

Genome-Wide Linkage Scan to Identify Loci Associated with Type 2 Diabetes and Blood Lipid Phenotypes in the Sikh Diabetes Study

Dharambir K. Sanghera^{1*}, Latonya F. Been¹, Sarju Ralhan², Gurpreet S. Wander², Narinder K. Mehra³, Jai Rup Singh⁴, Robert E. Ferrell⁵, Mohammed I. Kamboh⁵, Christopher E. Aston¹

1 Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **2** Section of Cardiology, Hero DMC Heart Institute, Ludhiana, Punjab, India, **3** Department of Transplant Immunology and Immuno-genetics, All India Institute of Medical Sciences and Research, New Delhi, India, **4** Central University, Bathinda, Punjab, India, **5** Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Abstract

In this investigation, we have carried out an autosomal genome-wide linkage analysis to map genes associated with type 2 diabetes (T2D) and five quantitative traits of blood lipids including total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, very low-density lipoprotein (VLDL) cholesterol, and triglycerides in a unique family-based cohort from the Sikh Diabetes Study (SDS). A total of 870 individuals (526 male/344 female) from 321 families were successfully genotyped using 398 polymorphic microsatellite markers with an average spacing of 9.26 cM on the autosomes. Results of non-parametric multipoint linkage analysis using S_{all} statistics (implemented in Merlin) did not reveal any chromosomal region to be significantly associated with T2D in this Sikh cohort. However, linkage analysis for lipid traits using QTL-ALL analysis revealed promising linkage signals with $p \leq 0.005$ for total cholesterol, LDL cholesterol, and HDL cholesterol at chromosomes 5p15, 9q21, 10p11, 10q21, and 22q13. The most significant signal ($p = 0.0011$) occurred at 10q21.2 for HDL cholesterol. We also observed linkage signals for total cholesterol at 22q13.32 ($p = 0.0016$) and 5p15.33 ($p = 0.0031$) and for LDL cholesterol at 10p11.23 ($p = 0.0045$). Interestingly, some of linkage regions identified in this Sikh population coincide with plausible candidate genes reported in recent genome-wide association and meta-analysis studies for lipid traits. Our study provides the first evidence of linkage for loci associated with quantitative lipid traits at four chromosomal regions in this Asian Indian population from Punjab. More detailed examination of these regions with more informative genotyping, sequencing, and functional studies should lead to rapid detection of novel targets of therapeutic importance.

Citation: Sanghera DK, Been LF, Ralhan S, Wander GS, Mehra NK, et al. (2011) Genome-Wide Linkage Scan to Identify Loci Associated with Type 2 Diabetes and Blood Lipid Phenotypes in the Sikh Diabetes Study. PLoS ONE 6(6): e21188. doi:10.1371/journal.pone.0021188

Editor: Sunil K Ahuja, South Texas Veterans Health Care System, United States of America

Received: December 21, 2010; **Accepted:** May 22, 2011; **Published:** June 16, 2011

Copyright: © 2011 Sanghera et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was supported by NIH grants (K01TW006087 and R01DK082766) funded by the Fogarty International Center (FIC) and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and a seed grant from University of Oklahoma Health Sciences Center, Oklahoma City, OK and also in part by the Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Dharambir-sanghera@ouhsc.edu

Introduction

Type 2 diabetes (T2D) is a major public health problem of 21st century and the fifth leading cause of death worldwide. According to Global Burden of Disease Study predictions, India, China and USA will be the top three leading countries for the prevalence of diabetes [1]. The approximate estimate of 31.7 million people with diabetes in India in 2000 will increase to 79.4 million by year 2030 and the size of the USA population with diabetes, both diagnosed and undiagnosed, will rise from approximately 30 million now to 44 million by the year 2030 [2]. T2D is strongly linked to various metabolic disturbances including obesity, insulin resistance, dyslipidemias, and elevated blood pressure. Linkage and candidate-gene focused studies successfully identified some rare forms of T2D controlled by one or two genes such as the various forms of maturity onset diabetes of young (MODY), mitochondrial diabetes, and neonatal diabetes. However, no single locus was

noted to have strong and consistent evidence of linkage with the most common form of T2D in multiple populations [3].

Elevated serum lipid levels are important risk factors for the development of cardiovascular disease (CVD). The genetic basis of several monogenic forms of lipid disorders has been determined, including familial lipoprotein lipase (LPL) deficiency, apoC-II deficiency, defective apoB, familial hypercholesterolemia, and familial triglyceridemia [4]. However, genes associated with common forms of dyslipidemia in the general population remain elusive.

Recent genome-wide association studies (GWAS) performed for many complex traits are revolutionizing the dissection of genetic determinants of several complex traits including T2D and serum lipids. Although these studies are adding to the list of reliably associated common loci controlling T2D and blood lipids and even other complex traits, these loci explain only a small portion of the heritable component associated with these complex diseases.

Clearly, additional loci that can explain a large proportion of the variation await discovery.

Asian Indians, one quarter of the global population, have unusually high CVD mortality and very high prevalence of insulin resistance and T2D [5]. The increased susceptibility to early onset of T2D and premature CVD in Asian Indians was confirmed in several earlier studies [6,7,8,9]. Indians tend to develop T2D at a relatively earlier age of 40–45 that is about 10–15 year earlier than European populations [6,10,11,12]. However, the reasons underlying the increased morbidity and mortality associated with T2D and CVD and in people of South Asian ancestry are poorly understood. In this investigation, we have carried out an autosomal genome-wide linkage scan to map the genes associated with T2D and serum lipid levels using our large family-based cohort from the Sikh Diabetes Study (SDS) [13]. This non-smoking, primarily vegetarian, endogamous caste group has high prevalence of diabetes and CVD with young age-of-onset. To our knowledge, this is the first report of genome-wide linkage studies on T2D and quantitative lipid traits in a population from South Asian Continent.

Methods

Study Population, Ascertainment Criteria, and Recruitment

This study was carried out on an endogamous community of Khatri Sikhs living in Northern Indian states of Punjab, Haryana, and New Delhi. The Khatri population was chosen because of its relatively higher prevalence of diabetes as compared to other Sikh castes. Khatri Sikhs are more affluent and live in cities and are traders by profession. In general, Sikhs do not smoke for religious and cultural reasons and about 50% of the study participants are life-long vegetarians. A total of 1,115 individuals from 338 families were extensively phenotyped [13]. DNA samples of 870 individuals (526 male/344 female) comprising 685 T2D cases and 185 normal glucose tolerant (NGT) relatives were successfully genotyped and used in this investigation. The T2D cases were 25 years or older and mean age at the time of recruitment (mean \pm standard deviation [SD]) was 54.2 ± 11.0 years. Average age of unaffected relative was 46.0 ± 14.7 years with a minimum age of 19 years. Only individuals who reported that all four grandparents were Khatri Sikhs of North Indian origin, who had Khatri surnames, and who spoke the Punjabi language were included. In addition, probands were required to have two or more full siblings with diabetes, or at least one living parent, and more than two siblings available for sampling. Excluded from the sample were half-siblings, adopted individuals, and individuals of South, East and Central Indian origin; individuals with type 1 diabetes (T1D) or a family member with T1D; individuals with rare forms of T2D such as maturity-onset diabetes of young (MODYs), or secondary diabetes (e.g., due to hemochromatosis or pancreatitis). Clinical characteristics of the SDS participants used for this investigation are summarized in **Table 1**. All blood samples were obtained at the baseline visit. All participants provided a written consent following an informed consent procedures approved by Institutional Review Boards (IRBs). All SDS protocols and consent documents were reviewed and approved by the University of Oklahoma Health Sciences Center (OUHSC) (IRB # 13302, approved till August 31, 2011) and the University of Pittsburgh (IRB # 021234) as well as the Human Subject Protection (Ethical) committees at the participating hospitals and institutes in India. The Ethical committees of local institutions in India were Hero DMC Heart Institute, Ludhiana, and Guru Nanak Dev University, Amritsar. Each Institute in India also

Table 1. Characteristics of Study Population Stratified by Gender and Disease (Mean \pm SD).

		T2D cases	Unaffected Relatives	<i>p</i> value**
		685 (412M/273F)	185 (116M/69F)	
Age at recruitment (year)	M*	52.42 \pm 11.36	44.71 \pm 15.20	<0.0001
	F	55.93 \pm 10.88	48.60 \pm 13.60	<0.0001
Age at onset (years)	M	45.80 \pm 10.65	–	–
	F	48.62 \pm 10.40	–	–
Duration of T2D (years)	M	7.82 \pm 7.42	–	–
	F	7.33 \pm 6.65	–	–
BMI (kg/m ²)	M	26.90 \pm 4.23	26.89 \pm 4.53	0.876
	F	28.50 \pm 5.20	27.58 \pm 4.79	0.140
WAIST (cm)	M	95.50 \pm 10.40	93.2 \pm 11.55	0.040
	F	92.70 \pm 11.00	88.1 \pm 10.60	<0.0001
HIP (cm)	M	95.8 \pm 8.20	96.0 \pm 8.50	0.875
	F	99.3 \pm 11.00	97.6 \pm 9.70	0.135
WHR [†]	M	0.99 \pm 0.07	0.97 \pm 0.07	<0.0001
	F	0.94 \pm 0.07	0.90 \pm 0.07	<0.0001
Fasting Glucose (mg/dl)	M	185.81 \pm 70.66	95.30 \pm 11.38	<0.0001
	F	193.23 \pm 74.58	97.77 \pm 9.31	<0.0001
Total Cholesterol (mg/dl)	M	177.18 \pm 44.34	174.54 \pm 45.86	0.525
	F	187.24 \pm 47.42	177.58 \pm 38.94	0.049
Triglycerides (mg/dl)	M	197.63 \pm 113.50	160.71 \pm 85.49	<0.0001
	F	172.66 \pm 95.34	155.74 \pm 69.72	0.078
HDL-cholesterol (mg/dl)	M	38.11 \pm 12.44	39.21 \pm 10.01	0.321
	F	41.76 \pm 12.59	43.70 \pm 11.37	0.187
LDL-cholesterol (mg/dl)	M	99.98 \pm 35.68	98.36 \pm 32.96	0.619
	F	110.17 \pm 39.81	102.89 \pm 35.56	0.098
VLDL-cholesterol (mg/dl)	M	39.75 \pm 24.05	32.45 \pm 18.43	0.001
	F	34.77 \pm 19.18	31.13 \pm 13.84	0.056

*M - male, F- female; [†]Waist to hip ratio; **Difference between T2D cases and unaffected relatives.

doi:10.1371/journal.pone.0021188.t001

separately obtained Federal Wide Assurance (FWA) from the Office of Human Research Protection (OHRP) from the US Department of Health and Human Services (DHHS). All the key investigators and key personnel working for SDS obtained online training for Human Participant Protection Education for Research.

SDS Families

A total of 557 families were investigated and 236 families were excluded because they did not meet the eligibility criteria for the study. A total of 321 families containing 870 individuals (526 male/344 females), who were successfully genotyped (call rate \geq 95%), were used for linkage analysis of T2D. These 321 diabetic families comprised 275 affected sibling pairs, 59 affected cousin pairs, 127 affected parent-child pairs, 1 affected grand parent-child

Table 2. Description of SDS Pedigrees.

	Phenotyped*			Genotyped and phenotyped for T2D**			Genotyped and phenotyped for lipid levels**		
	Total	Male	Females	Total	Male	Females	Total	Male	Females
Number of families	338	–	–	321	–	–	316	–	–
Family size	6.51(*)	–	–	2.71(**)	–	–	2.68	–	–
Generations (average)	2.46	–	–	2.49	–	–	2.49	–	–
Number of individuals in pedigrees	2,199(*)	1248	951	870(**)	526	344	846(**)	511	335
Founders	979(*)	–	–	85	–	–	82	–	–
Number of affecteds in pedigrees	1,202(*)	710	492	685	412	273			
Number of individuals with blood available	1,115	684	431						
Number of affecteds with blood available	868	530	338						
Number of unaffected with blood available	247	154	93						

*—including deceased; **—excluding deceased.

doi:10.1371/journal.pone.0021188.t002

pair, and 61 affected avuncular pairs. We collected an average of 6.5 participants per family with family size ranging from 3 to 105 members. The average number of generations per family was 2.5. Of these 321 families, 316 families containing 846 individuals (511 male/335 female) were used in the linkage analysis of blood lipids (**Table 2**).

Phenotypes

The diagnosis of T2D was confirmed by (a) searching medical records for indications of symptoms of diabetes or measures of blood glucose levels, (b) use of diabetic medication, and (c) measuring fasting glucose levels following the guidelines of American Diabetes Association [14]. A medical record indicating either (1) a fasting plasma glucose level >126 mg/dl after a minimum 12-h fast or (2) a 2-h post glucose level >200 mg/dl [2-h oral glucose tolerance test (OGTT)] on more than one occasion with symptoms of diabetes. In the absence of medical record information, we confirmed self-reported T2D cases by performing a 2-h OGTT. The 2-h OGTTs were performed following the criteria of the World Health Organization (WHO) (75 g oral load of glucose). The NGT diagnosis was based on a fasting glycemia <108 mg/dl or a 2-h glucose <140 mg/dl. The average age at diagnosis was 47 years and duration of diabetes was about 7.5 years. Since T2D remains asymptomatic for several years, an average Asian Indian patient with new onset of diabetes might actually had diabetes 4–7 years before diagnosis [15]. This is in sharp contrast to the mean age at onset of 60 years or above in developed countries [10,11,16,17].

Body mass index (BMI) was calculated as [weight (kg)/height (meter)²], and waist-to-hip ratio (WHR) was calculated as the ratio of abdomen or waist circumference to hip circumference. Despite having comparable BMI (27.5±4.0 T2D cases vs. 27.3±4.7 controls), patients had a pronounced abdominal adiposity as reflected by their significantly higher WHR (0.97±0.07 vs. 0.94±0.07; p<0.0001) than controls. Interestingly, WHR in Khatri Sikh men (BMI 26–27 kg/m²) was higher than obese Mexican American men (BMI>32 kg/m²); Sikhs (0.97±0.05) vs. Mexican Americans (0.95±0.06) [18]. Perhaps central obesity is the underlying cause of high risk to insulin resistance and high prevalence of T2D and CVD in Indians.

Education (highest level completed) was scored 1–4 where 1 = primary or none, 2 = high school, 3 = bachelor degree, and 4 = post graduate degree. Job-grade was scored 1–3 based on

education and economic status where 1 = high income, 2 = middle-income, 3 = lower-middle and lowest income class; category 1 was used as a reference group. Smoking information was collected on past smoking, current smoking status, length of time, number of cigarettes smoked/day. Alcohol consumption was scored 0–4 where 0 = no alcohol, 1 = 50 to 100 ml/day, 2 = 100 to 400 ml/day, 3 = 400 to 1000 ml (1L)/day 4 = >1 L/day. Physical activity was scored 1–3 based on level of activity performed where 1 = very active, 2 = moderately active, 3 = quite inactive. About 83% of T2D patients were taking oral hypoglycemic agents. Some were maintaining glycemic control by diet and exercise. The individuals on lipid-lowering medications were not included in the analysis. Further recruitment details are available elsewhere [13].

Metabolic Estimations

Serum lipids [total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, very low-density lipoprotein (VLDL) cholesterol, and triglycerides] were quantified using standard enzymatic methods (Roche, Basel, Switzerland). Fasting serum insulin was measured by radio-immuno assay (Diagnostic Products, Cypress, USA). All quantitative parameters were determined by following manufacturer's instructions using a Hitachi 902 auto-analyzer (Roche, Basel, Switzerland).

Marker Genotyping

DNA was extracted from buffy coats using QiaAmp blood kits (Qiagen, Chatworth, USA) or by the salting out procedure [19]. 870 samples were successfully genotyped for 398 polymorphic microsatellite markers with an average spacing of 9.26 cM on the autosomes by the National Heart Lung and Blood Institute's (NHLBI) Mammalian Genotyping Service (<http://www.marshmed.org/genetics>). A total of 870 (526 male, 344 female) samples were used in linkage analysis of T2D and 846 (511 male, 335 female) samples were used in linkage analysis of lipid levels after excluding those with call rate <95%, relationship errors, gender errors, and those with missing phenotypes.

Error Checking and Data Handling

A variety of statistical software was used to complete this study. To set up the files for analysis, we extensively used the statistical

software R (version 2.0.1). Data cleaning was performed following several steps. To check for inconsistencies in the self-reported family structures, we carried out relationship testing using PREST [20] and RELPAIR [21,22]. PEDCHECK [23] was used to detect Mendelian inconsistencies in genotype combinations within a family. PEDSTATS (version 0.6.9) [24] was used to obtain counts of individuals included in the analysis.

Phenotype Normalization and Adjustment for Covariates

To adjust for the confounding effects of environmental influence on the lipid traits, we included information on age, age², sex, BMI, dietary and lifestyle factors (smoking, alcohol consumption, and physical activity), socio-economic status (education and job-grade) as covariates. To select significant covariate, both stepwise regression and backward elimination were used in genetic models. Significant covariates considered for selection in the model were age, age², sex, job grade, level of alcohol consumption. Additionally, analysis was performed including and excluding BMI in the model despite its elimination in stepwise regression. Univariate analysis was performed to obtain summary statistics for each trait (online supplementary Table S2). A classical multiple linear regression model: $Y = X\beta + \epsilon$, was used where Y is the response and X is the design matrix and β is the vector of regression coefficients. To reduce collinearity between age and age², these variables were mean centered. Since most of the traits were right skewed, they were transformed using the Box-Cox transformation method (Figure S1). Box-Cox method provides optimal value of the transformation parameters and increases the applicability and usefulness of statistical techniques based on the normality assumption and can significantly improve the linear fit of Y against X . After transformation we checked for the outliers. Regression parameters were estimated after exclusion of outliers (points outside of mean ± 3 SD), and residuals were computed for all participants. After building the model, the data were rechecked for further outliers using the jackknife method. Influential observations also were taken care of. High Leverage point and Cook's distance for each observation were also measured. To test the significance of the parameters, a significance level of 0.05 was used throughout the analyses. To sample a set of putatively unrelated individuals for the regression analysis, we took all phenotyped founders. If a family had no phenotyped founders, then we sampled one phenotyped non-founder from that family. Although BMI was not a significant covariate in step-wise selection, the entire QTL-ALL analyses were performed both including and excluding BMI as a covariate in the model.

Genome-wide Linkage Analysis for T2D

We used the S_{all} statistic [25] as implemented in Merlin [26] to perform linkage analysis for the trait T2D. This non-parametric method has excellent power and is robust across a wide variety of disease models [27,28]. Using FastSLINK [29,30,31], we simulated genetic data, for a 2-allele marker for an allele frequency of 0.01 and a penetrance vector of (0.054, 0.50, 0.70), which implies that the relative risk to siblings, λ_{sib} , is a relatively low 1.49. Then a S_{all} LOD score was computed using Merlin. From this simulated data we have 98% power to detect a LOD ≥ 2 ; and 90% to detect a LOD ≥ 3 in this cohort. Notably, these power estimates are conservative as we excluded the biggest two families which were too complex to run through Merlin.

QTL-ALL Analysis for Mapping Lipid Traits

In this study, families containing individuals affected by T2D are preferentially over-sampled, so this sample is non-randomly ascertained with respect to T2D. Therefore, to the extent that lipid

traits are correlated with T2D status, the sample is also non-randomly ascertained with respect to the lipid traits. Thus, it would not be appropriate to use the usual variance-component based linkage analysis methods on these data. Instead we used score-based linkage statistics as implemented in the QTL-ALL program for this data set [32]. We decided to use the statistic SCORE.MAX, which is recommended in most circumstances, and which has been shown to work well even on non-randomly ascertained data. The current version of QTL-ALL can handle only nuclear family pedigrees. So the Mega2 program was used to convert the multi-generation families to single generation nuclear families [33].

Results

Family Structure Error, Gender Error and Genotype Error Checking

Family structure data and X-linked genotypes at 27 markers were combined to detect possible gender errors by looking for males who are more heterozygous than expected and females who are more homozygous than expected. Five males were heterozygous at more than two markers; 16 women were more than 80% homozygous. All suspect participants were rechecked to ensure there was no misreporting of gender. We used RELPAIR and PREST to check the accuracy of self-reported family relationships. Misclassification of relationship for half-siblings as full-sibling, and unrelated as cousins, were detected and resolved. Participants with unresolved relationship errors were removed from families before analysis. We also used PEDCHECK to check Mendelian inconsistencies at each marker and erroneous data were omitted from further analysis. **Table 1** shows the clinical and physical characteristics of the SDS participants used in the analysis.

Linkage Analysis for T2D

As shown in online **Figure S1**, non-parametric multipoint linkage analysis did not show any chromosomal region to be significantly associated with T2D in this Sikh cohort. Adjusting for age, BMI, and gender did not alter linkage signal significantly and consequently were not included as covariates in the results presented. We found little evidence of linkage with T2D with maximum LOD of 1.24 reached on chromosome 2p24 near microsatellite markers SRAP and X130YG9P. No other region revealed any signal (LOD > 1.00) associated with T2D in these families.

Influence of Environmental Factors on Lipid Traits

Univariate analysis of the lipid traits showed some individuals with very high or very low outlier values, which were removed from the analysis. As needed a Box-Cox transformation was used to make the error distribution of the data more normal (**online Figure S2**). Regression models were then fitted for the transformed traits. In the variable selection step, in most cases forward stepwise-regression and backward elimination agreed with each other. **Table 3** shows the final models selected after detecting the significant covariates for each lipid trait analyzed. Total serum cholesterol levels were influenced by economic status. The correlation between VLDL cholesterol and triglycerides was very high (0.98) and level of alcohol consumption was a significant factor for influencing both serum triglycerides and VLDL-cholesterol levels. Gender was a significant covariate for serum HDL cholesterol, and age, age², and socio-economic status (job grade) were significant predictors of serum LDL cholesterol levels (**Table 3**). All the estimated coefficients are presented in online **Table S1**. High leverage points and Cook's distance were

Table 3. Final model variables in the five lipid traits.

Covariate Trait	Job grade	Alcohol consumption	Sex	Age	Age ²
Total Cholesterol	*				
Triglycerides		*			
HDL Cholesterol			*		
LDL Cholesterol	*			*	*
VLDL Cholesterol		*			

*represents significant covariate used for each lipid trait.
doi:10.1371/journal.pone.0021188.t003

calculated to detect influential observations and poorly fitted observations. After removing the maximum Cook's distance points, there was no significance change in the model. Calculated jackknife statistics was also within the acceptance region. Residuals of each trait were calculated and these residuals were used for the final QTL-ALL analysis.

QTL-ALL Analysis for Mapping Lipid Traits

QTL-ALL analysis, using the Score.Max statistics, was performed for the five quantitative traits. An overview of the linkage results for the significant signals associated with serum lipid associated traits is given in **Figure 1** and **Table 4**. Several QTLs with $p \leq 0.005$ were detected on chromosomes 5p, 9q, 10q, 10p, and 22q. The strongest linkage signal ($p = 0.0011$) was detected on chromosome 10q21.2 near D10S1225 for serum HDL cholesterol. Suggestive evidence of linkage for total cholesterol was observed on chromosome 5 near marker D5S2488 ($p = 0.0031$), and on chromosome 22 near marker TCTA015M ($p = 0.0016$). Two signals, one near marker D9S1122 ($p = 0.0039$) on chromosome 9 and other near D10S1426 ($p = 0.0045$) on chromosome 10, were detected for LDL cholesterol. A peak for HDL ($p = 0.031$) was seen near marker D9S934 on chromosome 9. No significant signal for serum triglycerides was observed (**online Figure S3**). Because obesity is a major risk factor for CVD and T2D risk, and affects lipid levels, we also tested linkage signals including and excluding BMI. Our results did not change after including BMI in the model.

Discussion

Our study represents the first large scale genome-wide effort to identify chromosomal regions with putative loci affecting T2D and lipid traits in a unique community of Asian Sikhs from Northern India. This diabetic cohort from a genetically homogenous subgroup was collected with the initial goal of identifying T2D predisposing genes. However, the results of our non-parametric linkage scan did not identify any chromosomal region to be significantly linked to T2D (**online Figure S1**). Note that the non-parametric method for linkage (used in our study) only considers allele sharing between affected individuals, therefore, the ambiguous phenotype of unaffected members is unlikely to have led to the failure to detect linkage in this large sample. These results reaffirm the highly complex nature of T2D phenotype. Essentially, our study failed to identify genes associated with T2D even when a homogenous population was used to control genetic heterogeneity associated with T2D phenotype and a sample collected from one geographic location was used to reduce environmental heterogeneity. These findings suggest that the genes responsible for T2D in Sikhs have small effects, as seen in other

ethnic groups, and are difficult to detect using linkage analysis. It can be argued that in comparison to random-mating population, higher identity by descent (IBD) sharing in this inbred population might have reduced the power of detecting significant linkage. In this scenario, one would expect to see increased average IBD leading to false positive indications of linkage. On the contrary, we found the opposite with no substantial increase in IBD among affected individuals and thus no linkage. At the same time, we believe that our linkage data may still contain considerably useful information that could enable the discrimination of causal variant from a near-by variant that is merely in linkage disequilibrium (LD) [34]. Interestingly, our case-control association studies have confirmed some Caucasian GWAS loci (*TCF7L2*, *PPARG*, *KCNJ11*, *FTO* and *KCNQJ*) associated with T2D in this population [35,36,37,38]. Therefore, further fine mapping especially in the elevated regions using high-density SNP panel and whole genome sequencing may identify rare and functional variants with large effects contributing to T2D. These investigations also may answer the questions of 'missing heritability' which is expected to lie in the 'rare' variants and which the GWA studies are unable to explain [39].

The other aim of this investigation was to identify genomic regions affecting lipid-related phenotypes in this cohort. We performed QTL-ALL analysis on this non-randomly ascertained dataset, which revealed several suggestive linkage signals associated with serum lipid levels (**Table 4**). Classical multiple linear regression models were used to adjust for environmental effects on the serum lipid traits. In view of strong environmental component associated with T2D and lipid metabolism, we have carefully analyzed the environmental factors, particularly the unique life style factors such as diet, physical activity, obesity, job status, socioeconomic status, gender, and medication that could potentially influence these traits. As explained in the Results section, the significant covariates with potential to modify linkage effect were identified and included in the analysis model. The strongest evidence of linkage ($p = 0.0011$) for HDL cholesterol was detected on chromosome 10q21.1–21.2. Suggestive evidence of linkage to ApoA-I was observed on chromosome 10q21.1 in the Quebec Family Study (QFS) [40]. The same region containing proto-cadherin 15 (*PCDH15*) gene (10q21.1) has been associated with multiple lipid traits in Finnish and Dutch multigenerational dyslipidemic families [41]. Another strong GWAS candidate gene linked with metabolic traits is solute carrier family 16, member 9 (*SCL16A9*) that also maps to chromosome 10q21.2 [42]. It is a proton-linked monocarboxylate transporter and catalyzes the rapid transport of many monocarboxylates across the plasma membrane. Chromosome 10 also carried a signal for LDL cholesterol at 10p11.23 ($p = 0.0045$) in our study. The closest candidate gene at this region is *KIAA1462* (10p11.23) that encodes a yet uncharacterized protein. However, a recently published GWAS showed an unambiguous evidence for association of rs3739998 ($p = 7.2 \times 10^{-8}$) within this gene with CVD and myocardial infarction in German MI Family cohort (GerMIFS) III (KORA) [43]. Interestingly, some common variants in *PCDH15* and *SCL16A9* and *KIAA1462* genes are also associated with multiple lipid traits including HDL cholesterol, LDL cholesterol, and triglycerides (p values of 0.004 to 0.0001) in our provisional results of lipid GWAS being performed on the population originated from the same Asian Indian community (unpublished results).

A linkage peak for total serum cholesterol ($p = 0.0031$) was detected near marker D5S2488 at the proximal region of chromosome 5p15.33. This region was previously linked to LDL cholesterol in the NHLBI Family Heart Study [44] and HDL

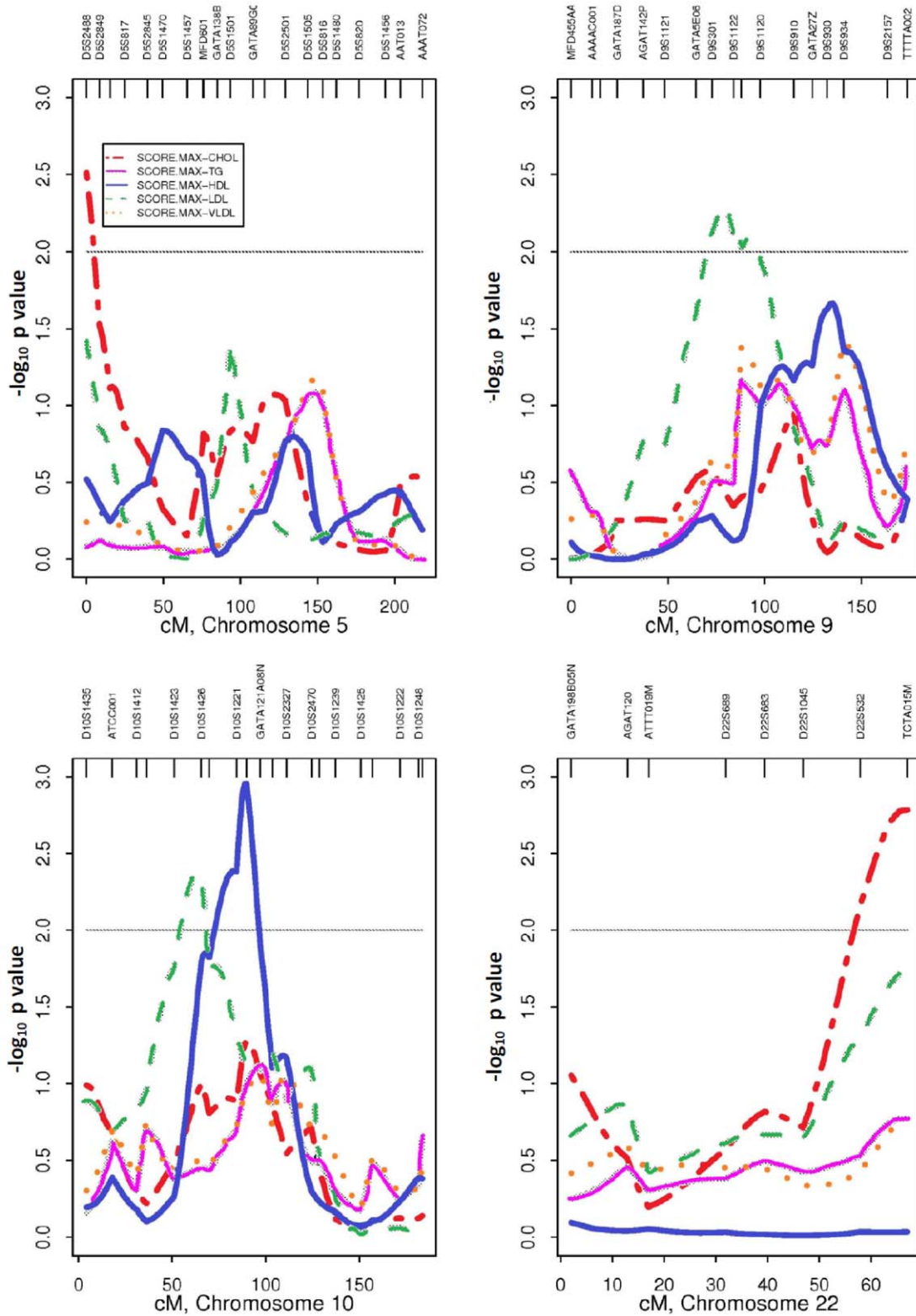


Figure 1. Genome-wide linkage scan to detect susceptibility loci for five blood lipid phenotypes using QTL-ALL analysis using 316 pedigrees. Linkage plots show significant signals at four chromosomal regions with allele sharing LOD ($-\log_{10} p$ value) on Y axis and chromosome distance (cM) on X axis. Significant linkage includes chromosome 5 near marker D5S2488 ($p=0.0031$) for total cholesterol; chromosome 9 near marker D9S1122 ($p=0.0039$) for LDL cholesterol; chromosome 10q21.2 near D10S1225 ($p=0.0011$) for HDL cholesterol; and chromosome 22 near marker TCTA015M ($p=0.0016$) for total cholesterol. doi:10.1371/journal.pone.0021188.g001

Table 4. Susceptibility regions for serum lipid levels with Score.Max p values of ≤ 0.005 .

Trait	Chromosome	Cytogenetic position	Physical Position*	Closest Marker	Genetic Position (cM)	Score.Maxp Value
Total Cholesterol	5	5p15.33	180390	D5S2488	0.00	0.0031
LDL Cholesterol	9	9q21.13	78878414	D9S1122	75.88	0.0039
LDL Cholesterol	10	10p11.23	30535660	D10S1426	65.61	0.0045
HDL Cholesterol	10	10q21.1	57199892	D10S1221	84.44	0.0041
HDL Cholesterol	10	10q21.2	64425005	D10S1225	89.69	0.0011
Total Cholesterol	22	22q13.32	47925896	TCTA015M	66.96	0.0016

*NCBI Build 36.1 positions from the UCSC browser.
doi:10.1371/journal.pone.0021188.t004

cholesterol in the Hypertension Genetic Epidemiology Network Blood Pressure Study [45]. Additionally, meta-analysis of linkage scans from four studies revealed a modest signal for LDL cholesterol (LOD 1.6) on chromosome 5p15.33 [46]. The suggestive linkage for LDL cholesterol at 9q21.13 near marker D9S1122 in our cohort was also associated with the triglyceride phenotype in a linkage study performed in families with myocardial infarction [47].

The linkage signal at chromosome 22q13.32 near marker TCTA015M ($p = 0.0016$), detected for total cholesterol was linked with familial hypercholesterolemia in a Utah study [48]. The chromosomal region 22q11–13 was also reported to effect HDL cholesterol in the Old Order Amish [49]. Notably, the strongest candidate gene in this region is *PPAR α* , which is a ligand-activated nuclear transcription factor and controls extracellular and intracellular lipid metabolism, and also inhibits progression of atherosclerotic lesions [50]. Lipid-lowering drugs of fibrate class are synthetic ligands of *ppar α* [51]. Variants in this gene were reported to be associated with T2D and CVD [52]. Another gene *CELSRI* (located at 22q11–13) is associated with ischemic stroke in recent Japanese GWAS [53]. Furthermore, a single nucleotide polymorphism (SNP) near *CELSR2* on chromosome 1p13 (homologous to *CELSRI*) is associated with LDL cholesterol and myocardial infarction in a meta-analysis study by Myocardial Infarction Genetics Consortium [54].

Our study does not represent a common replication attempt to identify lipid loci in an independent population. Rather, this investigation has been carefully carried out in this unique family-based cohort using a conservative statistical approach applying score-based statistics to map quantitative lipid traits in a non-randomly ascertained dataset. Exceeding our expectations, this study has identified linkage regions, primarily HDL cholesterol (10q21.1–21.2) and total cholesterol (22q13.32) that were previously reported for lipid traits or CVD. The most interesting part of this study is that some of these linkage signals also harbor important candidate loci (e.g., *KIAA1462*, *PCDH15*, *PPAR α* , *SLC16A9*, and *CELSRI*) implicated with lipid traits in recent GWAS and meta-analysis studies and also some of these regions overlap with prior linkage studies [55,56,57]. Therefore, our findings suggest that these regions might contain some novel genes for blood lipids rather than chance findings, and perhaps some of the loci may have larger effects in this Khatri Sikh cohort. Notably, the presence of HDL cholesterol signal on chromosome 10q21.2 is particularly important in view of low HDL cholesterol-associated CVD risk in Asian Indian men, in general, and may strongly relate to gene-environmental interaction which is enhanced by rapidly emerging western lifestyle [58,59]. Further fine mapping with more efficacious strategy using SNP-based

arrays (which would also help determine LD over small intervals), sequencing, and functional studies should allow rapid detection of novel target genes of therapeutic importance under these candidate regions.

Conclusions

Unlike previous studies, our genome-wide linkage scan could not identify any significant chromosomal region associated with T2D in this unique family cohort of Punjabi Sikhs with increased risk to developing T2D and cardiovascular illnesses. Our study, however, for the first time provides an evidence of linkage for loci controlling quantitative lipid traits at four chromosomal regions in this Asian Indian population. The strongest linkage signal was seen for HDL cholesterol on chromosome 10q21.2. Our data also revealed linkage signals for total cholesterol on chromosome 5p15.33 and 22q13.32, and for LDL cholesterol on 10p11.23 and 9q21.13. Some of these regions have been linked to lipid-related traits in recent GWA studies and contain other plausible candidate genes. The strongest peak for HDL cholesterol ($p = 0.0011$ at 10q21.2) suggests that this region may contain novel gene(s) influencing serum HDL cholesterol levels and other lipid traits. Further denser and more informative genotyping in each of these regions would be important to discover functional loci influencing blood lipids.

Supporting Information

Figure S1 Genome-wide non-parametric linkage scans for type 2 diabetes using 321 diabetic pedigrees and 398 microsatellite markers (9.26 cM). Individual plot shows linkage signals (Kong and Cox LOD score) on Y axis and microsatellite markers on X axis. None of the chromosome regions revealed any signal associated with T2D in these pedigrees.

(TIF)

Figure S2 Plot of Box-Cox coefficient lambda and the distribution of five quantitative traits including total cholesterol, triglycerides, HDL cholesterol, LDL cholesterol, and VLDL cholesterol before and after transformation.

(TIF)

Figure S3 Genome-wide autosomal linkage scan for five blood lipid phenotypes. Individual plot shows allele sharing LOD ($-\log_{10} p$ value) on Y axis and chromosome distance (cM) on X axis.

(TIF)

Table S1 Linear regression model for quantitative traits.

(DOC)

Acknowledgments

This study would not have been possible without the generous support of the families who participated in this study and the non-governmental, social, and religious organizations who assisted in the recruitment and ascertainment of the SDS families. Authors are also thankful to Mr. Jagtar S. Sanghera for his repeated help in subject recruitment and providing guidance and support in logistical and ethical issues. We are thankful to National Heart Lung and Blood Institute (NHLBI)'s Mammalian Genotyping Service (Contract Number HV48141), for genotyping our study population.

References

- Murray CJ, Lopez AD (1997) Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 349: 1498-1504.
- Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 27: 1047-1053.
- Zeggini E (2007) A new era for Type 2 diabetes genetics. *Diabet Med* 24: 1181-1186.
- Hegele RA (2009) Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* 10: 109-121.
- Basnyat B, Rajapaksa LC (2004) Cardiovascular and infectious diseases in South Asia: the double whammy. *British Medical Journal* 328: 781.
- Mather HM, Keen H (1985) The Southhall Diabetes Survey: prevalence of known diabetes in Asians and Europeans. *Br Med J (Clin Res Ed)* 291: 1081-1084.
- Snehalatha C, Satyavani K, Sivasankari S, Vijay V, Ramachandran A (1999) Insulin secretion and action in different stages of glucose tolerance in Asian Indians. *Diabet Med* 16: 408-414.
- McKeigue PM, Miller GJ, Marmot MG (1989) Coronary heart disease in south Asians overseas: a review. *J Clin Epidemiol* 42: 597-609.
- Banerji MA, Chaiken RL, Gordon D, Kral JG, Lebovitz HE (1995) Does intra-abdominal adipose tissue in black men determine whether NIDDM is insulin-resistant or insulin-sensitive? *Diabetes* 44: 141-146.
- McKeigue PM, Marmot MG, Syndercombe Court YD, Cottier DE, Rahman S, et al. (1988) Diabetes, hyperinsulinaemia, and coronary risk factors in Bangladeshis in east London. *Br Heart J* 60: 390-396.
- McKeigue PM, Pierpoint T, Ferrie JE, Marmot MG (1992) Relationship of glucose intolerance and hyperinsulinaemia to body fat pattern in south Asians and Europeans. *Diabetologia* 35: 785-791.
- Ramachandran A, Jali MV, Mohan V, Snehalatha C, Viswanathan M (1988) High prevalence of diabetes in an urban population in south India. *Bmj* 297: 587-590.
- Sanghera DK, Bhatti JS, Bhatti GK, Ralhan SK, Wander GS, et al. (2006) The Khatri Sikh Diabetes Study (SDS): study design, methodology, sample collection, and initial results. *Hum Biol* 78: 43-63.
- American Diabetes Association (2004) Diagnosis and classification of diabetes mellitus. *Diabetes Care* 27 Suppl 1: S5-S10.
- Votey SR PA (2004) Diabetes Mellitus, Type 2-A Review. *eMedicine Journal* [serial online] <http://www.emedicine.com/emerg/topic134.htm>. Accessed December 5, 2006.
- Ramachandran A SC, Viswanathan V (2003) Explosion of type 2 diabetes in the Indian subcontinent. *International Diabetes Monitor* 15(5): 1-6.
- Nakagami T, Qiao Q, Carstensen B, Nhr-Hansen C, Hu G, et al. (2003) Age, body mass index and Type 2 diabetes-associations modified by ethnicity. *Diabetologia* 46: 1063-1070.
- Tulloch-Reid MK, Williams DE, Looker HC, Hanson RL, Knowler WC (2003) Do measures of body fat distribution provide information on the risk of type 2 diabetes in addition to measures of general obesity? Comparison of anthropometric predictors of type 2 diabetes in Pima Indians. *Diabetes Care* 26: 2556-2561.
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16: 1215.
- McPeck MS, Sun L (2000) Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet* 66: 1076-1094.
- Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. *Am J Hum Genet* 61: 423-429.
- Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. *Am J Hum Genet* 67: 1219-1231.
- O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63: 259-266.
- Wigginton JE, Abecasis GR (2005) PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 21: 3445-3447.
- Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. *Biometrics* 50: 118-127.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30: 97-101.
- Davis S, Weeks DE (1996) Comparison of nonparametric statistics for detecting linkage in affected-sib-pair data. *Am J Hum Genet* 59: A216.

Author Contributions

Conceived and designed the experiments: DKS REF MIK. Performed the experiments: LFB. Analyzed the data: CEA. Contributed reagents/materials/analysis tools: DKS CEA LFB. Wrote the paper: DKS SR GSW NKM JRS REF MIK LFB CEA. National Heart Lung and Blood Institute (NHLBI)'s Mammalian Genotyping Service (Contract Number HV48141) performed genotyping for this study.

- Sengul H, Weeks DE, Feingold E (2001) A survey of affected-sibship statistics for nonparametric linkage analysis. *American Journal of Human Genetics* 69: 179-190.
- Ott J (1989) Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci U S A* 86: 4175-4178.
- Weeks DE (1990) A likelihood-based analysis of consistent linkage of a disease locus to two nonsynthetic marker loci: osteogenesis imperfecta versus COL1A1 and COL1A2. *Am J Hum Genet* 47: 592-594.
- Cottingham RW, Jr., Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* 53: 252-263.
- Bhattacharjee S, Kuo CL, Mukhopadhyay N, Brock GN, Weeks DE, et al. (2008) Robust score statistics for QTL linkage analysis. *Am J Hum Genet* 82: 567-582.
- Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE (2005) Mega2: data-handling for facilitating genetic linkage and association analyses. *Bioinformatics* 21: 2556-2557.
- Bourgain C, Genin E, Cox N, Clerget-Darpoux F (2007) Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? *Eur J Hum Genet* 15: 260-263.
- Sanghera DK, Nath SK, Ortega L, Gambarelli M, Kim-Howard X, et al. (2008) TCF7L2 polymorphisms are associated with type 2 diabetes in Khatri Sikhs from North India: genetic variation affects lipid levels. *Ann Hum Genet* 72: 499-509.
- Sanghera DK, Demirci FY, Been L, Ortega L, Ralhan S, et al. (2010) PPARG and ADIPOQ gene polymorphisms increase type 2 diabetes mellitus risk in Asian Indian Sikhs: Pro12Ala still remains as the strongest predictor. *Metabolism* 59: 492-501.
- Sanghera DK, Ortega L, Han S, Singh J, Ralhan SK, 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 Med Genet* 9: 59.
- Been LF, Ralhan S, Wander GS, Mehra NK, Singh J, et al. (2011) Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3,310 subjects from India and the US. *BMC Med Genet* 12: 18.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.
- Bosse Y, Chagnon YC, Despres JP, Rice T, Rao DC, et al. (2004) Compendium of genome-wide scans of lipid-related phenotypes: adding a new genome-wide search of apolipoprotein levels. *J Lipid Res* 45: 2174-2184.
- Huertas-Vazquez A, Plaisier CL, Geng R, Haas BE, Lee J, et al. (2010) A nonsynonymous SNP within PCDH15 is associated with lipid traits in familial combined hyperlipidemia. *Hum Genet* 127: 83-89.
- Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, et al. (2009) A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* 42: 137-141.
- Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, Konig IR, et al. (2011) Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J* 32: 158-168.
- Coon H, Eckfeldt JH, Leppert MF, Myers RH, Arnett DK, et al. (2002) A genome-wide screen reveals evidence for a locus on chromosome 11 influencing variation in LDL cholesterol in the NHLBI Family Heart Study. *Hum Genet* 111: 263-269.
- Coon H, Leppert MF, Eckfeldt JH, Oberman A, Myers RH, et al. (2001) Genome-wide linkage analysis of lipids in the Hypertension Genetic Epidemiology Network (HyperGEN) Blood Pressure Study. *Arterioscler Thromb Vasc Biol* 21: 1969-1976.
- Heijmans BT, Beckman M, Putter H, Lakenberg N, van der Wijk HJ, et al. (2005) Meta-analysis of four new genome scans for lipid parameters and analysis of positional candidates in positive linkage regions. *Eur J Hum Genet* 13: 1143-1153.
- Broeckel U, Hengstenberg C, Mayer B, Holmer S, Martin IJ, et al. (2002) A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 30: 210-214.
- Hunt SC, Hopkins PN, Bulka K, McDermott MT, Thorne TL, et al. (2000) Genetic localization to chromosome 1p32 of the third locus for familial hypercholesterolemia in a Utah kindred. *Arterioscler Thromb Vasc Biol* 20: 1089-1093.

49. Pollin TI, Hsueh WC, Steinle NI, Snitker S, Shuldiner AR, et al. (2004) A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis* 173: 89–96.
50. Staels B, Dallongeville J, Auwerx J, Schoonjans K, Leitersdorf E, et al. (1998) Mechanism of action of fibrates on lipid and lipoprotein metabolism. *Circulation* 98: 2088–2093.
51. Forman BM, Chen J, Evans RM (1997) Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. *Proc Natl Acad Sci U S A* 94: 4312–4317.
52. Lacquemant C, Lepretre F, Pineda Torra I, Manraj M, Charpentier G, et al. (2000) Mutation screening of the PPARalpha gene in type 2 diabetes associated with coronary heart disease. *Diabetes Metab* 26: 393–401.
53. Yamada Y, Fuku N, Tanaka M, Aoyagi Y, Sawabe M, et al. (2009) Identification of CELSR1 as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. *Atherosclerosis* 207: 144–149.
54. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41: 334–341.
55. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41: 56–65.
56. Kathiresan S, Melander O, Guiducci C, Surti A, Burt NP, et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40: 189–197.
57. Lilja HE, Suviolahti E, Soro-Paavonen A, Hiekkalinna T, Day A, et al. (2004) Locus for quantitative HDL-cholesterol on chromosome 10q in Finnish families with dyslipidemia. *J Lipid Res* 45: 1876–1884.
58. Balarajan R (1991) Ethnic differences in mortality from ischaemic heart disease and cerebro-vascular disease in England and Wales. *Br Med J* 302: 560–564.
59. McKeigue PM, Ferrie JE, Pierpoint T, Marmot MG (1993) Association of early-onset coronary heart disease in South Asian men with glucose intolerance and hyperinsulinemia. *Circulation* 87: 152–161.