Woman diabetes 27 of 63 (43%) Woman no diabetes 368 of 1891 (19%)
Whitehall II [28]		
Our study	Men diabetes 27.9±4 Men no diabetes 25.2±2.8 Woman diabetes 29.9±5.8 Woman no diabetes 25.4±4.8	Men diabetes 481 of 1114 (43.2%) Men no diabetes 272 of 1277 (21.3%) Woman diabetes 730 of 1479 (49.4) Woman no diabetes 392 of 1769 (22.2%)

VITA

Husamettin Gul, M.D., MSc.,

Associate Professor

Pharmacology & Toxicology

Born: 1969

Married, three children

Present Position: Assoc.Prof. of Pharmacology & Toxicology in Gulhane Military Medical Academy, School of Medicine, Dept. of Toxicology

Education

- 1993 M.D., Medicine, Gulhane Military Medical Academy, School of Medicine
- 1999 Specialist (equivalent to Ph.D.), Pharmacology and Toxicology, Gulhane Military Medical Academy, School of Medicine
- 1999-2003, Postdoctoral Research, Dept. of Pharmacology and Toxicology, Gulhane Military Medical Academy, School of Medicine
- 2003-2004, Assistant Professor, Gulhane Military Medical Academy, School of Medicine
- 2004-present Assoc. Prof., Department of Pharmacology and Toxicology, Gulhane Military Medical Academy, School of Medicine
- 2001-2003, MS in Informatics, Middle East Techical University
- 2003- Ph.D. education in Medical Informatics, Middle East Techical University
- 2010 January – 2011 July: The University of North Carolina (UNC), USA
- 2011 - Gulhane School of Medicine, Dept. of Pharmacology & Toxicology

Permanent Mailing Address:

Gulhane Military Medical Academy, School of Medicine,
Dept. Of Pharmacology & Toxicology
06018 Etlik Ankara, TURKEY
Phone: +90 (312) 304 4784
e-mail: hgul23@gata.edu.tr

Dissertation of Specialist: “Functional characterization of serotonin (5-HT) receptors in human various isolated arteries”

Affiliations:

- Turkish Pharmacological Society
- Turkish Medical Informatics Society
- Society of Toxicology (SOT)

Scientific Interest

- Toxicological screening of human samples by Gas Chromatography/ Mass Spectrometry (GC/MS)
- Drug analysis by GC, GC/MS and LC-MS/MS
- Experimental diabetes, diabetic neuropathy
- Functional studies in human isolated arteries, especially evaluation of contractile 5-HT receptors.

Experience in Using Medical Apparatus

- Gas Chromatography/ Mass Spectrometry (GC/MS)
- Gas Chromatography, FID
- Isolated organ bath techniques

Awards and Honors

2005, Turkish Diabetes Foundation, Best Article Awards **(3th)**

2005, 2003-2004 Novartis Pharmacological Article Awards in Turkey **(2th)**

2003, National Cukurova Coloproctology Symposium Oral Presentation Awards **(1th)**

2002, GMMA Scientific Article Awards **(3th)**

2002, Turkish Pharmacological Society - Servier Young Investigator Awards

2001, Diabetes Foundation, Best Article Awards **(Honorable Mention)**

2000, 1999-2000 Novartis Pharmacological Article Awards in Turkey **(2th)**

1999, 25th National Congress of Physiology, Young Investigator Awards

1997, Turkish Diabetes Foundation, Prof.Celal Oker Research and Development Fund, Diabetes Research Awards **(3th)**

Papers Published in International Journals (PubMed)

1. **Gül. H**, Odabaşı Z, Yıldız O, Ozata M, Deniz G, Işımer A, Vural O. Beneficial effect of trythropin releasing hormone on neuropathy in diabetic rats. Diabetes Research Clinical Practice 1999, 44(2): 93-100.
2. **Gul H**, Yıldız O , Dogrul A, Yesilyurt O, Isimer A. The interaction between IL-1b and morphine: possible mechanism of the deficiency of morphine-induced analgesia in diabetic mice. Pain, 2000, 89(1):39-45.
3. **Gul H** and Yıldız O. Amplification of sumatriptan-induced contractions with phenylephrine, histamine and KCl in the isolated human mesenteric artery: in-vitro evidence for sumatriptan-induced mesenteric ischaemia. Naunyn Schmiedeberg's Arch. Pharmacol., 2002, 366: 254-261.
4. **Gul H**, Yıldız O, Sımsek A, Balkan M, Ersoz N, Cetiner S, Isımer A, Sen D. Pharmacological characterization of contractile serotonergic receptors in human isolated mesenteric artery. J. Cardiovasc. Pharmacol., 2003, 41: 307-315.
5. Yıldız O, **Gul H**, Kilciler M, Onguru O, Ozgok IY, Aydın A, Isımer A, Harmankaya AC. Increased vasoconstrictor reactivity and decreased endothelial function in high grade varicocele: functional and morphological study. Urological Research, 2003, 31(5): 323-328.
6. Dogrul A, **Gul H**, Akar A, Yıldız O, Bilgin F, Guzeldemir E. Topical Cannabinoid Analgesia: Synergy With Spinal Sites. Pain, 2003, 105(1-2): 11-6.
7. Yesilyurt O, Dogrul A, **Gul H**, Seyrek M, Kusmez O, Ozkan Y, Yıldız O.Topical cannabinoid enhances topical morphine antinociception. Pain, 2003, 105(1-2): 303-308.
8. Yagci G, **Gul H,** Simsek A, Varol N, Onguru O, Yıldız O, Balkan M, Zeybek N, Sen D. Beneficial effects of N-acteylcysteine on sodium taurocholate-induced pancreatitis in rats. Journal of Gastroenterology. 2004, 39(3): 268-276.
9. Dogrul A, **Gul H**, Yildiz O, Bilgin F, Guzeldemir ME. Cannabinoids blocks tactile allodynia in diabetic mice without attenuation of its antinociceptive effect. Neurosci Lett. 2004; 368(1):82-6.
10. Yildiz O, Seyrek M, Un I, **Gul H**, Uzun M, Yildirim V, Bolu E. Testosterone Relaxes Human Internal Mammary Artery In Vitro. J Cardiovasc Pharmacol. 2005 Jun;45(6):580-585.
11. Yildiz O, Seyrek M, Un I, **Gul H**, Candemir G, Yildirim V. The Relationship Between Risk Factors and Testosterone-Induced Relaxations in Human Internal Mammary Artery. J Cardiovasc Pharmacol. 2005 Jan;45(1):4-7.
12. Dogrul A, Gulmez SE, Deveci MS, **Gul H**, Ossipov MH, Porreca F, Tulunay FC. The local antinociceptive actions of nonsteroidal antiinflammatory drugs in the mouse radiant heat tail-flick test. Anesth Analg. 2007 Apr;104(4):927-35.
13. Yildiz O, Ulusoy HB, Seyrek M, **Gul H**, Yildirim V. Dexmedetomidine Produces Dual alpha(2)-Adrenergic Agonist and alpha(1)-Adrenergic Antagonist Actions on Human Isolated Internal Mammary Artery. J Cardiothorac Vasc Anesth. 2007 Oct;21(5):696-700.
14. Gokhan Eraslan, Sahan Saygi, Dinc Essiz, Abdurrahman Aksoy, **Husamettin Gul**, Enis Macit. Evaluation of aspect of some oxidative stress parameters using vitamin E, proanthocyanidin and N-acetylcysteine against exposure to cyfluthrin in mice. Pesticide Biochemistry and Physiology 88 (2007) 43–49.

15. Odabasi E, **Gul H**, Macit E, Turan M, Yildiz O. Lipophilic components of different therapeutic mud species. J Altern Complement Med. 2007 Dec;13(10):1115-8.

16. Akay C, Kalman S, Dündaröz R, Sayal A, Aydin A, Ozkan Y, **Gül H**. Serum aluminium levels in glue-sniffer adolescent and in glue containers. Basic Clin Pharmacol Toxicol. 2008 May;102(5):433-6.

17. Ulusoy HB, **Gul H,** Seyrek M, Yildiz O, Ulku C, Yildirim V, Kuralay E, Celik T, Yanarates O. The concentration-dependent contractile effect of methylene blue in the human internal mammary artery: a quantitative approach to its use in the vasoplegic syndrome. J Cardiothorac Vasc Anesth. 2008 Aug;22(4):560-4.

18. Gulec M, Ogur R, **Gul H**, Korkmaz A, Bakir B. Investigation of vasoactive ion content of herbs used in hemorrhoid treatment in Turkey. Pak J Pharm Sci. 2009 Apr;22(2):187-92.

19. Yucel O, Kunak ZI, Macit E, Gunal A, Gozubuyuk A, **Gul H**, Genc O. Protective efficiacy of taurine against pulmonary edema progression: experimental study. J Cardiothorac Surg. 2008 Oct 28;3:57.

20. Celik T, Iyisoy A, **Gul H**, Isik E. Clopidogrel resistance: a diagnostic challenge. Int J Cardiol. 2009 Jan 9;131(2):267-8.

21. Dogrul A, **Gul H**, Yesilyurt O, Ulas UH, Yildiz O. Systemic and spinal administration of etanercept, a tumor necrosis factor alpha inhibitor, blocks tactile allodynia in diabetic mice. Acta Diabetol. 2011 Jun;48(2):135-42.

22. Karapirli M, Kizilgun M, Yesilyurt O, **Gul H**, Kunak ZI, Akgul EO, Macit E, Cayci T, Gulcan Kurt Y, Aydin I, Yaren H, Seyrek M, Cakir E, Yaman H. Simultaneous determination of cyclosporine A, tacrolimus, sirolimus, and everolimus in whole-blood samples by LC-MS/MS. ScientificWorldJournal. 2012;2012:571201. Epub 2012 May 2.

23. **Gul H**, Uysal B, Cakir E, Yaman H, Macit E, Yildirim AO, Eyi YE, Kaldirim U, Oztas E, Akgul EO, Cayci T, Ozler M, Topal T, Oter S, Korkmaz A, Toygar M, Demirbag S. The protective effects of ozone therapy in a rat model of acetaminophen-induced liver injury. Environ Toxicol Pharmacol. 2012 Jul;34(1):81-6.

24. Nakamura J, **Gul H**, Tian X, Bultman SJ, Swenberg JA. Detection of PIGO-deficient cells using proaerolysin: a valuable tool to investigate mechanisms of mutagenesis in the DT40 cell system. PLoS One. 2012;7(3):e33563. Epub 2012 Mar 12.

25. Agackiran Y, Gul H, Gunay E, Akyurek N, Memis L, Gunay S, Sirin YS, Ide T. The efficiency of proanthocyanidin in an experimental pulmonary fibrosis model: comparison with taurine. Inflammation. 2012 Aug;35(4):1402-10.

26. Lu K, **Gul H**, Upton PB, Moeller BC, Swenberg JA. Formation of hydroxymethyl DNA adducts in rats orally exposed to stable isotope labeled methanol. Toxicol Sci. 2012 Mar;126(1):28-38. Epub 2011 Dec 8.

Papers Published in National Journals

1. **Gül H.**, O. Yıldız, Z. Odabaşı, M. Özata, G. Deniz, A. Işımer. Deneysel diyabetik nöropatide Thyrotropin-releasing hormone (TRH) uygulamasının elektrofizyolojik etkileri. Year Book of Turkish Diabetology 1997-1998, 13: 157-165, 1997

2. S. Şenöz, **Gül H.**, M. Özata, O. Yıldız. Streptozosin diyabetik sıçanlarda aminoguanidin, enalapril, asetilsalisilik asit ve doğal antioksidanların nefropati tedavisindeki yeri. Türk Diabet Yıllığı, 1998-1999, sayfa:190-196.

3. **Hüsamettin Gül.** Hipertansiyon Tedavisinde Diüretikler. Türkiye Klinikleri, 2007;3(18):17-27.

4. Müjgan Güler, Enis Macit, Bengü Bakdık, **Hüsamettin Gül**, Bülent Çiftçi, Halil Yaman, Ebru Ünsal, Yurdanur Erdoğan, Şahan Saygı. Serum Rifampisin Seviyesinin Antitüberküloz Tedaviye Yanıttaki Rolü. Solunum Hastalıkları, 2007; 18: 14-19.

5. Cemal AKAY, İsmail Tuncer DEĞİM, Ahmet SAYAL, Ahmet AYDIN, Yalçın ÖZKAN, **Hüsamettin GÜL.** Rapid and Simultaneous Determination of Acetylsalicylic Acid, Paracetamol, and Their Degradation and Toxic Impurity Products by HPLC in Pharmaceutical Dosage Forms. Turk J Med Sci 2008; 38 (2): 167-173

Poster Presentations in International Congress

1. **Gül H.**, O. Yıldız, Z. Odabaşı, M. Özata, G. Deniz, A. Işımer. Electrophysiologic effects of thyrotropin releasing hormone (TRH) on experimental diabetic neuropathy (Poster). Abstract Book,. 3th Military Ballkan Congress, Athens, Greece 1998, s:305.

2. **Gül H.**, O. Yıldız, Z. Odabaşı, M. Özata, G. Deniz, A. Işımer.. The effects of thyrotropin releasing hormone in experimental diabetic neuropathy (Poster). 34th Annual Meeting of the EASD, 8th 12th September 1998, Barcelona, Spain, Diabetologia, 41: Suppl. 1, A271.

3. **Gül H.**, O.Yıldız, A. Doğrul, T. Ide, A. Işımer. Role of IL-1 b and TNF-a in the deficiency in-morphine-induced analgesia indiabetic mice (Poster). 2nd European Congress of Pharmacology. Fundamental & Clinical Pharmacology, 1999, 13/Suppl.1; PM27.

4. **Gül H.**, O. Yıldız, N. Ersöz, S. Çetiner, A. Işımer. Characterizatıon of contractile 5-HT receptors in human isolated mesenteric artery, 5th Balkan Military Medical Congress, 25-28 September, 2000, Ankara, Türkiye. (Poster), Abstract book, page: 146 –P147.

5. **Gül H.**, O. Yıldız, M. Balkan, C. Yigitler, S. Çetiner, A. Işımer, Amplifıcation of responses to sumatriptan by various agoniısts in human isolated mesenteric artery, 5th Balkan Military Medical Congress, 25-28 September, 2000, Ankara, Türkiye. (Poster), Abstract book, page: 114 – P49.

6. **Gül H.**, O. Yıldız, N. Ersöz, S. Çetıner, A. Işımer. Variability of 5-HT1B/1D receptor-induced contractile responses in human isolated mesenteric artery, 5th Balkan Military Medical Congress, 25-28 September, 2000, Ankara, Türkiye. (Poster), Abstract book, page: 227 - P385.

7. **Gül H.**, O. Yıldız, M. Kilciler, A. Aydın, Y. Özgök, A. Işımer, A.Ç. Harmankaya. Does endothelial dysfunction develop in high grade varicocele?. 6th Balkan Military Medical Congress, 1-3 October, 2001, Plovdiv, Bulgaria. (Poster), Abstract book, P212.

8. **Gül H.**, Yildiz O, Kilciler M, Onguru O. Increased vasoconstrictor reactivity and decreased endothelial function in high grade of varicocele: functional and morphological study. 14th World Condress of Pharmacology, July 7-12, 2002, San Francisco, (Poster), Pharmacologist, Vol 44, No: 2; Suppl 1 pA30, No: 26.9.

9. Bicak, M., **Gul, H.**, Ozkan, M., Yildiz, O., Ekiz, K., Saygı, Ş., Demirci, N., “Comparison the effects of steroidal therapy by measuring exhaled carbon monoxide (CO) in bronchial asthma and COPD.” 6th EACPT Congress, The Proceedings of the Sixth Congress of the European

Association for Clinical Pharmacology and Therapeutics, P.No:373, page:172, Istanbul, Turkiye, 24-28 June 2003.

10. **Gul, H.**, Yesilyurt, O., Dogrul, A., Yildiz, O., Bilgin, F., Guzeldemir, E., “Cannabinoid analgesia is preserved in diabetic mice” 8th Congress of Balkan Military Medical Committee, Abstract book, page:212, PS.I.78, Cluj Napoca, Romania, 2003.

11. Saygi, S., Ates, Y., Aslan, M., Tuzun, A., **Gul, H.**, “Herbal remedies induced hepatotoxicity” 8th Congress of Balkan Military Medical Committee, Abstract book, page:350, PS.II.30, Cluj Napoca, Romania, 2003.

12. Yildiz, O., Seyrek, M., **Gul, H.**, Bolu, E., Ozal, E., Kuralay, E., Yildirim, V., “Testosterone relaxes human internal mammary and radial arteries in vitro” 8th Congress of Balkan Military Medical Committee, Abstract book, page:566, PS.III.61, Cluj Napoca, Romania, 2003.

13. Sahan Saygi, **Gul H**, Enis Macit, Bilgin Comert, “Nonfatal acute poisoning with high doses of amitriptylline, acetaminophen and codeine” 9th Congress of Balkan Military Medical Committee, Abstract book, page:378, P-247, Antalya, Turkey, 21-24 June, 2004.

14. **Gul H**, Oguzhan Yildiz, Melik Seyrek, Ismail Un, “Protein kinase C activation may contribute to testosterone-induced relaxation in human internal mammary artery” 9th Congress of Balkan Military Medical Committee, Abstract book, page:422, P-291, Antalya, Turkey, 21-24 June, 2004.

15. Oguzhan Yildiz, Melik Seyrek, **Gul H**, Ismail Un, “Testosterone relaxes human internal mammary artery in vitro” 9th Congress of Balkan Military Medical Committee, Abstract book, page:477, P-346, Antalya, Turkey, 21-24 June, 2004.

16. Yildiz O, Seyrek M, **Gul H**, Ulusoy HB, Yildirim V. Direct contracting effect of dexmedetomidine on human internal mammary artery. 11th Congress of Balkan Military Medical Committee, Abstract book, page: 291, P-094D, Athens, Greece 2006. (poster)

17. S. Saygi, E. Macit, **H. Gul**, Z. Sezer. Retrospective evaluation of toxicologic analysis performed at GATA Medical School Department of Analytical Toxıcology. 6th International Congress of Turkish Society of Toxicology. Abstract Book page:129, poster no: P-054, November 2-5, 2006, Antalya TURKEY

18. Mujgan Z.Guler, Enis Macit, Bengu Baktik, **Husamettin Gul**, Bulent Ciftci, Halil Yaman, Ebru Unsal, Yurdanur Erdogan, Sahan Saygi. The role of serum rifampicin level on antituberculosis treatment outcome. European Respiratory Society, ERS Munich 2006 Annual Congress, September 2-6, 2006, poster no: E4890, page: 848s.

19. Enis Macit, Emin Ozgur Akgul, Z. Ilker Kunak, Recai Ogur, Husamettin Gul, I. Tayfun Uzbay. Determining of Chemicals with Expectorant Activity in a Traditional Medicine Used for Treatment of Chronic Obstructive Diseases in Turkey by Gas Chromatography/Mass Spectrometry. Third Symposium on the Practical Applications of Mass Spectrometry in the Biotechnology & Pharmaceutical Industries. Abstract Book page:48, poster no: P-11, September 6-8, 2006. La Jolla, California, USA.

20. Oguzhan Yıldız, Hasan Basri Ulusoy, Melik Seyrek, **Husamettin Gul**, Vedat Yıldırım. Dexmedetomidine-induced contraction in human internal mammary artery: involvement of alfa-adrenoceptor subtypes. Page: 178-179; Poster No.: P110209. Acta Pharmacologica Sinica, 2006 July, Supplement 1: 1-489, Abstracts of the 15th World Congress of Pharmacology, July 2-7, 2006 Beijing, China.

21. Enis Macit, Şahan Saygı, **Husamettin Gul**. Comparison of the solid phase and liquid-liquid extraction method for toxicological screening using Gas Chromatography / Mass Spectrometry. Page: 362; Poster No.: P310041. Acta Pharmacologica Sinica, 2006 July, Supplement 1: 1-489, Abstracts of the 15th World Congress of Pharmacology, July 2-7, 2006 Beijing, China.

22. Enis Macit, Şahan Saygı, **Husamettin Gul**. Superiority of liquid-liquid extraction for toxicological screening in Gas Chromatography / Mass Spectrometry. Page: 364-365; Poster No.: P310054. Acta Pharmacologica Sinica, 2006 July, Supplement 1: 1-489, Abstracts of the 15th World Congress of Pharmacology, July 2-7, 2006 Beijing, China.

23. G. Sezer, Y. Tekol, Z. Sezer, **H. Gul**, I.T. Uzbay. Effect of Venlafaxine On Rat Formalin Test. European Journal of Pain 11(S1) (2007) page:S175, poster no: 395.

24. **GUL Husamettin**, MACIT Enis, UYSAL Bulent, TOPAL Turgut, YILDIZ Oguzhan. Gypsophila L. saponin induced toxicity is not associated with the increased erythrocyte membrane permeability. 12th Congress of Balkan Military Medical Committee, Abstract book, page:290, June 24-28, 2007; Poiana Brasov, Romania.

25. YILDIZ Oguzhan, SEYREK Melik, **GUL Husamettin**, KURALAY Erkan, CELIK Turgay, YILDIRIM VEDAT. Contractile Effect of Methylene Blue in Human Internal Mammary Artery. 12th Congress of Balkan Military Medical Committee, Abstract book, page:368, June 24-28, 2007; Poiana Brasov, Romania.

26. Enis Macit, **Hüsamettin Gül**, Tayfun Uzbay, Abdurrahman Aksoy, Hakan Çermik, Oğuzhan Yıldız, Zeki İlker Kunak, Turgut Topal, Gyposphila L. Induces Hepatocellular Damage In Balb/c Mice. PREP 2008: 21st International Symposium, Exhibit and Workshops on Preparative/Process Chromatography, Ion Exchange, Adsorption/Desorption Processes and Related Separation Techniques. Poster, P-110-M, Page: 45, June 15-18, 2008 San Jose, California, USA

27. **Hüsamettin Gül**, Abdurrahman Aksoy, Enis Macit, Ahmet Aydın, Oğuzhan Yıldız, Turgut Topal, Gökhan Eraslan, Ahmet Sayal, Şahan Saygı. Gyposphila L. Containing Diet Reduces Glutathione Peroxidase Activity In Balb/c Mice. PREP 2008: 21st International Symposium, Exhibit and Workshops on Preparative/Process Chromatography, Ion Exchange, Adsorption/Desorption Processes and Related Separation Techniques. Poster, P-125-M, Page: 50-51, June 15-18, 2008 San Jose, California, USA

28. **H. Gul**, K. Lu, P.Upton, J.A. Swenberg. Formaldehyde-induced Hydroxymethyl Adducts in Rats Exposed to Isotope Labeled Methanol. 2011 SOT Annual Meeting, March 6-10, 2011, Washington D.C., Abstract Number:1691.

Poster Presentations in National Congress

1. S. Şenöz , **H. Gül**, M. Özata, O Yıldız. Streptozosin diyabetik sıçanlarda aminoguanidin, enalapril, asetilsalisilik asit ve doğal antioksidanların nefropati tedavisindeki yeri (Poster). 34 ncü Ulusal Diyabet Kongresi ve 3 ncü Uluslararası Obezite Sempozyumu Kongresi, 1-3 Mayıs 1998 Ankara, Program ve Özet Kitapçığı, sayfa:89.

2. **Gül, H.**, "İnsan internal meme arterinde serotonin reseptör alttiplerinin fonksiyonel olarak tanımlanması", 25 nci Ulusal Fizyoloji Kongresi, Bildiri Özetleri Kitapçığı, S9 (Sözel Bildiri), sayfa:24, Fırat Universitesi, ELAZIĞ, 1999.

3. O. Yıldız, **H. Gül**, U. Demirkılıç, A. Işımer. İnsan internal meme arterinde 5-HT1B/1D ve 5-HT2A reseptör cevaplarının değişkenliği. XV. Ulusal Farmakoloji Kongresi, Manavgat – Antalya, 1-5 Kasım 1999. (Poster). Özet Kitabı, s. P – 06-01, 1999.

4. **H. Gül**, O. Yıldız, U. Demirkılıç, A. Işımer, İnsan internal meme arterinde spontan ve agonist ile indüklenen fazik kasılmaların diltiazem ile önlenmesi. XV. Ulusal Farmakoloji Kongresi, Manavgat – Antalya, 1-5 Kasım 1999. (Poster). Özet Kitabı, s. P –06-01 , 1999.

5. **H. Gül**, O. Yıldız, M. Balkan, N. Ersöz, S. Çetiner, A. Işımer, İzole insan mezenter arterinde 5-HT 1B/1D reseptör cevapları: Mezenter iskemisine yeni bir bakış. 17. Gastroenteroloji Haftası, 3-8 Ekim 2000, Antalya (Sözlü). The Turkish Journal of Gastroenterology, Volume 11/ Supplement 1, Oral presentations, p:17, S:44.

6. **H. Gül**, O. Yıldız, A.Şimşek, N. Ersöz, S. Çetiner, A. Işımer, D. Şen. İnsan izole mezenter arterinde çeşitli önkasıcı maddeler varlığında 5-HT1B/1D reseptör cevaplarının amplifikasyonu. XVI. Ulusal Farmakoloji Kongresi, Kuşadası, 1-5 Ekim 2001. (Poster). Özet Kitabı, PY-65, 2001.

7. **H. Gül**, O. Yıldız, A.Şimşek, N. Ersöz, S. Çetiner, A. Işımer, D. Şen. İnsan izole mezenter arterinde serotonin reseptörlerinin farmakolojik olarak tanımlanması. XVI. Ulusal Farmakoloji Kongresi, Kuşadası, 1-5 Ekim 2001. (Poster). Özet Kitabı, PY-66, 2001.

8. **H. Gül**, O. Yıldız, A.Şimşek, N. Ersöz, S. Çetiner, A. Işımer, D. Şen. İnsan izole mezenter arterinde 5-HT 1B/1D reseptör cevaplarının değişkenliği. XVI. Ulusal Farmakoloji Kongresi, Kuşadası, 1-5 Ekim 2001. (Poster). Özet Kitabı, PY-67, 2001.

9. **H. Gül**, O. Yıldız, M. Kilciler, A. Aydın, Y. Özgök, A. Işımer, A.Ç. Harmankaya. İnsan varikosel veninde yüksek grade'lerde endotel fonksiyonlarında bozulma mı meydana geliyor? XVI. Ulusal Farmakoloji Kongresi, Kuşadası, 1-5 Ekim 2001. (Poster). Özet Kitabı, PY-71, 2001.

10. Ateş, Y., Aslan, M., Saygı, Ş., Tüzün, A., **Gül., H.**, "Bitkisel ilaç kullanımına bağlı karaciğer toksisitesi", Ulusal Toksikoloji ve Klinik Toksikoloji Sempozyumu, Türkiye Klinikleri Farmakoloji, Cilt I, Sayı I, Sayfa:122, P.No:71, Dokuz Eylül Üniversitesi Tıp Fakültesi, 8-9 Mayıs 2003, Balçova–İZMİR, 2003.

11. Saygi,Ş., Cömert, B., **Gül, H**. İntihar amaçlı trisiklik antidepresan kullanımına bağlı fatal ve nonfatal 2 olgu. Ulusal Toksikoloji ve Klinik Toksikoloji Sempozyumu, Türkiye Klinikleri Farmakoloji, Cilt I, Sayı I, Sayfa:104, P.No:30, Dokuz Eylül Üniversitesi Tıp Fakültesi, 8-9 Mayıs 2003, Balçova–İZMİR, 2003.

12. Yıldız, O., Seyrek, M., **Gül, H.**, "Testosterone relaxes human internal mammary and radial arteries in vitro", 17th National Congress of Pharmacology, 1st Clinical Pharmacology Symposium, Abstract Book, s.179, P–102, Belek – ANTALYA, 17-21 October, 2003.

13. Dogrul, A., **Gül, H.**, Yıldız, O., Bilgin, F., Güzeldemir, E., "Cannabinoids blocks tactile allodynia in diabetic mice without attenuation of its antinociceptive effect", 17th National Congress of Pharmacology, 1st Clinical Pharmacology Symposium, Abstract Book, s.211, P–129, Belek – ANTALYA, 17-21 October, 2003.

14. Doğrul, A, **Gül, H**, Akar A, Bilgin F, Yıldız, O, Güzeldemir ME. Topikal Kanabinoid Analjezi. 6. Ulusal Ağrı Kongresi, Program & Özet Kitabı, sayfa:150, P59. 16-19 Mayıs 2003 (Poster), İstanbul.

15. Balkan M, **Gül H**, Yıldız O, Çetiner S, Işımer A, Şen D. Serotonin reseptörlerinin mezenter iskemisindeki rolü. I. Çukurova Koloproktoloji Sempozyum ve Kursu Konuşma Metinleri ve Bildiri Kitabı, Sözlü Sunum (S 01), Sayfa:125, Çukurova Üniversitesi, ADANA, 2003.

16. Kaya Kuru, **Hüsamettin Gül**, Güney Gürsel, Kemal Arda, Erkan Mumcuoğlu, Nazife Baykal. "Sağlık Hizmetlerinde Kaynakların Doğru Kullanımında Bilgisayar Benzetim Yönteminin Kullanılması: Bir Poliklinik Çalışması", 2. Ulusal Tıp Bilişimi Kongresi / Medical Informatics '05 Turkey Bildiri Kitabı, Sözel Bildiri, s.14-20, 17-20 Kasım 2005, Belek-ANTALYA.

17. **Hüsamettin Gül**, Kaya Kuru, Güney Gürsel, Özkan Yıldız. "Elektronik Reçetenin Avantajları, Kullanımında Karşılaşılabilecek Sorunlar ve Giderilme Yöntemleri", 2. Ulusal Tıp Bilişimi Kongresi / Medical Informatics '05 Turkey Bildiri Kitabı, Sözel Bildiri, s.134-139, 17-20 Kasım 2005, Belek-ANTALYA.

18. **Hüsamettin Gül**, Kaya Kuru, Güney GÜRSEL. "Çok Yönlü, Kullanıcı Tarafından Yönlendirilebilen, Esnek ve Ölçeklenebilir, Bir Toksikoloji Karar Destek Aracı". 3. Ulusal Tıp Bilişimi Kongresi/ Medical Informatics '06 Turkey Bildiri Kitabı, Sözel Bildiri, s.118-119, 16-19 Kasım 2006, ANTALYA.

19. Seyrek, M., Yıldız, O., **Gül, H.**, Yıldırım, V., "İnsan İnternal Meme Arterinde Testosteronun Oluşturduğu Gevşeme Cevabının Kardiyak Risk Faktörleriyle İlişkisi", 18. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.165, P–26, İZMİR, 2005.

20. **Gül, H.**, Seyrek, M., Yıldız, O., "Elektronik Reçete", 18. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.369, P–155, İZMİR, 2005.

21. Abdurrahman AKSOY, Şahan SAYGI, Gökhan ERASLAN, **Hüsamettin GÜL**, Enis MACİT, Hasan AYÇİÇEK, Halil YAMAN. Determination of Histamine Levels of Mackerel Fish (Secomber Scomburs) and Fermented Turkish Sausage Consumed In Ankara (Ankara Piyasalarında Tüketime Sunulan Uskumru Balığı ve Sucuklarda Histamin Düzeyleri). Birinci Ulusal Veteriner Farmakoloji ve Toksikoloji Kongresi, Kongre Kitabı, sayfa 320-321; 22-24 Eylül 2005, ANKARA.

22. Seyrek, M., Yıldız, O., **Gül, H.**, Yıldırım, V., "İnsan İnternal Meme Arterinde Testosteronun Oluşturduğu Gevşeme Cevabının Kardiyak Risk Faktörleriyle İlişkisi", 18. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.165, P–26, İZMİR, 28 Eylül – 1 Ekim, 2005.

23. **Gül, H.**, Seyrek, M., Yıldız, O., "Elektronik Reçete", 18. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.369, P–155, İZMİR, 28 Eylül – 1 Ekim, 2005.

24. Oğuzhan Yıldız, Hasan Basri Ulusoy, Melik Seyrek, **Hüsamettin Gül**, Vedat Yıldırım. Deksmedetomidinin İnsan İnternal Meme Arterine Alfa-1 Adrenerjik Agonist ve Alfa-2 Adrenerjik Antagonist Olarak Çift Yönlü Etkisi. 40. Türk Anesteziyoloji ve Reanimasyon Kongresi, 25-28.10.2006, Istanbul, Türk Anesteziyoloji ve Reanimasyon Derneği Dergisi, Eylül-Ekim 2006, Cilt:34, Supplement:1, P.141, sayfa:127-128.

25. **Hüsamettin Gül**, Ergin Soysal, Nazife Baykal. "Tıp Doktorlarına Yönelik Elektronik İlaç Karar Destek Sistemi Geliştirilmesi ve Karşılaşılan Güçlükler", 19. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.354, P–082, Trabzon, 24-27 Ekim, 2007.

26. Hasan Basri Ulusoy, Oğuzhan Yıldız, **Hüsamettin Gül**, Melik Seyrek, Vedat Yıldırım, Suat Doğancı, Erkan Kuralay, Turgay Çelik. "Metilen Mavisinin İnsan İnternal Meme Arterinde Kasıcı Etkisi: Vazoplejik Sendromda Kullanımına Kantitatif Bir Yaklaşım". 19. Ulusal Farmakoloji Kongresi, Özet Kitabı, s.377, P–103, Trabzon, 24-27 Ekim, 2007.

27. **Hüsamettin Gül**, Ergin Soysal, Nazife Baykal. “Türkiye’de Mobil İlaç Karar Destek Sistemi Geliştirilmesi İçin Bir Çalışma”. 4. Ulusal Tıp Bilişimi Kongresi/ Medical Informatics ’07 Turkey Bildiri Kitabı, s.123-124, 15-18 Kasım 2007, ANTALYA.

Reviewers

1. Diabetologia
2. TAF Preventive Medicine Bulletin (TAF Prev Med Bull)
3. Renal Failure
4. Turkish Journal of Biology
5. Digestive Disease and Sciences
6. Food and Chemical Toxicology

Citation

My papers has been referenced 300 times by 06.27.2013