Title
Assessing the genetic architecture of metabolic diseases using candidate gene and genome-wide approach

Author
Chan, Kei Hang Katie

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Assessing the genetic architecture of metabolic diseases using candidate gene and genome-wide approach

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Kei-hang Katie Chan

2012

ABSTRACT OF THE DISSERTATION

Assessing the genetic architecture of metabolic diseases using candidate gene and genome-wide approach

by

Kei-hang Katie Chan

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2012

Professor Eric Sobel, Co-chair

Professor Simin Liu, Chair

Much work has targeted the detection of disease genes through genetic mapping for metabolic diseases such as type 2 diabetes (T2D), cardiovascular diseases (CVD), and other diabetes-related traits such as body mass index (BMI) and hemoglobin (HbA1C) levels. However, the etiology of metabolic diseases remains partially understood hampering the development of more personalized diagnosis, treatment and prevention strategies.

This dissertation examines the association between genetic variants with risk of metabolic diseases and diabetes-related quantitative traits in both candidate gene and genome-wide scan settings. In particular, we assessed the association of genetic loci related to adiposity, inflammation, and lipid storage, with the risk of diabetes using a candidate gene approach. We also investigated biological pathways that may give rise to the development of vascular disease

(T2D and/or CVD) and also further investigated genetic variants related to BMI and HbA1C levels using a genome-wide approach. Chapter 1 introduces general background on the evolution of genetic research in the arena of metabolic diseases. Chapter 2 investigates common variants in the genomic region of *FABP4*, *CRP*, *TNF*, *IL6* and *PPARG* in relation to diabetes risk among postmenopausal women enrolled in the Women's Health Initiative Observational Study (WHI-OS). Chapter 3 examines whether common variants involved in vascular disease risk are clustered in multiple pathways among African and Hispanic American participants in the WHI SNP Health Association Resource (SHARe) cohort. Chapter 4 examines the association between genetic variants with BMI and HbA1C levels using a family-based genome-wide association approach among participants in the Framingham Heart Study (FHS).

Our main findings are: 1) Candidate gene-based studies indicate that variation exists across even the candidate gene regions. *FABP4* genotypes were associated with reduced VCAM-1 levels, though none of the common genetic variants in the *FABP4* gene examined were associated with risk of T2D. We also observed modest associations between *TNF* genetic variants and circulating concentrations of TNF-α-R2, although common variants of *CRP*, *TNF*, and *IL6* genes were not associated with T2D risk. Using the example of the *PPARG* gene with T2D risk, however, we replicated the association between the *PPARG Pro12Ala* genetic variant with diabetes risk and found that haplotype-based analysis is more powerful than single-SNP analysis for identifying genetic variants. 2) Using a pathway-based analytical approach and genome-wide scan data collected among African and Hispanic American postmenopausal women, we observed that genetic variants associated with vascular disease appeared to cluster into several biological pathways including the glycerolipid metabolism and PPAR signaling pathways. 3) We found strong associations between SNPs near the *LOC100507205* locus and

BMI in the family-based Framingham Heart Study with three generations. We also replicated five well-validated genes that have been previously reported to be significantly associated with the BMI trait. These findings contribute to the growing body of literature in identifying genetic variants in the development of metabolic disease, further future work (e.g. in the area of structure and functional variants) are warranted to improve understanding of the genetic architecture for metabolic outcomes. Increasing integration of cutting edge genomic science into population-based epidemiologic investigation will accelerate and improve our understanding of the genetic susceptibility of complex diseases. The work described in this dissertation represents a tip of our effort toward the ultimate improvement of the diagnosis, treatment and prevention of metabolic diseases in human populations.

The dissertation of Kei-hang Katie Chan is approved.

Zuo-Feng Zhang

Onyebuchi A. Arah

Qing Zhou

Eric Sobel, Committee Chair

Simin Liu, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

VITA

2000-2004	Bachelor of Information Engineering Department of Electrical and Electronic Engineering The University of Hong Kong (HKU) , Hong Kong, China
2002-2003	Global Engineering Exchange Scheme Drexel University, Philadelphia, PA
2002-2003	HKU Worldwide Exchange Scholarship
2004-2006	Master in Public Health (Epidemiology and Biostatistics) Keck School of Medicine The University of Southern California (USC) , Los Angeles, CA
2005-2007	Research Assistant Department of Preventive Medicine and Department of Molecular Epidemiology, USC, Los Angeles, CA
2006	Research Assistant Fellowship Department of Preventive Medicine, USC, Los Angeles, CA
2006 Summer	Intern at the Department of Ethics, Trade, Human Rights and Health Law in the Sustainable Development and Healthy Environments World Health Organization, Geneva, Switzerland
2008-2012	Graduate Student Researcher Center for Metabolic Diseases, UCLA, Los Angeles, CA
2008-2009	University Fellowship in Epidemiology
2009-2010	Weisman Memorial Fellowship
2010-2012	Burroughs-Wellcome Fellowship, Burroughs Wellcome Fund Inter-school Training Program in Metabolic Diseases

PUBLICATONS AND PRESENTATIONS

K.K. Chan, E. Sobel, S. Liu. Genome-wide Analysis of copy number variation in the Women Health's Initiative SNP Health Association Resource (WHI-SHARe): an example in type 2 diabetes. The Genomics and Randomized Trials Network, or Genome-wide Association Research Network into Effects of Treatment (GARNET) Steering Committee Meeting (December 2011). Omni Shoreham Hotel Washington, DC, U.S.A.

K.K. Chan, K. Brennan, N. You, X. Lu, Y. Song, Y. Hsu, G. Chaudhuri, L. Nathan, L. tinker, S. Liu. Common variations in the genes encoding C-reactive protein, tumor necrosis factor-alpha, and interleukin-6, and the risk of clinical diabetes in the Women's Health Initiative Observational Study. Clin Chem. 2011 Feb; 57(2):317-25

K.K. Chan. The past, present and future of Genome-wide Association Study (GWAS). The 10th Emerging Information and Technology Conference (EITC-2010). James H. Clark Center, Stanford University, Stanford, CA, U.S.A.

K. K. Chan, Y. Song, Y. Hsu, N. You, L. Tinker, S. Liu. Common Genetic Variants in Fatty Acid Binding Protein-4 (FABP4) and Clinical Diabetes Risk in the women's Health Initiative Observational Study. Obesity (Silver Spring). 2010 Jan 28.

K. K. Chan, Y. Song, Y. Hsu, N. You, L. Tinker, S. Liu. Relation of Genetic Variants in Fatty Acid Binding Protein-4 (FABP4)/ Adipocyte P2 (aP2) and Diabetes Risk in the Women's Health Initiative Observational Study. 52th American Society of Human Genetics (ASHG) Annual Meeting.

K.K. Chan, K. Kaneko, T.Y. Liang, P. Woo, Q.B. Wu, A. Zaw, P. Sun, P.H. Palmer, C.A. Johnson. Westernization and Alcohol Drinking Patterns Among Chinese Adults. 37th Asia-Pacific Academic Consortium for Public Health (APACPH).

K.K. Chan, C. Chang, F.H.Y. Chan. Fourier and Spectral Envelope Analysis of Medically Important Bacterial and Fungal Sequences (Undergraduate Final Year Project). 47th IEEE International Midwest Circuit and systems proceedings, July 2004, pp. III-175 - III-178.

S.K.P Lau, P.C.Y. Woo, M.Y. Mok, J.L.L. Teng, V.K.P. Tam, **K.K. Chan**, K.Y. Yuen. Haemophilus segnis is an important cause of 'non-influenzae' Haemophilus bacteremia: characterization by 16S RNA gene sequencing. Journal of Clinical Microbiology, Feb. 2004, pp. 877-880.

Chapter 1:

Introduction

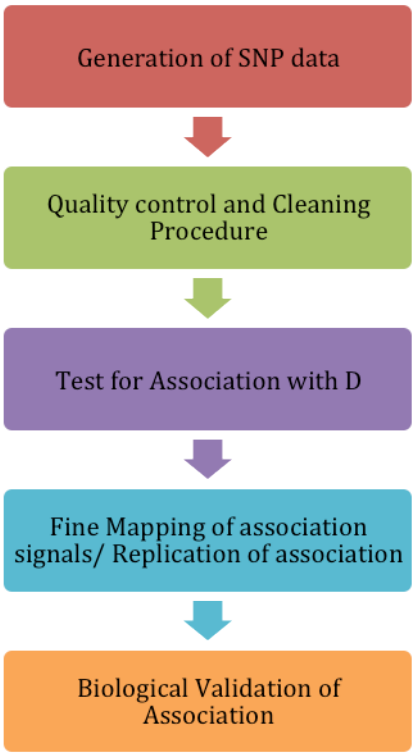
1.1 Background and Significance

Genetic variants play an important role in the pathogenesis of metabolic diseases such as adiposity[1] and T2D[2-4]. Type 2 diabetes (T2D) is a multifaceted metabolic disorder characterized by hyperglycemia fundamentally caused by impaired insulin action and secretion[5]. More than 300 million people worldwide have diabetes[6], which poses major challenges for public health. Multiple lines of evidence suggest the notion that genetic factors play a critical role in the etiology of metabolic diseases, particularly for T2D[2,4]. First, twin studies show that estimates for concordance rates for T2D have ranged from 0.26 to 0.76 in monozygotic twins[7-10]. Second, familial aggregation of the disease shows another source of evidence for a genetic role in the pathogenesis of T2D because families share both environments and genes. Third, the disease prevalence is observed to vary extensively among ethnic groups that share similar environment. Fourth, studies show a genetic basis for measures of some intermediate phenotypes leading to T2D, including insulin sensitivity and insulin secretion[2]. Therefore, an understanding of genetic determinants may contribute to better prevention, diagnosis, and treatment of disease. During the past several decades, linkage analysis and candidate gene approaches have been adopted to identify relevant disease genes, particularly for monogenic 'Mendelian' diseases[11,12]. Association approaches are powerful and have better resolution than linkage approaches. A large body of research through the candidate gene approach has identified many novel genes and loci related to metabolic diseases including glucokinase hexokinase 4 regulator (*GCKR*), glucose-6-phosphatase, catalytic, 2 (*G6PC2*) and ATP-binding cassette, sub-family B, member 11 (*ABCB11*), a gene that influences fasting glucose levels; fat mass and obesity associated (*FTO*) and melanocortin 4 receptor (*MC4R)* known to be linked with obesity or adiposity, as well as

many other variants that affect diabetes, triglyceride, HDL-cholesterol, and low-density lipoprotein (LDL)-cholesterol levels[13].

In recent years, genome-wide association studies (GWAS), which do not require an initial hypothesis, have been utilized for identifying genetic variants associated with different disease phenotypes[14]. GWAS have provided insights about diseases, in particular: (1) there are single nucleotide polymorphism (SNP) variants common in the population (i.e. with allele frequency >5%) that are robustly associated with disease for many diseases that has been investigated; (2) most of these variants are located in genes that play a role in biological pathways that were previously not known to be related with disease or are not in a known protein-coding regions[15,16]; (3) the associated SNPs are typically of modest effect sizes with odds ratios of risk alleles in the range of 1.1 to 1.5; (4) cumulative effects of many different SNPs associated with a disease usually explain only a small fraction of the familial risk; (5) not all disease traits are similar in genetic architecture[15,16]. Typical process involved GWAS is shown in Figure 1. In the first step, SNPs across the genome are genotyped. Second, data are subjected to quality control and data cleaning procedure after the generation of SNP data. Third, each SNP is then tested for association with a disease trait in a form of Manhattan plot. Fourth, SNPs or loci are selected for replication in an independent sample set. Last, additional genotype and more functional and biological work may be needed in independent replication cohorts to determine whether an association with a disease is genuine or not.

Figure 1: General workflow of a GWAS



GWAS is known to have greater power to identify genetic variants that confer modest disease risks than linkage and candidate genes analysis, even when a large number of markers is tested across the genome[14]. In comparing different GWAS on the same disease, the most significant SNPs in one study may not necessarily show up as the most significant SNPs in another study. Also, SNPs that are genuinely associated with disease may not be detected by any GWAS because of the small effect sizes of the SNPs and the lack of power of any individual study. GWAS have extended the breadth of genetic information, but most published GWAS list only the 20-50 most-significant SNPs and their neighboring genes (i.e. the "most-significant SNPs" approach), while paying little attention to the rest[17-23].

More recently, international efforts of GWAS have identified more than 40 genetic variants that affect T2D risk[24,25], providing novel biological insight into the pathogenesis of

metabolic diseases. More work needs to be done in analyzing the genomic data that may shed some light on improving the understanding of genetic architecture of human diseases.

1.2 Specific Aims

The primary objective of this dissertation is to examine the genetic architecture of metabolic diseases, mainly T2D, by adopting both candidate gene and GWAS approach and investigating SNPs as well as structural variants, especially CNVs. In the chapters that follow, we will discuss the methodology and findings regarding our specific aims to: 1) examine the genetic associations of variants in the *FABP, CRP, TNF-α, IL-6*, and *PPARG* genes with T2D risk and diabetes-associated biomarkers among postmenopausal women enrolled in the multi-ethnic Women's Health Initiative Observational Study (WHI-OS); 2) using a pathway-based GWAS approach, examine whether the common variants involved in Vascular Disease (VD, i.e. T2D and/or cardiovascular disease) risk are involved in multiple pathways among African and Hispanic American participants in the WHI SNP Health Association Resource (SHARe) cohort; 3) investigate the association between genetic variants with two diabetes-related quantitative traits, i.e. body mass index and glycated hemoglobin levels, using a family-based GWAS approach among participants in the Framingham Heart Study (FHS).

1.3 References

1. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet;42:949-60.

2. Das SK, Elbein SC. The Genetic Basis of Type 2 Diabetes. Cellscience 2006;2:100-31.

3. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38:320-3.

4. Owen KR, McCarthy MI. Genetics of type 2 diabetes. Curr Opin Genet Dev 2007;17:239-44.

5. Bae JS, Cheong HS, Kim JH, et al. The genetic effect of copy number variations on the risk of type 2 diabetes in a Korean population. PLoS One 2011;6:e19091.

6. Diabetes. WHO Media Centre Fact sheet N°312 2011.

7. Barnett AH, Eff C, Leslie RD, Pyke DA. Diabetes in identical twins. A study of 200 pairs. Diabetologia 1981;20:87-93.

8. Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD. Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. Diabetologia 1999;42:146-50.

9. Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 1987;30:763-8.

10. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia 1999;42:139-45.

11. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95-108.

12. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.

13. Lusis AJ, Attie AD, Reue K. Metabolic syndrome: from epidemiology to systems biology. Nat Rev Genet 2008;9:819-30.

14. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356-69.

15. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009;360:1759-68.

16. Visscher PM, Montgomery GW. Genome-wide association studies and human disease: from trickle to flood. JAMA 2009;302:2028-9.

17. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631-7.

18. Salonen JT, Uimari P, Aalto JM, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. Am J Hum Genet 2007;81:338-45.

19. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-6.

20. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-5.

21. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881-5.

22. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770-5.

23. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336-41.

24. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci 2010;1212:59-77.

25. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. Curr Diab Rep 2009;9:164-71.

Chapter 2A:

Common Genetic Variants in Fatty Acid-Binding Protein-4 (*FABP4*) and Clinical Diabetes Risk in the Women's Health Initiative Observational Study

2A.1 Introduction

Fatty acid-binding protein-4 (*FAPB4*), also known as adipocyte FABP (AFABP) and adipocyte P2, is highly expressed in adipocytes and macrophages. *FAPB4* is regulated during adipocytes differentiation, and its mRNA is transcriptionally controlled by fatty acids, nuclear hormone receptors, perosisome proliferator-activated receptor-γ agonists, insulin and liver X receptor[1-4], all of which have been shown to play an important regulatory role in inflammation and energy metabolism. Recently, deficiency of *FAPB4* has been correlated with plasma lipid levels, especially as a protective factor against atherosclerosis and coronary heart disease risk[5-9]. In animal models, a modest increase in insulin sensitivity has been exhibited in obese mice with *FAPB4* deficiency[10-12]. However, little is known about the association between the genetic variants in *FAPB4* and diabetes (T2D) risk in human population.

Therefore, we comprehensively assess the genetic associations of variants in the *FAPB4* gene with T2D risk and diabetes-associated biomarkers in a nested case-control study of postmenopausal women aged between 50 and 79 years who enrolled in the Women's Health Initiative Observational Study (WHI-OS). We selected and genotyped a total of 11 haplotype-tagging single-nucleotide polymorphisms (tSNPs) spanning 41.3 kb across *FAPB4* in all samples.

Furthermore, we investigated whether and to what extent the *FAPB4* variants affect circulating levels of inflammatory cytokines [tumor necrosis factor-α (TNF-α), interleukin-6 (IL-6), and high-sensitivity C-reactive protein] and endothelial adhesion molecules [including E-selectin, intercellular adhesion modelecule-1 (ICAM-1), and vascular cell adhesion molecule-1 (VCAM-1)]. These data may help to elucidate the relevant metabolic mechanisms underlying the *FAPB4* gene and T2D[13,14].

2A.2 Research Design and Methods

2A.2.1 Study participants

The WHI-OS is a longitudinal study designed to examine the association between clinical, socioeconomic, behavioral, and dietary risk factors with subsequent incidence of health outcomes, including cardiovascular disease and diabetes. Details regarding the case-control study design have been described elsewhere[15,16]. The study has been reviewed and approved by human subjects review committees at each participating institution, and signed informed consent was obtained from all the participants.

Between September 1994 and December 1998, the WHI-OS enrolled 93,676 postmenopausal women aged 50-79, and ~82,069 had no prior history of cardiovascular disease, cancer, or diabetes at baseline. WHI-OS participants were followed by annual mailed self-administered questionnaires and completed annual medical histories. Incident diabetes cases were identified on the basis of clinical cases that were diagnosed during the follow-up, with primary selection from those reporting treatment with hypoglycemic medication (insulin or oral hypoglycemic agents) and hypoglycemic medication confirmed at the clinic visit at the 3rd year of follow-up. Following the principle of risk-set sampling, for each new case, controls were selected randomly from women who remained free of diabetes at the time the case was identified during follow-up. A total of 1,529 cases were matched with 2,147 controls on age (± 2.5 years), racial/ ethnic group, clinical center (geographic location), time of blood draw (± 0.10h), and length of follow-up. The ethnic groups represented in this study include whites (n=1,899), African Americans (n=1,117), Hispanic/Latino Americans (n=419), and Asian American/Asian Pacific Islanders (n=241). The 1:2 matching ratio was used for minorities to strengthen the power in these smaller samples[16].

2A.2.2 Serum marker measurements

Serum inflammatory cytokines (TNF-α receptor 2, IL-6, and high-sensitivity C-reactive protein) and endothelial adhesion molecules (including E-selectin, ICAM-1, and VCAM-1) were measured for each participant[13,14]. Fasting blood specimens were collected from each participant at baseline and processed locally into separate aliquots including serum, plasma, and buffy coat according to a standardized protocol. The frozen aliquots were then shipped to a central repository, where they were held for long-term storage at -70^{o}C. All biochemical assays were carried out by laboratory staff blinded to case/control status. Blood samples from cases and their matched controls were handled one and the same, shipped in the same batch, and assayed in random order in one run to reduce systematic bias and inter-assay variation. Tumor necrosis factor α receptor 2 (TNF-α-R2) was measured by an enzyme-linked immunosorbent assay (R&D Systems, Minneapolis, Minnesota). Interleukin 6 (IL-6) was measured by an ultrasensitive enzyme-linked immunosorbent assay (R&D Systems). High0sensitivity C-reactive protein (hsCRP) was measured on a chemistry analyzer (Hitachi 911; Roche Diagnostics, Indianapolis, Indiana) using an immunoturbidimetric assay with reagents and calibrators (Denka Seiken Co Ltd, Niigata, Japan)[13]. Soluble E-selectin, ICAM-1 and VCAM-1 were measured by an enzyme-linked immunosorbent assay (R&D Systems, Minneapolis, Minnesota)[14]. The coefficients of variation were 3.5% for TNF-α receptor 2, 7.6% for IL-6, 1.61% for high-sensitivity C-reactive protein, 6.5% for E-selectin, 6.7% for ICAM-1, and 8.9% for VCAM-1.

2A.2.3 SNP frequency estimation and tagging SNP selection

We implemented a two-stage approach to choose tSNPs for genotyping in our large case-control samples[17]. The first stage consists of comprehensive common SNP discovery by

genotyping a total of 25 SNPs in 244 samples randomly selected from the WHI-OS source population. The second stage involved selecting the tSNPs on the basis of linkage disequilibrium (LD) patterns.

In the first stage, we surveyed common genetic variation using the National Center for Biotechnology Information database SN supplemented by HapMap database[18]. Our goal was to capture an initial set of SNPs with high SNP density covering the *FABP4* gene as well as its 30-kbp 5'-upstream and 30-kbp 3'-downstream regions. In total, an initial set of 25 SNPs was selected on the basis of the following criteria: (i) functionality priority: nonsynonymous coding SNPs (cSNPs) and splicing-site SNPs (ssSNPs) were kept in the following order: cSNPs > ssSNPs > 5'-upstream SNPs > 3'-downstream SNPs > intronic SNPs; (ii) minor allele frequency (MAF) ≥ 5% in at least one ethnic group; and (iii) relatively evenly spaced across the genomic region[19].

In the second stage, we identified tSNPs on the basis of LD patterns of those 25 SNPs among 61 women from each ethnic population. Pairwise LD between SNPs was assessed using Lewontins D' statistic and the squared correlation statistic r^2 [20]. The Haploview program was used to calculate the LD coefficient and define haplotype blocks[21,22]. We chose all common tSNPs with special focus on African and white American samples. First, we selected tSNPs n the African-American sample using the r^2-based Tagger program[23]. tSNPs in the African American sample were chosen by finding the minimum set of SNPs with $r^2 \geq 0.08$ and MAF ≥ 5%. We then used backward trimming to reach a minimum set of tSNPs for other ethnic groups to ensure a sufficient, yet nonredundant, parsimonious set of tSNPs. From the initial dense set of 25 SNPs, a total of 10 tSNPs were eventually selected and genotyped in all case-control samples. An additional functional SNP, T87C[8,24,25], was also genotyped in all samples.

2A.2.4 SNP genotyping method

For these 11 SNPs, large-scale genotyping was performed using the TaqMan allelic discrimination method. Specific primers and probes were custom designed by Applied Biosystems (ABI, Foster City, CA). Following PCR amplification, end-point fluorescence was read using the ABI Primer 7900 HT instrument, and genotypes were scored using SDS 2.2.2 Allelic Discrimination Software (Applied Biosystems). We genotyped 5% blind duplicated samples randomly selected to evaluate reproducibility. SNPs wit higher genotyping discordant rate, higher missing genotype rate or deviations from Hardy-Weinberg equilibrium at $P < 0.001$ level were excluded.

2A.2.5 Statistical Analysis

We first estimated the MAF in the control samples for each ethnic group. The Hardy-Weinberg equilibrium test for each of the 11 SNPs was performed using the χ^2 test (degrees of freedom = 1). We also tested for heterogeneity of genotype distributions across ethnicities by the χ^2 test (degrees of freedom = 3; SAS, version 9.2, SAS Institute, Cary, NC).

In both single-SNP and haplotype-based analyses, we employed conditional multivariable logistic regression to calculate odds ratios and 95% confidence intervals for each genetic variant with T2D risk. We made adjustments for covariates and matching factors (such as age, clinical enter, time of blood drawn, and ethnicity), BMI, cigarette smoking (never, past, and current), alcohol intake (never, past, and current), hormone replacement therapy usage (never, past, and current), and total metabolic equivalent value (the energy expended by a person at rest; 1 metabolic equivalent = 1 kcal/kg body weight/h) from recreational physical activity per week at baseline.

In single-SNP analyses, each SNP was coded as an additive, dominant, or recessive genetic model; in the estimation of allelic association with T2D risk, the likelihood ratio test (LRT) was used to test the interaction effect between the genotypes and ethnicity on T2D risk.

In haplotype-based analyses, the Haploview program was sued to define the haplotype block patterns among SNPs[21,22]. Only haplotypes with estimated frequencies ≥ 5% in the combined cases and controls were included for analyses. We tested for ethnic differences in haplotype-associated risks by performing an LRT following the inclusion of an interaction term between the risk haplotypes and ethnicity in the multivariable model. To examine the association between the resulting haplotype/haplotype combinations and T2D risk, the estimate of haplotype dosage was treated as a surrogate variable for the true haplotype. Global LRTs were used to examine whether the frequency distributions of the common haplotypes differed between cases and controls. We also adjusted for covariates that were adjusted in single-SNP association test. To increase the genomic coverage, we employed a sliding-window (window width = 3 SNPs) haplotype-based analysis. For each window, an omnibus LRT was used, which was a χ^2 test (degrees of freedom = number of haplotypes in a particular window = 1). The test used a measure derived on the basis of difference of the logarithmic likelihood of two conditional logistic regression models: (i) the reduced model that does not contain the haplotype covariates, and (ii) the full model that contains the haplotype covariates.

In association tests with serum biomarkers analyses, we transformed all serum biomarker levels in log scale o enhance compliance with normality assumption. We then calculated the geometric mean differences and standard error by genotypes. To determine the effect of the genetic variant on each phenotype level, we calculated the geometric mean difference of these phenotype levels by genotypes using general linear models. Because the inheritance model used

in the single-SNP analysis was the additive model, the result displayed a geometric mean biomarker level increase per copy of the minor allele. All the linear models included matching factors and some other covariates (as mentioned previously) in each f the four ethnic groups. The regression models were performed among cases and controls separately. An LRT was also used to test the interaction effect between the genotypes and ethnicity on serum biomarkers. We also performed some subgroup analyses (biomarker groups) for *FABP4*-T2D.

To account for potential false positive due to multiple comparisons in this study, we calculated the false discovery rate (FDR) by incorporating all P values from multiple tests performed for SNPs and haplotypes in the association tests. The FDR statistics were obtained for each P value, and the FDR statistics with $q \leq 0.05$ were considered as significant[26]. Proc Multtest procedure in SAS 9.2 was used to obtain the adjusted P values.

2A.3 Results

2A.3.1 Estimations of MAF and LD structures of 11 tSNPs in the *FABP4* gene among controls

Figure 1 and **Table** 1 showed the characteristics of the 11 SNPs (one of them did not have an rs number). A total of 4 SNPs (rs2290201, rs8192688, rs2305319, and rs1054135) out of 11 (1 approximately every 1.3kb) were located within the *FABP4* gene (5.3 kbp long). All of the genotyped SNPs did not show statistically significant deviation from Hardy-Weinberg equilibrium at $P < 0.001$ between white, Hispanic, and Asian American/ Asian Pacific Islanders controls. In addition, 2 out of 11 SNPs, that is, rs1486004 (5' flanking region) and T87C (5' promoter region), showed statistically significant deviation from Hardy-Weinberg equilibrium among African American controls.

Figure 2 illustrates the LD structure and haplotype blocks, on the basis of 11 SNPs stratified by ethnicity among controls. With the exception of rs1054135, moderate to strong, pairwise LD was observed between most of the other genotyped SNPs. The locations and lengths of defined haplotype blocks varied slightly between ethnic groups. This partly reflects the differences in allele frequencies between groups. On the whole, three blocks with slightly different boundaries between four ethnic groups could be readily defined. The LD pattern within Block 1 (rs1486004, rs7017115, rs1843560, and rs2200477) was similar across all ethnic groups; however, the block did not include rs1486004 among African Americans. There was evidence for high LD pattern within Block 2 (rs2290201, rs8192688, and rs2305319), which consisted of SNPs within the *FABP4* genomic region. This pattern was similar among whites, African Americans, and American Hispanics. The LD pattern within Block 3 (rs7835371 and rs3824088) was similar between whites, American Hispanics, and Asian Americans/Asian Pacific Islanders. The following SNPs pairs were almost in perfect LD: rs1486004 and rs1843560 (each 5.92 kbp apart) among all groups (D' = 0.969-0.976, r^2 = 0.895-0.93), excluding African American (D' = 0.851 and r^2 = 0.454); rs1486004 and rs2200477 (each 6.35 kbp apart) among all groups (D' = 0.982-1.00, r^2 = 0.932-1.00), excluding African Americans (D' = 0.986, r^2 = 0.651); rs7017115 and rs1843560 (each 4.34 kbp apart) among all groups (D' = 0.974-1.00, r^2 = 0.851-0.91), excluding African Americans (D'= 0.992, r^2 = 0.577); and rs1843560 and rs2200477 (each 4.31 bp part) among all groups (D' = 0.967-1.00, r^2 = 0.903-0.93), excluding African Americans (D' = 0.994, r^2 = 0.418).

2A.3.2 Single-marker association analysis

The association of each SNP with T2D risk in each ethnic group or in the pooled samples was investigated under the additive, dominant, and recessive genetic models. (The results under the additive genetic models are shown in **Table 2**). No evidence of significant associations between all SNPs and T2D risk were found in all ethnic groups. Similar null associations were observed under either the dominant or the recessive genetic model (data not shown).

2A.3.3 Haplotype association analysis

We reconstructed haplotypes on the basis of haplotype block structure in each ethnic group. As shown in **Supplementary Table S1**, haplotypes with all major alleles, that is all 0, occurs most frequently among the participants. Apart from the haplotype with all the major alleles of four SNPs in Block 1 [rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)], 1-1-1-1 is the next most frequent among whites (29.6%), African Americans (41.1%), Hispanic Americans (35.5%), and Asian American/Asian Pacific Islanders (13.4%) in Block 1. In Block 2 [rs2290201(C/T)-rs8192688(C/T)-rs2305319(A/G)], apart from the haplotype with all major alleles, only 1-0-0 had frequency larger than 5%. In Block 3 [rs783537(A/T)-rs3824088(A/G)], 1-1 is the next most frequent among whites (9.2%), African Americans (21.2%), Asian American/Asian Pacific Islanders (52.2%), and Hispanics (17.2%). There was statistical evidence showing frequency difference of haplotype in Black 1 (0-0-0-0 and 1-1-1-1), Block 2 (0-0-0 and 1-0-0), and Block 3 (0-0 and 1-1) between ethnic groups ($P < 0.05$). Such ethnic differences still showed statistical significance after adjusting for multiple testing.

We used 11 SNPs to deduce the common haplotype within each block with a frequency of ≥ 5% in the combined data of all controls, in spite of ethnicity. As shown in **Table 3**, we observed two common haplotypes in Block 1, three common haplotypes in Block 2, and three common haplotypes in Block 3. We first performed global tests to find differences in the overall haplotype frequency between cases and controls among each ethnic group and did not observe statistically significant haplotype effects in all blocks. The haplotypes in the table did not appear to be significantly associated with T2D disease risk. There was also no statistically significant interaction between different ethnic groups.

We also assessed the associations between ethnicity specific haplotypes of the gene and T2D risk. As shown in **Supplementary Table S2**, all odds ratios were not statistically significant among each ethnic group, with the exception of a marginally decreased T2D risk of the 1-0 haplotype (odds ratio: 0.53, 95% confidence interval: 0.27-1.06, P = 0.07). This exception was observed for all carriers vs. all others in Block 3 [rs7835371(A/T)-rs3824088(A/G)] among Hispanics Americans. Matching-adjusted models were also analyzed, and the results were similar to the models adjusted for other covariates, as shown in the **Supplementary Table S2**.

In addition to the previous analyses, the sliding window (with window width = 3 SNPs) was used to analyze haplotype-disease associations. The 11 SNPs generated a total of 9 window frames. No significant association was found between the haplotypes and T2D risk.

2A.3.4 Serum biomarkers

We also analyzed the genotype associations with inflammatory and endothelial biomarkers (including TNF-α receptor 2, IL-6, and high-sensitivity C-reactive protein) and endothelial adhesion molecules (including E-selectin, ICAM-1, and VCAM-1). With the

exception of VCAM-1, none of the inflammatory and endothelial biomarkers showed consistently significant associations. **Tables 4-6** show the geometric mean differences in VCAM-1 levels according to the *FABP4* genotype in each ethnic group for cases and controls separately. After controlling for the covariates (age, clinical center, time of blood draw, ethnicity, and other confounders, including hormone replacement therapy use, alcohol consumption, cigarette smoking, BMI, physical activity), we studied the genetic and lifestyle predictors of T2D.

Among incident cases of the African American group, plasma VCAM-1 level were -1.08 ng/ml (s.e. = 1.03 ng/ml, P = 0.01) lower in subjects with SNP rs1486004 T allele, -1.07 ng/ml (s.e. = 1.03 ng/ml, P = 0.03) lower in subjects with rs7017115 G allele, -1.07 ng/ml (s.e. = 1.03 ng/ml, P = 0.05) lower in subjects with rs1843560 G allele, -1.09 ng/ml (s.e. = 1.04 ng/ml, P = 0.02) lower in subjects with rs2200477 G allele, and -1.12 ng/ml (s.e. = 1.04 ng/ml, P = 0.002) lower in subjects with rs2290201 T allele. After adjusting for multiple comparisons, rs2290201 still showed a significant decreasing trend in the geometric mean differences of plasma VCAM-1 level (adjusted P = 0.02). Among African American controls, plasma VCAM-1 level were -1.05 ng/ml (s.e. = 1.02 ng/ml, P = 0.02) in SNP rs2200477. However, this decreasing trend was no longer significant after adjusting for multiple testing (**Table 5**). The interaction between SNP and VCAM-1 levels was significant ($P < 0.05$) for both cases and controls, even after multiple testing adjustment.

Data from **Tables 4 and 5** showed that only four SNPs in Block 1 [rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)] were significantly associated with lower VCAM-1 levels among African American women. We further performed a haplotype-VCAM-1 association analysis in this group (**Table 6**). The plasma VCAM-1 level was elevated by 1.09

ng/ml (s.e. = 1.04 ng/ml, P = 0.01) in haplotype 0-0-0-0, and the level was lowered by -1.08 ng/ml (s.e. = 1.03 ng/ml, P = 0.01) in haplotype 1-1-1-1. After adjusting for multiple testing, both trends remained significant (adjusted P = 0.04).

2A.4 Discussion

Based on association tests in 1,529 T2D cases and 2,147 matched controls from a multiethnic cohort of American postmenopausal women, there were little evidence supporting the influence of the *FABP4* gene on T2D risk. None of the genotyped SNPs located near/within the *FABP4* gene showed any significant association with T2D. Further, our haplotype-based analyses, as well as sliding window haplotype analysis did not reveal any significant findings. These null results were consistent across different ethnic groups, including whites, African Americans, Hispanics, and Asian American/Asian Pacific Islanders, which results strongly suggest that common genetic variants in *FABP4* may not confer a susceptibility to T2D.

The SNPs genotyped in our study covered 7 kbp upstream and 29 kbp downstream region around the *FABP4* gene. We may not have enough coverage to detect functional variants around the *FABP4* genomic region. We selected the common variant on the basis of available genetic information, and we may not be able to identify potentially function variants, especially rare variants with relatively low frequencies (< 5%) that were also not in high LD with any chosen in our study. Sample sizes were relatively small in Hispanic Americans and Asian American/Asian Pacific Islanders, which may not afford sufficient power to detect any moderate association for the *FABP4* variants and T2D risk[27,28].

In addition to T2D, the variants in FABP have been associated with diseases (obesity[1,7,29], atherosclerosis[4,9], and cardiovascular disorders[30]) that share some metabolic traits (e.g. insulin

resistance) and molecular pathways (e.g. inflammatory activity and macrophage cholesterol trafficking) that also lead to T2D[1,3]. Biological data, mainly from animal models, highlight adipocyte P2 in several key pathways in the pathogenesis of T2D, such as lipolytic response, lipolysis-associated insulin secretion[10,11], plasma glucose and insulin level[12], and cytokine secretion[3]. Nevertheless, we should interpret these results cautiously because different cell types, experimental conditions, and quantitative methods were adopted in previous studies. The human *FABP4* gene, located on chromosome 8q21, encodes a 131 amino-acid precursor protein, including at least two different isoforms regarding alternative translation or splicing processes[31]. It is possible that specific FABP4 protein isoform may function as a tissue-specific metabolic modulator.

In the literature, most studies regarding the association between *FABP4* and T2D in humans have focused on the serum AFABP rather than its genotype. *FABP4* plasma concentrations were reported to be increased with the early presence of metabolic syndrome (MS) components, inflammation, as well as oxidation markers in Spanish and white T2D individuals[6,32]. Serum AFABP has also been found to be associated with glucose dysregulation and predictive of T2D development in a Chinese cohort study[24]. One study reported a promoter polymorphism, T87C, of the AFABP gene. It reduced adipose tissue AFABP mRNA expression and was associated with lower risk for T2D and cardiovascular disease[25]. However, limited investigation has been done on the direct association of the *FABP4* gene with T2D. Our findings on T87C were not consistent with previous studies[25]. This is likely due to limited sample size, especially after stratifying according to ethnic group. Our large multiethnic case-control study of postmenopausal women showed some suggestive evidence of the relation of the *FABP4* genotype and an intermediate phenotype, such as VCAM-1 levels. Endothelial activation, as

indicated by elevated levels of soluble adhesion molecules, has been associated with T2D[33]. In a prospective analysis with the same source population, high circulating levels of endothelial biomarkers, including VCAM-1, were significantly associated with risk of T2D[14]. In the present study, we found that T2D cases with polymorphisms at SNPs (rs1486004, rs7017115, rs1843560, rs2200477, and rs2290201) showed significantly lower VCAM-1 levels among African Americans alone. Consistent results were also shown in the haplotype analysis. The polymorphism at SNP rs2200477 showed significantly lower VCAM-1 level among African American controls as well. However, because the SNPs are intronic SNPs, it is likely that they are in LD with the nearby causal SNPs and also may simply represent false-positive associations because of multiple testing. Most P values for significant SNP and haplotype associations were above 5% after performing a false-discovery rate for multiple testing. Further studies are needed to confirm these findings[28].

Our study demonstrated substantial heterogeneity in the frequencies of both the alleles and haplotypes in the *FABP4* gene among different ethnic groups. Participants from different geographic locations might have different T2D risk, mainly due to their different environmental exposures or different genetic background. Ethnicity may contribute to the *FABP4* genetic variants, affecting T2D risk differently. Thus, it has been suggested that population stratification may lead to false-positive results[17,34]. We have cautiously selected the control women to be representative of the WHI-OS source population. We have also conducted stratified analyses on the basis of ethnicity, in order to address the potential bias from population stratification. However, at the same time, these ethnicity-stratified analyses lack the power to detect potential associations for the *FABP4* SNPs and haplotypes with T2D. Due to the comprehensive screening of SNPs, this may be the first study to determine the haplotype structure of *FABP4*.

We performed our comprehensive association analyses of informative SNPs (MAF ≥ 5%), as well as haplotypes based on the LD patterns in each ethnic group constructed from the HapMap database. The analyses provided powerful evidence against a main-effect association between the overall risk of T2D and variants in *FABP4* that are common among the four ethnic groups, although the lack of association between *FABP4* variants and T2D may also be due to insufficient power and inability to detect changes in phenotypes. If the effect size of each *FABP4* variant or haplotype is modest, it would require very large samples to achieve sufficient power for detection. Further studies, like large-scale association and genome-wide association studies, will be necessary to confirm the null association between genetic variants of *FABP4* and T2D risk and examine the potential effect modification from biomarkers.

The strengths of our study include the well-established T2D incidence ascertainment methods, which employed standard protocols to define cases and controls following the principle of risk-set sampling; excluded all the prevalent cases from the original case-control sampling set; and matched each case-control pair on age, ethnicity, clinical center, time of blood draw, as well as follow-up time. Finally, our findings may be generalizable to women of similar age and ethnically diverse background because our study included ethnically diverse women from 40 states in the United States.

In conclusion, our large, multiethnic, case-control study of postmenopausal women did not provide evidence to support the notion that common genetic variants in *FABP4* may contribute significantly to the pathogenesis of T2D. However, we cannot exclude the possibility of a modest genetic effect as well as genotype-phenotype association. There is some suggestive evidence for an association between the *FABP4* genotypes and VCAM-1 levels in African American women alone, although further replication studies are warranted.

Table 1. The location, relative distances, minor allele frequencies (MAFs) of 11 SNPs chosen in *FABP4* genomic region.

SNP ID	Location [a]	Relative distance (bp) [a]	Allele[b]	Minor allele frequency (MAF%)					P-value for heterogeneity across ethnicity
				White (n=939)	Black (n=746)	Asian (n=165)	Hispanic (n=279)	Pooled (n=2129)	
rs1486004	5' flanking	0	C/T	31.2	67.3	14.1	39.1	43.5	<0.0001
rs7017115	5' flanking	1584	A/G	30.5	43.5	13.5	35.1	34.4	0.001
rs1843560	5' flanking	5919	C/G	31.4	56.9	14.6	38.0	39.9	<0.0001
rs2200477	5' flanking	6349	C/G	31.6	75.8	14.1	40.4	46.9	<0.0001
T87C	5' promoter	29811	T/C	2.05	0.75	0	1.09	1.31	0.70
rs2290201	Intron 1	30645	C/T	28.8	65.3	68.2	41.5	46.4	<0.0001
rs8192688	Intron 1	32498	C/T	16.6	9.12	0.30	10.8	12.0	0.002
rs2305319	Intron 2	33473	A/G	17.0	16.6	7.01	16.6	16.0	0.17
rs1054135	Exon 4	34587	A/G	6.8	24.3	11.8	9.64	13.7	0.003
rs7835371	3' UTR	60172	A/T	17.0	30.1	67.4	29.6	27.2	<0.0001
rs3824088	3' UTR	64218	A/G	8.7	23.0	51.6	17.2	18.2	<0.0001

[a] Location and relative distance between SNPs are based on the contig position of contig NT_008183.18.

[b] Major/minor allele.

[c] P-values were estimated by a chi-square test (df=3) for genotype distribution across the four ethnic groups.

Table 2. Single-SNP association studies of the 11 SNPs in the *FABP4* genomic region with T2D risk.

		Adjusted OR (95% CI) [a]				
SNP ID	**Allele**	**White (947/952)**	**Black (366/751)**	**Asian (76/165)**	**Hispanic (140/279)**	**Pooled (1529/2147)**
rs1486004	T/C	0.98 (0.80-1.19)	0.91 (0.74-1.12)	0.99 (0.49-1.99)	0.86 (0.59-1.25)	0.95 (0.83-1.07)
rs7017115	A/G	1.01 (0.82-1.24)	1.07 (0.85-1.35)	1.39 (0.72-2.67)	0.84 (0.58-1.24)	1.01 (0.89-1.16)
rs1843560	G/C	0.98 (0.80-1.20)	0.96 (0.77-1.21)	1.00 (0.51-1.96)	0.79 (0.54-1.15)	0.95 (0.83-1.08)
rs2200477	G/C	0.94 (0.77-1.15)	0.79 (0.62-1.01)	1.22 (0.64-2.32)	0.89 (0.62-1.27)	0.90 (0.79-1.03)
T87C	T/C	1.80 (0.87-3.75)	0.62 (0.14-2.65)	---[b]	0.69 (0.11-4.17)	1.26 (0.74-2.16)
rs2290201	C/T	0.95 (0.77-1.17)	1.04 (0.84-1.30)	1.46 (0.90-2.38)	0.98 (0.69-1.40)	1.02 (0.90-1.16)
rs8192688	C/T	0.92 (0.72-1.18)	1.15 (0.80-1.65)	---[b]	0.83 (0.46-1.51)	0.97 (0.81-1.17)
rs2305319	A/G	1.04 (0.82-1.34)	1.02 (0.77-1.35)	1.41 (0.57-3.52)	0.76 (0.47-1.23)	0.99 (0.84-1.17)
rs1054135	G/A	0.85 (0.59-1.24)	1.07 (0.84-1.37)	1.26 (0.62-2.55)	0.67 (0.35-1.27)	1.01 (0.84-1.22)
rs7835371	T/A	1.02 (0.79-1.31)	1.02 (0.81-1.28)	1.32 (0.82-2.12)	0.85 (0.55-1.33)	1.03 (0.89-1.19)
rs3824088	A/G	1.08 (0.76-1.54)	0.98 (0.75-1.28)	1.26 (0.75-2.12)	1.38 (0.76-2.49)	1.09 (0.91-1.31)

[a] ORs are estimated using conditional logistic regression adjusted for age, clinical center, time of blood draw, ethnicity and other confounders including HRT use, alcohol consumption, cigarette smoking, BMI, and physical activity. The numbers of participants (cases/controls) were included in the parenthesis.

[b] Result is difficult to interpret because of small sample within strata.

Table 3. Haplotype-based associations between *FABP4* common haplotypes and T2D risk.

Haplotype	Haplotype specific OR(95% CI) [a, b, c] White (947/952)	Black (366/751)	Asian (76/165)	Hispanic (140/279)	Pooled (1529/2147)	P-value for ethnic interaction
Block 1						
rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)						
0-0-0-0	1.04 (0.85-1.27)	1.21 (0.96-1.53)	0.77 (0.42-1.41)	1.13 (0.79-1.62)	1.08 (0.95-1.23)	0.61
1-1-1-1	1.01 (0.82-1.24)	1.07 (0.86-1.33)	1.09 (0.55-2.17)	0.85 (0.58-1.24)	1.00 (0.88-1.14)	0.61
P-values for global testing	0.51	0.33	0.93	0.53		
Block 2						
rs2290201(C/T)-rs8192688(C/T)-rs12305319(A/G)						
0-0-0	1.05 (0.86-1.29)	0.98 (0.79-1.22)	0.76 (0.47-1.21)	1.04 (0.73-1.48)	0.99 (0.87-1.13)	0.84
1-0-0	0.90 (0.66-1.22)	1.02 (0.83-1.25)	1.18 (0.76-1.83)	1.18 (0.77-1.81)	1.04 (0.90-1.20)	0.65
1-1-1	0.98 (0.76-1.26)	1.16 (0.81-1.67)	––– [b]	0.79 (0.43-1.45)	1.01 (0.84-1.22)	0.54
P-values for global testing	0.54	0.89	0.37	0.80		
Block 3						
rs7835371(A/T)-rs3824088(A/G)						
0-0	0.96 (0.75-1.23)	0.98 (0.78-1.22)	0.77 (0.48-1.23)	1.17 (0.75-1.83)	0.96 (0.83-1.11)	0.86
1-1	1.07 (0.77-1.49)	1.06 (0.82-1.38)	1.29 (0.77-2.15)	1.20 (0.70-2.06)	1.11 (0.93-1.32)	0.89
1-0	1.01 (0.70-1.47)	0.92 (0.62-1.38)	1.12 (0.56-2.22)	0.58 (0.30-1.10)	0.91 (0.73-1.15)	0.36
P-values for global testing	0.45	0.95	0.23	0.15		

[a] ORs are estimated using conditional logistic regression adjusted for age, clinical center, time of blood draw, and other confounders including HRT use, alcohol consumption, cigarette smoking, BMI and physical activity.

[b] Result is difficult to interpret because of small sample size within strata

Table 4. Geometric mean differences[a] in Vascular cell adhesion molecule levels according to tagged SNPs and ethnicity among cases.

	White (n=947)		Black (n=366)	
SNP	mean difference (SE)	P for trend	mean difference (SE)	P for trend[b]
rs1486004	1.03 (1.02)	0.15	-1.08 (1.03)	**0.01**
rs7017115	1.02 (1.02)	0.38	-1.07 (1.03)	**0.03**
rs1843560	1.02(1.02)	0.35	-1.07 (1.03)	**0.05**
rs2200477	1.02 (1.02)	0.29	-1.09 (1.04)	**0.02**
T87C	-1.04 (1.05)	0.46	1.20 (1.26)	0.43
rs2290201	1.01 (1.02)	0.76	-1.12 (1.04)	**0.002***
rs8192688	1.01 (1.02)	0.55	1.00 (1.06)	0.95
rs2305319	1.01 (1.02)	0.76	-1.04 (1.05)	0.41
rs1054135	1.01 (1.03)	0.87	1.01(1.04)	0.76
rs7835371	-1.02 (1.02)	0.37	1.01 (1.04)	0.84
rs3824088	-1.00 (1.03)	0.99	-1.05 (1.04)	0.23
	Asian/Pacific Islanders (n=76)		**Hispanic (n=140)**	
SNP	mean difference (SE)	P for trend	mean difference (SE)	P for trend
rs1486004	-1.15 (1.09)	0.09	1.01 (1.06)	0.80
rs7017115	-1.08 (1.08)	0.34	-1.04 (1.06)	0.48
rs1843560	-1.09 (1.08)	0.31	-1.02 (1.06)	0.77
rs2200477	-1.06 (1.08)	0.43	1.02 (1.05)	0.64
T87C	---[c]	.	-1.45 (1.50)	0.37
rs2290201	1.03 (1.06)	0.67	1.01 (1.06)	0.80
rs8192688	---[c]	.	-1.04 (1.10)	0.66
rs2305319	-1.10 (1.14)	0.49	1.04 (1.08)	0.62
rs1054135	-1.04 (1.09)	0.61	1.07 (1.11)	0.50
rs7835371	1.04 (1.06)	0.48	1.01 (1.07)	0.84
rs3824088	1.07 (1.06)	0.23	-1.02 (1.10)	0.82

[a] Geometric mean difference (SE) for each SNP was calculated using general linear regression models with adjustment for matching factors (age, clinical center, and time of

blood draw), and other confounders including cigarette smoking, alcohol consumption, hormone replacement therapy, and physical activity. Negative sign indicates decreasing level of plasma VCAM-1 with additional copy of risk allele in the corresponding SNP.

[b] Adjusted p-value = 0.02 after FDR

[c] Result is difficult to interpret because of small sample size within strata

Table 5. Geometric mean differences[a] in Vascular cell adhesion molecule levels according to tagged SNPs and ethnicity among controls.

	White (n=952)		Black (n=751)	
SNP	mean difference (SE)	P for trend	mean difference (SE)	P for trend
rs1486004	1.01 (1.02)	0.54	-1.03 (1.02)	0.10
rs7017115	1.00 (1.02)	0.83	-1.01 (1.02)	0.55
rs1843560	1.01 (1.02)	0.75	-1.03 (1.02)	0.11
rs2200477	1.02 (1.02)	0.18	-1.05 (1.02)	**0.02**[b]
T87C	-1.02 (1.05)	0.68	1.02 (1.12)	0.89
rs2290201	1.01 (1.02)	0.42	1.00 (1.02)	0.91
rs8192688	-1.01 (1.02)	0.78	-1.01 (1.04)	0.81
rs2305319	-1.00 (1.02)	0.83	-1.01 (1.03)	0.81
rs1054135	1.06 (1.03)	0.07	1.02 (1.02)	0.36
rs7835371	1.02 (1.02)	0.41	1.03 (1.02)	0.19
rs3824088	1.01 (1.03)	0.62	1.02 (1.03)	0.38
	Asian/Pacific Islanders (n=165)		**Hispanic (n=279)**	
SNP	mean difference (SE)	P for trend	mean difference (SE)	P for trend
rs1486004	-1.01 (1.06)	0.90	-1.02 (1.04)	0.55
rs7017115	-1.02 (1.06)	0.70	-1.02 (1.04)	0.56
rs1843560	-1.05 (1.06)	0.34	-1.01 (1.04)	0.74
rs2200477	-1.01 (1.06)	0.88	-1.02 (1.04)	0.66
T87C	---[c]	.	-1.20 (1.23)	0.37
rs2290201	1.00 (1.04)	0.93	1.05 (1.04)	0.17
rs8192688	-1.38 (1.39)	0.33	1.01 (1.06)	0.80
rs2305319	-1.03 (1.08)	0.72	1.01 (1.05)	0.87
rs1054135	1.02 (1.07)	0.72	1.03 (1.06)	0.65
rs7835371	1.02 (1.04)	0.67	1.07 (1.04)	0.10
rs3824088	-1.00 (1.04)	0.90	1.07 (1.05)	0.15

[a] Geometric mean difference (SE) for each SNP was calculated using general linear

regression models with adjustment for matching factors (age, clinical center, and time of blood draw), and other confounders including cigarette smoking, alcohol consumption, hormone replacement therapy, and physical activity. Negative sign indicates decreasing level of plasma VCAM-1 with additional copy of risk allele in the corresponding SNP.

[b] Adjusted p-value = 0.23 after FDR

[c] Result is difficult to interpret because of small sample size within strata

Table 6. Geometric mean differences in VCAM levels according to specific haplotype among black cases (n=417).

Haplotype[a]	Haplotype specific Geometric Mean Difference (95% CI)[b] mean difference[b,c]	p for trend
rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)		
0-0-0-0	1.09 (1.02-1.17)	**0.01***
0-0-0-1	-1.02(-1.20- -0.87)	0.78
1-0-0-1	-1.00(-1.11- -0.90)	0.95
1-0-1-1	1.00 (0.90-1.11)	0.97
1-1-1-1	-1.08 (-1.15- -1.02)	**0.01***

[a] Haplotype observed with ≥ 0.05 frequency in this ethnic group (0=major allele, 1=minor allele).
[*] Adjusted p-value <0.05 after FDR

Supplementary Table 1. Haplotype frequencies reconstructed from SNPs.

Haplotype [a]	Haplotype frequency (%)					P-value for frequency difference of haplotype among ethnic groups [b]
	White (947/952)	Black (366/751)	Asian (76/165)	Hispanic (140/279)	Pooled (1529/2147)	
Block 1 rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)						
0-0-0-0	67.5	24.9	84.1	57.7	54.6	<0.0001***,[c]
1-1-1-1	29.6	41.1	13.4	35.5	32.7	0.0025**,[d]
Block 2 rs2290201(C/T)-rs8192688(C/T)-rs12305319(A/G)						
0-0-0	71.8	33.9	30.9	59.2	56.1	<0.0001***,[e]
1-0-0	11.0	49.0	61.8	24.4	27.5	<0.0001***,[f]
Block 3 rs7835371(A/T)-rs3824088(A/G)						
0-0	83.2	67.8	31.7	72.0	73.9	<0.0001***,[g]
1-0	7.5	9.2	16.0	10.4	8.8	0.29
1-1	9.2	21.2	52.2	17.2	16.7	<0.0001***,[g]

[a] Only haplotypes with frequency > 5% are reported (0, major allele; 1, minor allele)

[b] * $p < 0.05$; ** $p < 0.01$; *** $P < 0.001$

[c] P<0.0001 without adjusted for multiple testing; by setting the p-value as 0.00005, FDR q-value=0.0002

[d] P=0.0025 without adjusted for multiple testing; FDR q-value=0.0069

[e] P<0.0001 without adjusted for multiple testing; by setting the p-value as 0.00005, FDR q-value=0.0001

[f] P<0.0001 without adjusted for multiple testing; by setting the p-value as 0.00005, FDR q-value=0.0001

[g] P<0.0001 without adjusted for multiple testing; by setting the p-value as 0.0005, FDR q-value<0.0001

Supplementary Table 2. Association between ethnic-specific haplotypes of FABP4 and T2D risk.

Haplotype[a]	Haplotype frequency		OR (95% CI) [b]		
	Cases	Controls	All carriers versus all others	Homozygous versus all non-carriers	P-values for global testing[c]
White	n=947	n=952			
Block 1	**rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)**				
0-0-0-0	68.3	66.7	1.16 (0.78-1.72)	1.00 (0.59-1.69)	0.51
1-1-1-1	29.8	29.4	1.03 (0.79-1.34)	0.89 (0.50-1.59)	
Block 2	**rs2290201(C/T)-rs8192688(C/T)-rs12305319(A/G)**				
0-0-0	72.6	71.0	1.08 (0.69-1.70)	1.14 (0.61-2.14)	0.54
1-0-0	10.4	11.6	0.95 (0.68-1.32)	0.37 (0.12-1.19)	
1-1-1	15.9	16.2	0.97 (0.73-1.28)	0.75 (0.31-1.82)	
Block 3	**rs7835371(A/T)-rs3824088(A/G)**				
0-0	83.6	82.8	0.74 (0.37-1.49)	0.67 (0.30-1.51)	0.45
1-0	6.7	8.3	0.95 (0.65-1.41)	1.60 (0.38-6.84)	
1-1	9.7	8.8	1.09 (0.76-1.55)	0.87 (0.36-2.09)	
Black	n=366	n=751			
Block 1	**rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)**				
0-0-0	26.4	24.1	1.25 (0.92-1.71)	0.99 (0.59-1.65)	0.24
0-0-1	16.9	19.0	0.76 (0.55-1.05)	0.87 (0.35-2.16)	
0-1-1	12.0	13.4	0.77 (0.53-1.11)	0.98 (0.51-1.86)	
1-1-1	44.5	43.4	1.05 (0.75-1.47)	1.04 (0.55-1.96)	
Block 2	**rs2290201(C/T)-rs8192688(C/T)-rs12305319(A/G)**				
0-0-0	32.5	34.5	1.08 (0.80-1.45)	0.76 (0.44-1.31)	0.89
1-0-0	49.7	48.6	1.20 (0.86-1.67)	1.31 (0.76-2.27)	
1-0-1	8.6	7.5	0.88 (0.58-1.33)	0.76 (0.38-1.51)	
1-1-1	9.0	9.0	1.18 (0.81-1.71)	0.49 (0.08-2.88)	
Hispanic	n=140	n=279			
Block 1	**rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)**				

0-0-0-0	55.4	58.9	0.97 (0.50-1.87)	0.85 (0.31-2.36)	0.53
1-1-1-1	38.1	34.2	0.72 (0.44-1.18)	1.13 (0.40-3.24)	
Block 2	**rs2290201(C/T)-rs8192688(C/T)-rs12305319(A/G)**				
0-0-0	61.0	58.3	1.34 (0.68-2.67)	0.99 (0.42-2.35)	0.80
1-0-0	24.2	24.5	1.43 (0.82-2.49)	1.44 (0.57-3.61)	
1-1-1	9.2	10.4	0.78 (0.42-1.46)	0.98 (0.32-2.99)	
Block 3	**rs7835371(A/T)-rs3824088(A/G)**				
0-0	75.7	70.1	1.65 (0.71-3.86)	3.17 (0.35-29.0)	0.15
1-0	8.0	11.6	0.53 (0.27-1.06)	0.56 (0.06-5.70)	
1-1	15.9	17.9	1.38 (0.74-2.56)	0.59 (0.15-2.26)	
Asian/Pacific Islander	n=76	n=165			
Block 1	**rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)**				
0-0-0-0	81.7	85.1	0.70 (0.19-2.59)	0.72 (0.18-2.87)	0.93
1-1-1-1	13.6	13.3	1.11 (0.51-2.43)	0.89 (0.07-12.0)	
Block 2	**rs7835371(A/T)-rs3824088(A/G)**				
0-0	29.4	32.8	0.76 (0.39-1.47)	0.56 (0.14-2.25)	0.23
1-0	16.3	15.9	1.55 (0.68-3.56)	0.71 (0.04-13.2)	
1-1	54.4	51.2	1.89 (0.79-4.50)	---[d]	

[a] Haplotypes with ≥0.05 frequency within each block were inferred in each of four ethnic groups (0, major allele; 1, minor allele).

[b] ORs for each haplotype was calculated using conditional logistic regression models with adjustment for matching factors (age, clinical center and time of blood draw) and other confounders including BMI, HRT, alcohol consumption, cigarette smoking and physical activity.

[c] Analyzed by global permutation test adjusted for multiple testing

[d] result is difficult to interpret because of small sample size within strata

Supplementary Table 3. Subgroup analysis for *FABP4*-T2D stratified by VCAM-1 among black.

SNP/Haplotype[a]		**Haplotype specific OR (95% CI)**[b,c]	
Low VCAM-1		***High VCAM-1***	
rs1486004	0.97 (0.74-1.27)	**rs1486004**	0.74 (0.43-1.26)
rs7017115	1.13 (0.85-1.52)	**rs7017115**	0.58 (0.32-1.05)
rs1843560	0.92 (0.69-1.23)	**rs1843560**	0.67 (0.37-1.22)
rs2200477	0.82 (0.59-1.14)	**rs2200477**	0.71 (0.40-1.27)
T87C	---[d]	**T87C**	---[d]
rs2290201	1.17 (0.87-1.57)	**rs2290201**	0.90 (0.50-1.61)
rs8192688	1.11 (0.71-1.74)	**rs8192688**	1.10 (0.43-2.82)
rs2305319	1.09 (0.77-1.54)	**rs2305319**	0.60 (0.30-1.20)
rs1054135	1.24 (0.90-1.71)	**rs1054135**	0.93 (0.51-1.68)
rs7835371	1.02 (0.76-1.38)	**rs7835371**	1.48 (0.81-2.71)
rs3824088	1.03 (0.72-1.47)	**rs3824088**	1.08 (0.44-2.62)
rs1486004(C/T)-rs7017115(A/G)-rs1843560(C/G)-rs2200477(C/G)			
0-0-0-0	1.25 (0.87-1.80)	**0-0-0-0**	1.04 (0.67-1.62)
0-0-0-1	1.17 (0.56-2.45)	**0-0-0-1**	0.82 (0.30-2.20)
1-0-0-1	0.78 (0.48-1.26)	**1-0-0-1**	1.40 (0.68-2.86)
1-0-1-1	0.68 (0.41-1.15)	**1-0-1-1**	1.35 (0.59-3.06)
1-1-1-1	1.09 (0.79-1.50)	**1-1-1-1**	0.76 (0.50-1.15)

[a] Haplotype observed with ≥ 0.05 frequency in this ethnic group (0=major allele, 1=minor allele).

[b] ORs are estimated using conditional logistic regression adjusted for age, clinical center, time of blood draw, and other confounders including HRT use, alcohol consumption, cigarette smoking, BMI and physical activity.

[c] VCAM-1 level was catergorized by its median value among African American controls.

[d] Result is difficult to interpret because of small sample within strata.

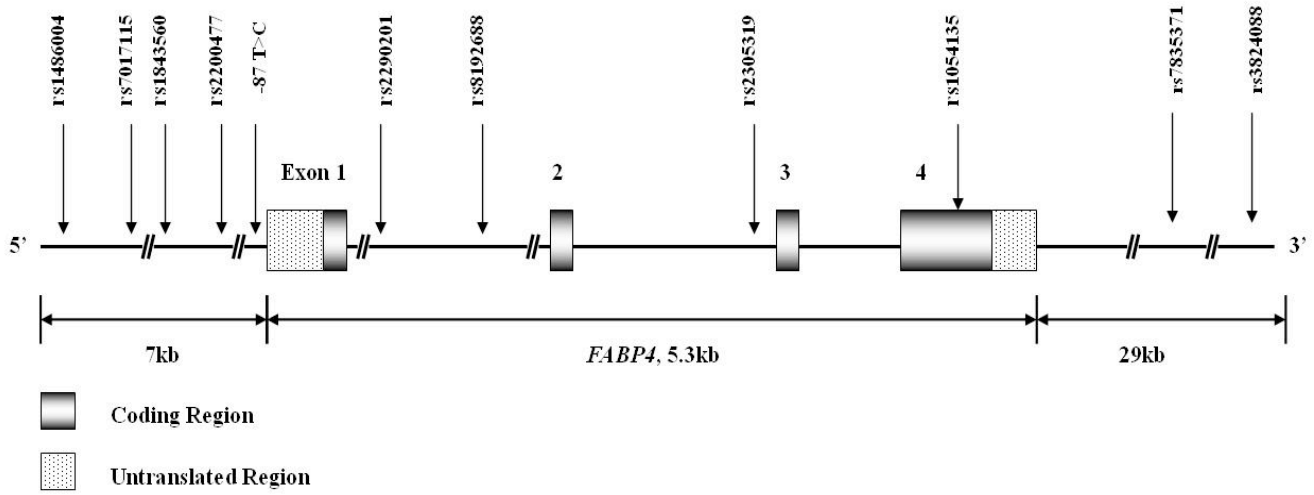


Figure 1. The schematic presentation shows the position of 11 SNPs that spans the *FABP4* genomic region. These markers within the *FABP4* region consisted of one SNP in the 5' promoter region (-87 T>C), one missense SNP (rs1054135) and three intronic SNPs (rs2290201, rs8192688, and rs2305319). The missense SNP was without high LD with other SNPs.

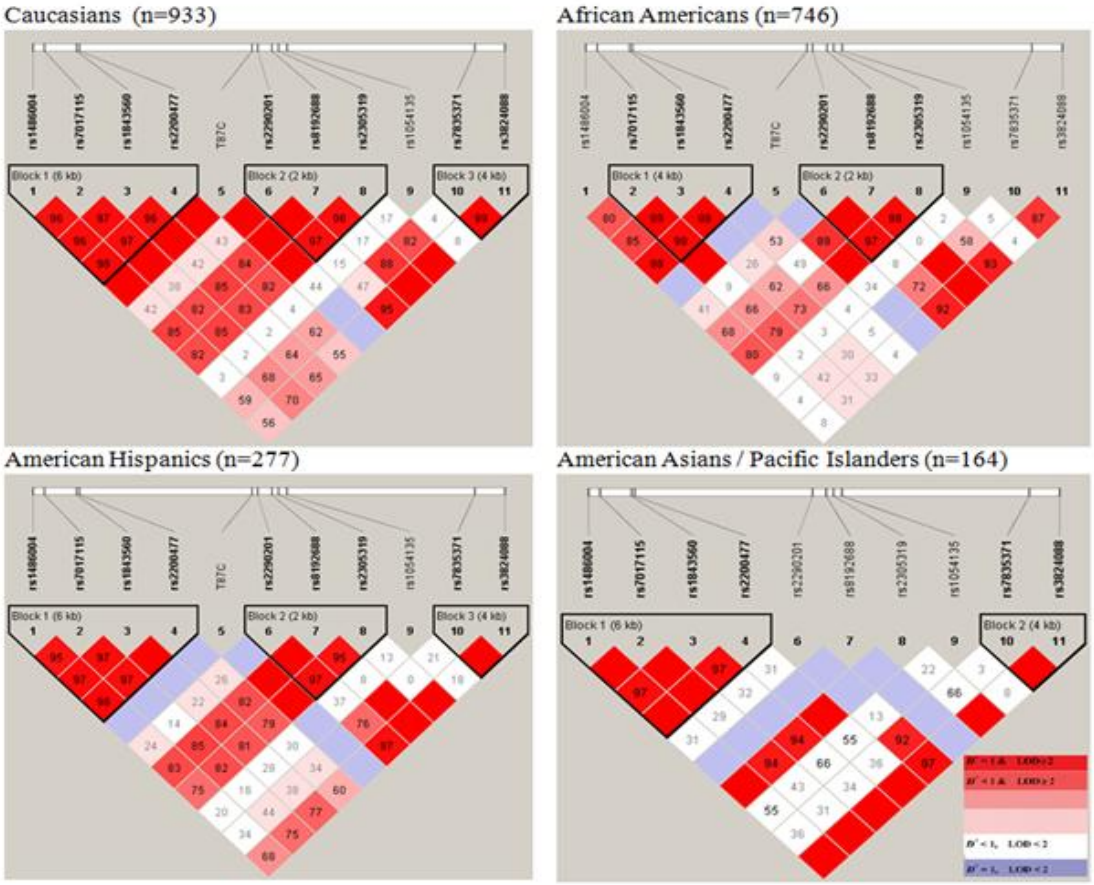


Figure 2. The haplotype blocks shown above define the LD structures between the 11 tSNPs near or within the *FABP4* gene from four ethnic groups (Caucasians, African Americans, American Hispanics, and American Asians/ Pacific Islanders). The upper diagram gives the relative physical position of each SNP. The pairwise LD between all tSNPs is indicated by the respective diamonds for each SNP combination (with red illustrating strong LD (D' > 0.8) and logarithm of odds score (LOD) ≥ 2. LD strength between the selected SNPs is determined by the 90% confidence limits of D' statistics.

This work was published in the *Obesity* journal on January 28, 2010.

2A.6 Reference

1. Furuhashi M, Hotamisligil GS. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. Nat Rev Drug Discov 2008;7:489-503.

2. Lapsys NM, Kriketos AD, Lim-Fraser M, et al. Expression of genes involved in lipid metabolism correlate with peroxisome proliferator-activated receptor gamma expression in human skeletal muscle. J Clin Endocrinol Metab 2000;85:4293-7.

3. Makowski L, Brittingham KC, Reynolds JM, Suttles J, Hotamisligil GS. The fatty acid-binding protein, aP2, coordinates macrophage cholesterol trafficking and inflammatory activity. Macrophage expression of aP2 impacts peroxisome proliferator-activated receptor gamma and IkappaB kinase activities. J Biol Chem 2005;280:12888-95.

4. Makowski L, Hotamisligil GS. The role of fatty acid binding proteins in metabolic syndrome and atherosclerosis. Curr Opin Lipidol 2005;16:543-8.

5. Boord JB, Fazio S, Linton MF. Cytoplasmic fatty acid-binding proteins: emerging roles in metabolism and atherosclerosis. Curr Opin Lipidol 2002;13:141-7.

6. Cabre A, Lazaro I, Girona J, et al. Fatty acid binding protein 4 is increased in metabolic syndrome and with thiazolidinedione treatment in diabetic patients. Atherosclerosis 2007;195:e150-8.

7. Maeda K, Cao H, Kono K, et al. Adipocyte/macrophage fatty acid binding proteins control integrated metabolic responses in obesity and diabetes. Cell Metab 2005;1:107-19.

8. Ordovas JM. Identification of a functional polymorphism at the adipose fatty acid binding protein gene (FABP4) and demonstration of its association with cardiovascular disease: a path to follow. Nutr Rev 2007;65:130-4.

9. Yeung DC, Xu A, Cheung CW, et al. Serum adipocyte fatty acid-binding protein levels were independently associated with carotid atherosclerosis. Arterioscler Thromb Vasc Biol 2007;27:1796-802.

10. Hotamisligil GS, Johnson RS, Distel RJ, Ellis R, Papaioannou VE, Spiegelman BM. Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. Science 1996;274:1377-9.

11. Scheja L, Makowski L, Uysal KT, et al. Altered insulin secretion associated with reduced lipolytic efficiency in aP2-/- mice. Diabetes 1999;48:1987-94.

12. Uysal KT, Scheja L, Wiesbrock SM, Bonner-Weir S, Hotamisligil GS. Improved glucose and lipid metabolism in genetically obese mice lacking aP2. Endocrinology 2000;141:3388-96.

13. Liu S, Tinker L, Song Y, et al. A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. Arch Intern Med 2007;167:1676-85.

14. Song Y, Manson JE, Tinker L, et al. Circulating levels of endothelial adhesion molecules and risk of diabetes in an ethnically diverse cohort of women. Diabetes 2007;56:1898-904.

15. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 1998;19:61-109.

16. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol 2003;13:S5-17.

17. Hsu YH, Niu T, Song Y, Tinker L, Kuller LH, Liu S. Genetic variants in the UCP2-UCP3 gene cluster and risk of diabetes in the Women's Health Initiative Observational Study. Diabetes 2008;57:1101-7.

18. A haplotype map of the human genome. Nature 2005;437:1299-320.

19. Wang L, Liu S, Niu T, Xu X. SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. BMC Bioinformatics 2005;6:60.

20. Lee CC, You NC, Song Y, et al. Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: findings from the Women's Health Initiative Observational Cohort. Clin Chem 2009;55:351-60.

21. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

22. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

23. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

24. Tso AW, Xu A, Sham PC, et al. Serum adipocyte fatty acid binding protein as a new biomarker predicting the development of type 2 diabetes: a 10-year prospective study in a Chinese cohort. Diabetes Care 2007;30:2667-72.

25. Tuncman G, Erbay E, Hom X, et al. A genetic variant at the fatty acid-binding protein aP2 locus reduces the risk for hypertriglyceridemia, type 2 diabetes, and cardiovascular disease. Proc Natl Acad Sci U S A 2006;103:6970-5.

26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57:289-300.

27. Song Y, Hsu YH, Niu T, Manson JE, Buring JE, Liu S. Common genetic variants of the ion channel transient receptor potential membrane melastatin 6 and 7 (TRPM6 and TRPM7), magnesium intake, and risk of type 2 diabetes in women. BMC Med Genet 2009;10:4.

28. Song Y, You NC, Hsu YH, et al. Common genetic variation in calpain-10 gene (CAPN10) and diabetes risk in a multi-ethnic cohort of American postmenopausal women. Hum Mol Genet 2007;16:2960-71.

29. Wellen KE, Hotamisligil GS. Inflammation, stress, and diabetes. J Clin Invest 2005;115:1111-9.

30. Xu A, Wang Y, Xu JY, et al. Adipocyte fatty acid-binding protein is a plasma biomarker closely associated with obesity and metabolic syndrome. Clin Chem 2006;52:405-13.

31. Prinsen CF, de Bruijn DR, Merkx GF, Veerkamp JH. Assignment of the human adipocyte fatty acid-binding protein gene (FABP4) to chromosome 8q21 using somatic cell hybrid and fluorescence in situ hybridization techniques. Genomics 1997;40:207-9.

32. Stejskal D, Karpisek M. Adipocyte fatty acid binding protein in a Caucasian population: a new marker of metabolic syndrome? Eur J Clin Invest 2006;36:621-5.

33. Boulbou MS, Koukoulis GN, Makri ED, Petinaki EA, Gourgoulianis KI, Germenis AE. Circulating adhesion molecules levels in type 2 diabetes mellitus and hypertension. Int J Cardiol 2005;98:39-44.

34. Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037-48.

Chapter 2B:

Common Variations in the Genes Encoding C-Reactive Protein, Tumor Necrosis Factor-α, and Interleukin-6, and the Risk of Clinical Diabetes in the Women's Health Initiative Observational Study

2B.1 Introduction

Inflammatory markers such as high-sensitivity C-reactive protein (hsCRP), tumor necrosis factor-α (TNF-α), and interkeukin-6 (IL-6) have been implicated s possible etiologic factors in the development of obesity, diabetes, and cardiovascular disease[1-10]. One of our previous studies reported that among postmenopausal women enrolled in the Women's Health Initiative Observation Study (WHI-OS) increased circulating concentrations of hsCRP, TNF-α, and IL-6 were significantly associated with an increased diabetes risk[11]. Recently, common genetic variants in the *CRP* (C-reactive protein, pentraxin-related) gene were associated with their corresponding plasma marker concentrations in Europeans[12], European Americans[13,14], African Americans[1,13], and Pima Indians[15]. To date, relatively few studies have investigated the associations of common variants in the genes encoding TNF-α ad IL-6 with their plasma concentrations or looked for direct association of *CRP*, *TNF* (tumor necrosis factor), of IL6 [interleukin 6 (interferon, beta 2)] gene variants with diabetes risk, especially in a multiethnic population.

We conducted a comprehensive assessment of the association of genetic variants for TNF-α and IL-6 with plasma concentrations of these 2 inflammation markers in a large case-control study nested within the WHI-OS. We also investigated the association of variations in the *CRP*, *TNF*, and *IL6* genes with diabetes risk in the same group of women.

2B.2 Research Design and Methods

2B.2.1 Study Participants

Details regarding the case-control study design of the WHI-OS have been described in the Section 2.2.1 in Chapter 2 (P.12-13)[16,17].

2B.2.2 Plasma marker measurements

Plasma concentrations of hsCRP, TNF-α receptor 2 (TNF-α_R2), and IL-6 were measured for each participant. TNNF-α-R2 is measured more reliably in frozen samples than TNF-α itself, and TNF-α_R2 concentrations correlate well with TNF-α concentrations[18,19]. In brief, the CVs were 1.6%, 3.5%, and 7.6% for hsCRP, TNF-α-R2, and IL-6, respectively[11].

2B.2.3 Haplotype-tagging single nucleotide polymorphism (SNP) selection and genotyping methods

Please refer to Section 2.2.3 in Chapter 2A (P.13-15) for details regarding the SNP selection and genotyping methods[20-22].

2B.2.4 Statistical Analysis

We first assessed the allele frequency and Hardy-Weinberg equilibrium (HWE) for each SNP among the controls for each ethnic group. Next, we used a χ^2 test to test for heterogeneity in genotype distributions across ethnic groups (SAS 9.2; SAS Institute). In multivariable regression models, we adjusted for matching factors (age, clinical center, time of blood draw, and ethnicity) and other covariates [body mass index, cigarette smoking (never, past, and current), alcohol intake (never, past, and current), family history of diabetes, hormone replacement therapy use (never, past, and current), and the total metabolic equivalent (MET) value from the individual's recreational physical activity per week at baseline]. To investigate the relationship between SNPs and plasma markers, we log-transformed the plasma marker data

with skewed distributions to improve compliance with the normality assumption. We calculated the differences in the mean logarithms of plasma marker concentrations according to each genotyped tSNP by fitting general linear models that treated plasma marker concentrations as dependent variables and tSNPs as independent variables. An additive model was used. The results of this analysis were expressed as an increase or decrease in the difference in the mean logarithms of the plasma marker per each additional copy of the reference allele. Likelihood ratio tests were used to test for the effects of genotype-ethnicity interaction on inflammatory marker concentrations.

In assessing the relationship between each SNP and diabetes risk, we used multivariable logistic regression (conditional on matching) to calculate odds ratios and 95% CIs. Each SNP was coded as an additive genetic model in estimating allelic association with diabetes risk; the likelihood ratio test was used to test the effect of genotype-ethnicity interaction on diabetes risk.

To account in the study for potential false positives due to multiple comparisons, we calculated the false-discovery rate (FDR) by incorporating all P values from multiple tests performed for the association of SNPs and plasma markers as well as the association of SNPs and diabetes risk[23]. The FDR statistics were obtained for each P value, and FDR statistics with q values < 0.05 were considered statistically significant.

2B.3 Results

2B.3.1 Estimation of allele frequencies

As shown in **Table 1**, the allele frequencies of 9 SNPs in the *TNF* gene and 13 SNPs in the *IL6* gene differed significantly by ethnicity. The 13 SNPs in the *CRP* gene have been published in another study done by our group and is shown in **Supplementary Table 1**[13]. **Figure**

1 and 2 schematically present the locations of the SNPs along the *TNF* and *IL6* genes according to the gene structure presented in NCBI Entrez Gene (http://www.ncbi.nlm.nih.gov/gene). **Supplementary Figure 1** shows the schematic diagram of the *CRP* genomic region. In the *TNF* gene, rs2239704, rs1041981, and rs3093661 in white women deviated significantly from the HWE among the controls. None of the 14 SNPs in the *IL6* gene showed any statistically significant deviation from HWE among the controls of each ethnic group.

2B.3.2 Associations of genetic variants with plasma biomarkers and diabetes risk

In **Table 2**, half of the 16 SNPs in the *TNF* gene were associated with plasma TNF-α-R2 concentrations in white women. For 4 SNPs, carriers of each additional copy of the reference allele had lower TNF-α-R2 concentrations [range for the decrease in mean logarithm per allele (SE), +0.03 (0.01) to -0.04 (0.02); all adjusted q values were < 0.05 after FDR]. In contrast, carriers of the reference alleles for the 4 other SNPs (rs909253, rs1041981, rs1800629, and rs2256974) had higher TNF-α-R2 concentrations [range of the increase in mean logarithm per allele (SE), 0.04 (0.01) to 0.05 (0.01); all adjusted q values were < 0.05 after FDR]. After adjusting for multiple testing, we found no significant association between any of the *IL6* gene variants and IL-6 concentration.

After adjusting for matching factors, other covariates, and multiple comparisons, as shown in **Table 3**, we found no evidence of any significant associations between any of the SNPs among the 3 genes (*CRP*, *TNF*, and *IL6*) and diabetes risk (all q values were > 0.05). Our findings were confirmed in our analysis of 4 additional models with various covariates (particularly the effect of controlling for family history of diabetes and body mass index) to

investigate the potential independent associations of the inflammation marker variants of interest with diabetes risk, which are shown in **Supplementary Tables 2a-d**.

2B.4 Discussion

In this large multiethnic cohort of postmenopausal women, 8 common variants of the gene encoding *TNF* were associated with the plasma TNF-α-R2 concentration in whites, whereas we found no association between common variants of the *IL6* gene and the plasma IL-6 concentration. No variants of the *CRP*, *TNF*, or *IL6* genes were significantly associated with increased diabetes risk after we corrected for multiple comparisons.

One of the 8 SNPs (rs1800629) in the *TNF* gene that we found to be associated with plasma TNF-α concentrations in whites was associated in a prior study with TNR-α concentration in the same ethnic group[24]. The null findings for a relationship between common variants of the same gene and diabetes risk were consistent with results from previous studies of European and Chinese populations[25-28]. The presence of the A allele of SNP rs1800629 in the *TNF* gene among Brazilian individuals older than 48 years has been associated with increased hsCRP concentrations[29]. Although our results were not statistically significant, the direction of the association between this *TNF* variant and hsCRP concentration was the same in our samples. Furthermore, as we demonstrated in our prior study, increased hsCRP concentrations were associated with an increased diabetes risk[11]. Therefore, if this *TNF* variant is in fact associated with increased hsCRP concentrations, then it may play an indirect role in the pathogenesis of type 2 diabetes. A meta-analysis has indicated that individuals who carry this TNR-α variant are at higher risk of developing obesity than control individuals, suggesting that the *TNF* gene is involved in the pathogenesis of the metabolic syndrome[30]. Obesity is a well-known risk factor

for the development of type 2 diabetes. On the other hand, another study indicated that this TNF-α variant was not associated with insulin resistance in young Asian Indians[31]. Taken together, the data indicate that the *TNF* gene may not lead directly to the development of diabetes, but it may play an interactive role with other factors, such as CRP, in the pathogenesis of diabetes.

We observed no significant associations in our samples between *IL6* variants and the plasma IL-6 concentration. One study showed the genetic variant rs10499563 to be significantly associated with increased IL-6 concentrations in individuals in an acute inflammatory state[30]. The presence of acute inflammation may affect the association between this genetic variant and the plasma IL-6 concentration. In general, we presume that the women in our study did not have acute inflammation at the time of blood draw, which may account for this discrepancy. A study of 1953 Korean men and women found that the rs1800796 G/G genotype was associated with increased serum IL-6 concentrations[32]. This result is consistent with our analysis, which showed that carriers of each additional copy of the G allele in this SNP were associated with increased IL-6 concentrations in the Asian population, although our result was not statistically significant. Inconsistent findings regarding the association between this gene and diabetes risk have been reported previously in several case-control, prospective population-based studies and meta-analyses[33-37]. A joint analysis of the data for the individual participants from 21 studies observed that the C allele of the rs1800795 SNP in the *IL6* gene was associated with a reduced risk of diabetes[34], whereas a meta-analysis indicated a null association between the same SNP in the *IL6* and diabetes risk[36]. In general, the literature lacks reports of studies that have examined the associations of common variants in the *TNF* and *IL6* gene regions with the corresponding plasma marker concentrations and diabetes risk, particularly in a multiethnic cohort.

Assuming an additive model, we observed no significant associations between the variants in the *CRP* gene and the risk of clinical diabetes, a result consistent with prior findings[38]. Although these genetic variants have substantial and independent associations with the plasma hsCRP concentration[13], our prospective data do not support a direct heritable role for CRP in the development of diabetes.

The lack of significant genetic associations in the current study may be due to insufficient statistical power, especially among the Hispanic and Asian women. Nevertheless, our study was well powered to detect effects for alleles shared across all ethnic groups. In fact, we had > 80% power to detect a relative risk of ≥ 1.25 for risk alleles with frequencies from 10% to 70%. Additionally, our study included only postmenopausal women, and therefore our results may not be generalized to men or younger women.

In conclusion, 8 common genetic variants of the *TNF* gene were associated with the plasma TNF-α-R2 concentration among whites in this large multiethnic case-control study of postmenopausal women, although these common *TNF* variants were not associated with a risk of clinical diabetes. Common *IL6* variants were not associated with IL-6 concentration or diabetes risk, nor were common *CRP* variants associated with the risk of clinical diabetes. Our data indicate modest associations between *TNF* gene variants and circulating concentrations of TNF-α-R2. Common variants of the genes encoding *CRP*, *TNF*, and *IL6* were not significantly associated with the risk of clinical diabetes in postmenopausal women.

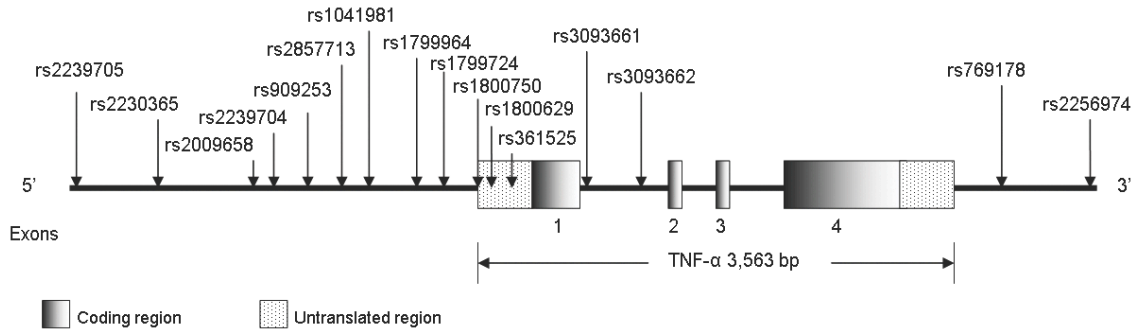


Figure 1. Human *TNF* gene (chromosome 6p21.3) and SNP locations.

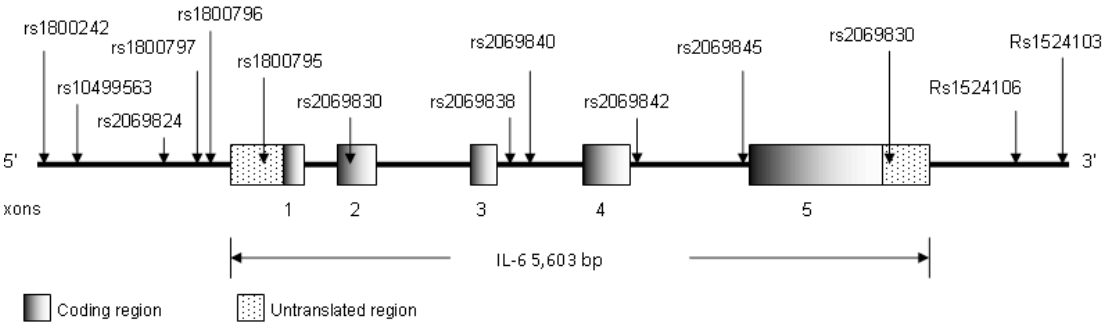


Figure 2. Human *IL6* gene (chromosome 7p21) and SNP locations.

Table 1. Location and allele frequencies of genotyped tSNPs in *TNF* and *IL6* genes in controls.

				Minor allele frequency (%)					
Gene	dbSNP ID [a]	Location	Allele[b]	Whites	Blacks	Hispanics	Asian/ Pacific Islanders	Pooled	P[c]
TNF-α				(n=939)	(n=737)	(n=275)	(n=156)	(n=2,107)	
	rs2239705	5' flanking	C/T	18.12	13.07	25.63	20.00	17.48	<0.0001
	rs2230365	5' flanking	C/T	15.09	6.68	17.63	25.15	13.26	<0.0001
	rs2009658	5' flanking	C/G	16.02	13.11	18.18	17.38	15.39	0.048
	rs2239704	5' flanking	G/T	41.77	28.21	39.96	40.30	36.71	<0.0001
	rs909253	5' flanking	T/C	34.22	48.59	36.05	37.35	39.73	<0.0001
	rs2857713	5' flanking	T/C	26.70	28.36	26.71	22.26	26.93	0.28
	rs1041981	5' flanking	C/A	33.58	48.25	35.61	38.04	39.27	<0.0001
	rs1799964	5' flanking	T/C	21.59	18.18	21.22	20.86	20.30	0.11
	rs1799724	5' flanking	C/T	10.74	3.95	16.91	17.38	9.69	<0.0001
	rs1800750	5' flanking	G/A	1.78	2.28	1.26	0	1.75	0.17
	rs1800629	5' flanking	G/A	15.88	12.72	9.39	2.42	12.90	<0.0001
	rs361525	5' flanking	G/A	5.67	4.62	3.06	3.44	4.80	0.22
	rs3093661	Intron1	G/A	3.98	4.10	2.36	3.46	3.77	0.50
	rs3093662	Intron1	A/G	8.15	8.67	6.27	3.61	7.73	0.045
	rs769178	3'flanking	C/A	10.15	4.03	17.33	16.67	9.45	<0.0001
	rs2256974	3'flanking	G/T	18.21	34.26	25.45	36.54	26.13	<0.0001
IL-6				(n=928)	(n=738)	(n=276)	(n=163)	(n=2,105)	
	rs1880242	5' flanking	G/T	53.17	79.22	56.34	26.52	60.69	<0.0001
	rs10499563	5' flanking	C/T	77.32	82.02	80.88	82.08	79.81	0.0069
	rs2069824	5' flanking	C/T	91.08	85.79	91.82	99.08	89.92	<0.0001
	rs1800797	5' flanking	A/G	62.2	90.34	78.88	96.04	76.88	<0.0001
	rs1800796	5' flanking	C/G	93.66	90.39	78.52	30.49	85.59	<0.0001
	rs1800795	5' flanking	C/G	61.29	90.19	78.68	97.26	76.58	<0.0001
	rs2069830	Exon2	C/T	0.05	9.14	0.36	0	3.26	<0.0001
	rs2069838	Intron3	C/T	0.32	6.65	1.08	0	2.62	<0.0001
	rs2069840	Intron3	C/G	34.72	18.71	31.93	7.72	26.59	<0.0001
	rs2069842	Intron4	A/G	99.79	92.81	99.09	100	97.26	<0.0001
	rs2069845	Intron4	A/G	42.00	34.35	29.86	3.05	34.7	<0.0001

rs2069861	3' flanking	C/T	8.27	1.96	3.97	0.31	4.88	<0.0001
rs1524106	3' flanking	C/A	67.4	47.1	45.31	8.84	52.84	<0.0001
rs1524103	3' flanking	G/C	25.54	41.6	31.52	19.63	31.5	<0.0001

[a] From the NCBI dbSNP.

[b] Reference alleles are indicated in parentheses.

[c] P values were estimated by a χ^2 test (df=3) for genotype distribution across the 4 ethnic groups.

Table 2. Differences in mean logarithms in plasma concentrations according to corresponding SNPs stratified by ethnicity.

Gene	SNP ID	White	Black	Hispanics	Asian/Pacific Islanders	All	P-value for ethnic interaction[b]
		Difference in mean logarithms (SE)[a]	Difference in mean logarithms (SE)	Difference in mean logarithms (SE)	Difference in mean logarithms (SE)	Difference in mean logarithms (SE)	
TNF-α		(n=1,592)	(n=912)	(n=328)	(n=209)	(n=3,041)	
	rs2239705	-0.02 (0.01)	-0.03 (0.02)	0.002 (0.03)	0.04 (0.04)	0.003 (0.01)	0.03
	rs2230365	**-0.04 (0.01)[c]**	-0.04 (0.03)	-0.02 (0.03)	-0.06 (0.03)	-0.02 (0.01)	0.86
	rs2009658	-0.03 (0.01)	-0.04 (0.02)	0.01 (0.03)	-0.05 (0.04)	-0.02 (0.01)	0.73
	rs2239704	-0.02 (0.01)	-0.002 (0.02)	-0.005 (0.02)	-0.01 (0.03)	0.001 (0.01)	0.25
	rs909253	**0.04 (0.01)[d]**	0.01 (0.02)	0.02 (0.02)	0.06 (0.03)	0.02 (0.01)	0.70
	rs2857713	**-0.03 (0.01)[e]**	-0.02 (0.02)	0.001 (0.03)	-0.07 (0.04)	**-0.03 (0.01)[l]**	0.38
	rs1041981	**0.05 (0.01)[f]**	0.003 (0.02)	0.01 (0.02)	0.06 (0.03)	0.02 (0.01)	0.51
	rs1799964	**-0.03 (0.01)[g]**	-0.02 (0.02)	0.01 (0.03)	-0.07 (0.04)	-0.02 (0.01)	0.65
	rs1799724	-0.03 (0.02)	-0.08 (0.04)	-0.01 (0.03)	0.05 (0.04)	-0.002 (0.01)	0.01
	rs1800750	0.03 (0.04)	0.04 (0.05)	0.08 (0.08)	-0.35 (0.31)	0.03 (0.03)	0.93
	rs1800629	**0.04 (0.01)[h]**	-0.01 (0.02)	0.03 (0.04)	0.01 (0.08)	0.02 (0.01)	0.07
	rs361525	-0.03 (0.02)	0.02 (0.03)	-0.02 (0.06)	-0.09 (0.08)	-0.02 (0.02)	0.58
	rs3093661	-0.04 (0.02)	-0.03 (0.04)	-0.08 (0.08)	-0.08 (0.08)	-0.05 (0.02)	0.44
	rs3093662	**-0.04 (0.02)[i]**	0.01 (0.03)	-0.05 (0.05)	-0.10 (0.08)	-0.03 (0.01)	0.44
	rs769178	-0.03 (0.02)	-0.07 (0.04)	-0.01 (0.03)	0.04 (0.04)	-0.001 (0.01)	0.02
	rs2256974	**0.04 (0.01)[j]**	0.01 (0.02)	0.02 (0.03)	0.05 (0.03)	0.01 (0.01)	0.54
IL-6		(n=1,608)	(n=918)	(n=327)	(n=207)	(n=3,060)	
	rs1880242	-0.01 (0.01)	-0.03 (0.03)	-0.06 (0.04)	-0.01 (0.05)	-0.01 (0.01)	0.64
	rs10499563	-0.003 (0.01)	-0.003 (0.02)	0.003 (0.03)	0.02 (0.05)	-0.002 (0.01)	0.69
	rs2069824	-0.01 (0.02)	-0.01 (0.03)	-0.02 (0.05)	-0.08 (0.09)	-0.01 (0.01)	0.28
	rs1800797	-0.05 (0.03)	-0.05 (0.07)	-0.10 (0.08)	-0.28 (0.26)	-0.04 (0.02)	0.02
	rs1800796	0.03 (0.05)	0.06 (0.07)	-0.10 (0.07)	0.02 (0.10)	0.04 (0.03)	0.07
	rs1800795	-0.04 (0.03)	-0.05 (0.07)	-0.10 (0.08)	-0.23 (0.27)	-0.03 (0.02)	0.01
	rs2069830	-0.32 (0.50)	0.05 (0.07)	-0.06 (0.53)	---[k]	0.09 (0.06)	0.62
	rs2069838	0.05 (0.27)	0.005 (0.08)	-0.36 (0.38)	---[k]	0.04 (0.07)	0.57
	rs2069840	-0.01 (0.03)	0.08 (0.05)	-0.17 (0.06)	-0.06 (0.17)	-0.01 (0.02)	0.66
	rs2069842	-0.16 (0.27)	-0.05 (0.08)	0.46 (0.26)	---[k]	-0.08 (0.06)	0.17

rs2069845	0.04 (0.03)	-0.001 (0.04)	0.12 (0.07)	0.10 (0.40)	0.04 (0.02)	0.01
rs2069861	0.05 (0.05)	0.13 (0.15)	0.21 (0.14)	---[k]	0.07 (0.04)	0.04
rs1524106	0.03 (0.03)	0.05 (0.04)	-0.01 (0.06)	-0.02 (0.15)	0.03 (0.02)	0.04
rs1524103	-0.03 (0.03)	-0.03 (0.04)	-0.05 (0.07)	0.002 (0.12)	-0.01 (0.02)	0.50

[a] Difference in mean logarithms and standard error per additional reference allele of each SNP was calculated using general linear regression models with adjustment for matching factors (age, clinical center, and time of blood draw), incidence of diabetes, and other confounders including BMI, hormone replacement therapy, alcohol consumption, cigarette smoking, family history of diabetes, and physical activity.

[b] P-value was estimated based on a log-likelihood ratio test for interaction between each genotype and ethnicity on plasma concentrations.

[c] P-value was 0.003 with q-value = 0.01 after FDR.

[d] P-value was 0.000041 with q-value = 0.0003 after FDR.

[e] P-value was 0.006 with q-value = 0.02 after FDR.

[f] P-value was 0.000006 with q-value < 0.0001 after FDR.

[g] P-value was 0.007 with q-value = 0.02 after FDR.

[h] P-value was 0.006 with q-value = 0.02 after FDR.

[i] P-value was 0.04 with q-value = 0.04 after FDR.

[j] P-value was 0.004 with q-value = 0.01 after FDR.

[k] Result is difficult to interpret because of small sample size within strata.

[l] P-value was 0.0009 with q-value = 0.01 after FDR.

Table 3. The multivariable adjusted odds ratio (95% CI) for diabetes risk associated with genetic variants, as calculated with the additive effect model.

		Adjusted OR (95% CI)[a]					P-value for ethnic interaction[c]
Gene	SNP ID	Whites	Blacks	Hispanics	Asian/Pacific Islanders	All	
CRP		(n=870/865)[b]	(n=303/638)	(n=115/242)	(n=67/154)	(n=1355/1899)	
	rs4275453	0.87(0.71- 1.07)	0.85(0.67 - 1.09)	0.93(0.57 - 1.52)	1.04(0.51 - 2.15)	0.85(0.74 – 0.98)[i]	0.92
	rs2808634	0.82(0.66 – 1.03)	0.74(0.51 - 1.07)	0.84(0.48 - 1.49)	1.93(0.73 - 5.11)	0.81(0.68 – 0.97)[j]	0.34
	rs3093059	0.86(0.59 - 1.25)	1.17(0.88 - 1.56)	0.62(0.30 - 1.25)	0.76(0.32 - 1.80)	1.03(0.84 - 1.26)	0.38
	rs2794521	0.82(0.65 – 1.02)	0.78(0.54 - 1.14)	0.76(0.41 - 1.38)	1.66(0.66 - 4.19)	0.81(0.68 – 0.97)[k]	0.43
	rs1417938	1.03(0.84 - 1.28)	1.36(0.93 – 2.00)	1.07(0.65 - 1.76)	0.75(0.23 - 2.47)	1.09(0.92 - 1.28)	0.91
	rs1800947	1.25(0.85 - 1.84)	0.52(0.15 - 1.81)	1.46(0.40 - 5.31)	1.00(0.33 – 2.96)	1.18(0.86 - 1.63)	0.55
	rs1130864	1.09(0.88 - 1.35)	1.18(0.86 - 1.63)	0.91(0.56 - 1.48)	1.15(0.40 - 3.31)	1.09(0.93 - 1.28)	0.76
	rs1205	1.10(0.89 - 1.37)	0.80(0.59 - 1.10)	1.19(0.74 - 1.92)	0.82(0.43 - 1.56)	1.01(0.86 - 1.17)	0.61
	rs3093075	0.90(0.61 - 1.32)	1.17(0.88 - 1.55)	0.57(0.328- 1.16)	1.09(0.47 - 2.51)	1.06(0.87 - 1.31)	0.76
	rs3093068	0.81(0.59 - 1.13)	1.30(0.97 - 1.73)	0.84(0.42 - 1.66)	1.11(0.47 - 2.61)	1.05(0.86 - 1.28)	0.62
	rs2808629	1.15(0.93 - 1.43)	0.73(0.54 – 0.99)[d]	1.30(0.81 - 2.10)	0.72(0.39 - 1.35)	1.00(0.86 - 1.16)	0.34
	rs2369146	0.79(0.62 – 1.00)	0.96(0.74 - 1.24)	1.23(0.76 – 2.00)	2.19(0.80 – 5.97)	0.92(0.78 - 1.08)	0.02[l]
	rs1470515	1.21(1.99 - 1.48)	0.70(0.52 - 0.94)[e]	0.97(0.6 2- 1.53)	0.95(0.50 - 1.83)	1.01(0.87 - 1.17)	0.29
TNF		(n=867/882)	(n=306/638)	(n=115/244)	(n=68/155)	(n=1356/1919)	
	rs2239705	1.02(0.79-1.31)	0.74(0.50-1.09)	0.96(0.61-1.53)	0.75(0.31-1.79)	0.92(0.76-1.10)	0.39
	rs2230365	1.06(0.81-1.40)	0.91(0.53-1.54)	0.77(0.42-1.44)	0.86(0.47-1.55)	0.96(0.78-1.17)	0.79
	rs2009658	0.96(0.73-1.27)	0.90(0.60-1.33)	0.75(0.42-1.33)	0.92(0.44-1.94)	0.92(0.76-1.11)	0.75
	rs2239704	1.09(0.90-1.31)	0.79(0.59-1.07)	1.33(0.92-1.92)	0.66(0.35-1.23)	1.00(0.88-1.15)	0.36
	rs909253	0.99(0.81-1.21)	1.40(1.06-1.85)	0.83(0.54-1.28)	1.77(0.92-3.38)	1.09(0.95-1.25)	0.52
	rs2857713	0.84(0.67-1.05)	0.94(0.70-1.27)	0.90(0.55-1.46)	0.97(0.50-1.91)	0.90(0.77-1.05)	0.27
	rs1041981	1.02(0.83-1.24)	1.39(1.06-1.83)	0.84(0.55-1.28)	1.79(0.93-3.43)	1.11(0.97-1.28)	0.60
	rs1799964	0.89(0.70-1.13)	0.86(0.61-1.20)	1.01(0.60-1.72)	1.01(0.50-2.03)	0.91(0.76-1.07)	0.28
	rs1799724	1.04(0.76-1.43)	0.62(0.30-1.28)	1.14(0.67-1.94)	0.80(0.34-1.92)	0.97(0.77-1.23)	0.57
	rs1800750	0.51(0.24-1.09)	1.22(0.55-2.69)	2.58(0.55-12.2)	---[g]	0.90(0.54-1.50)	0.01[m]
	rs1800629	0.91(0.70-1.18)	1.24(0.86-1.78)	1.19(0.56-2.52)	6.29(1.14-34.8)[h]	1.04(0.85-1.27)	0.02[n]
	rs361525	0.74(0.49-1.11)	1.04(0.57-1.87)	2.23(0.77-6.46)	1.56(0.39-6.17)	0.93(0.69-1.26)	0.09
	rs3093661	0.79(0.50-1.26)	0.66(0.33-1.32)	1.56(0.44-5.50)	0.48(0.08-2.92)	0.79(0.56-1.12)	0.95
	rs3093662	0.87(0.62-1.23)	0.70(0.43-1.12)	1.09(0.47-2.52)	1.24(0.33-4.63)	0.84(0.66-1.08)	0.61

	rs769178	1.06(0.77-1.47)	0.58(0.28-1.20)	1.09(0.64-1.86)	0.94(0.39-2.27)	0.98(0.77-1.24)	0.70
	rs2256974	1.10(0.85-1.42)	1.26(0.96-1.67)	0.80(0.49-1.30)	1.49(0.76-2.93)	1.14(0.97-1.34)	0.71
IL6		(n=876/888)	(v304/641)	(n=115/244)	(n=67/154)	(n=1362/1927)	
	rs1880242	1.01 (0.90-1.13)	1.04 (0.89-1.22)	1.14 (0.74-1.76)	1.14 (0.81-1.60)	1.03 (0.95-1.12)	0.36
	rs10499563	0.97 (0.88-1.08)	0.96 (0.83-1.11)	0.97 (0.76-1.23)	0.80 (0.58-1.11)	0.95 (0.89-1.03)	0.36
	rs2069824	0.93 (0.81-1.08)	1.06 (0.89-1.27)	0.66 (0.45-0.98)[f]	---[g]	0.98 (0.89-1.09)	0.57
	rs1800797	1.10 (0.90-1.34)	0.84 (0.57-1.26)	1.12 (0.62-2.04)	0.66 (0.09-4.77)	1.04 (0.89-1.23)	0.82
	rs1800796	1.18 (0.79-1.76)	1.34 (0.88-2.04)	1.02 (0.60-1.73)	0.90 (0.52-1.55)	1.10 (0.89-1.37)	0.66
	rs1800795	1.06 (0.87-1.29)	0.93 (0.61-1.41)	1.35 (0.76-2.41)	0.31 (0.01-6.92)	1.05 (0.89-1.23)	0.77
	rs2069830	1.81 (0.09-38.7)	0.99 (0.62-1.56)	---[g]	---[g]	1.00 (0.64-1.57)	0.66
	rs2069838	0.76 (0.15-3.88)	1.22 (0.76-1.97)	---[g]	---[g]	1.08 (0.69-1.69)	0.76
	rs2069840	1.16 (0.93-1.43)	0.89 (0.65-1.23)	0.94 (0.61-1.44)	0.64 (0.24-1.69)	1.06 (0.91-1.23)	0.28
	rs2069842	1.07 (0.11-10.4)	1.43 (0.87-2.37)	1.07 (0.14-8.55)	---[g]	1.26 (0.79-2.02)	0.98
	rs2069845	0.91 (0.75-1.11)	1.13 (0.87-1.48)	0.82 (0.51-1.32)	---[g]	0.95 (0.82-1.09)	0.49
	rs2069861	0.88 (0.62-1.25)	1.12 (0.43-2.91)	1.16 (0.43-3.16)	---[g]	0.98 (0.73-1.33)	0.92
	rs1524106	0.86 (0.70-1.06)	1.13 (0.88-1.44)	0.87 (0.58-1.31)	0.97 (0.40-2.34)	0.95 (0.83-1.10)	0.53
	rs1524103	1.17 (0.93-1.47)	0.94 (0.73-1.20)	1.16 (0.73-1.85)	0.75 (0.38-1.47)	1.02 (0.88-1.18)	0.33

[a] Odds Ratio (OR) per additional reference allele of each SNP was calculated for additive genetic effect model; ORs were estimated using conditional logistic regression models adjusted for matching factors (age, clinical center, time of blood draw, and ethnicity), BMI, cigarette smoking, alcohol intake, hormone replacement therapy, family history of diabetes and physical activity.

[b] Sample size for each ethnic group was shown for each plasma marker in the format of (cases/controls).

[C] P-value was estimated based on a log-likelihood ratio test for interaction between each genotype and ethnicity on diabetes risk.

[d] P-value was 0.04 with q-value = 0.28 after FDR.

[e] P-value was 0.02 with q-value = 0.25 after FDR.

[f] P-value was 0.04 with q-value = 0.56 after FDR.

[g] Result was difficult to interpret because of small sample size within strata.

[h] P-value was 0.04 with q-value = 0.46 after FDR.

[i] P-value was 0.03 with q-value = 0.12 after FDR.

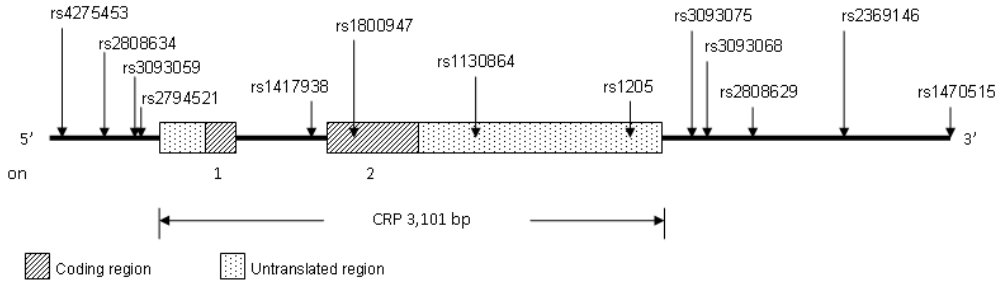
[j] P-value was 0.02 with q-value = 0.12 after FDR.

[k] P-value was 0.02 with q-value = 0.12 after FDR.

[l] The P-value became 0.24 after FDR.

[m] The P-value became 0.15 after FDR.

[n] The P-value became 0.15 after FDR.



Supplementary Figure 1. Human *CRP* gene (Chromosome 1q21-q23) and SNP locations. (Original source: Lee CC, You NC, Song Y, Hsu YH, Manson J, Nathan L, et al. Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: findings from the Women's Health Initiative Observational Cohort. Clin Chem 2009;55:351-60)

Supplementary Table 1. Location and minor allele frequencies (MAF) of genotyped tagging SNPs in CRP gene in controls. (Original source: Lee CC, You NC, Song Y, Hsu YH, Manson J, Nathan L, et al. Relation of genetic variation in the gene coding for C-reactive protein with its plasma protein concentrations: findings from the Women's Health Initiative Observational Cohort. Clin Chem 2009;55:351-60)

			Minor allele frequency (%)					
dbSNP ID[a]	Location	Allele[b]	Whites	Blacks	Hispanics	Asian/ Pacific Islanders	Pooled	P[c]
			(n=968)	(n=732)	(n=303)	(n=195)	(n=2,198)	
rs4275453	5' flanking	T/C	37.1	55.7	30.8	26.9	42.0	<0.0001
rs2808634	5' flanking	C/T	29.7	17.9	21.3	14.0	23.3	<0.0001
rs3093059	5' flanking	T/C	6.97	23.7	9.1	12.8	13.6	<0.0001
rs2794521	5' flanking	T/C	30.1	17.6	20.9	12.9	23.2	<0.0001
rs1417938	Intron1	T/A	29.9	11.7	30.5	7.8	21.9	<0.0001
rs1800947	Exon2	G/C	6.00	1.75	3.3	8.3	4.3	<0.0001
rs1130864	Exon2	C/T	29.3	16.6	31.9	8.2	23.5	<0.0001
rs1205	Exon2	C/T	32.5	22.0	35.4	64.4	31.7	<0.0001
rs3093075	3' flanking	C/A	6.77	24.3	9.6	11.8	13.7	<0.0001
rs3093068	3' flanking	C/G	8.42	24.1	9.0	12.5	14.3	<0.0001
rs2808629	3' flanking	G/A	32.5	22.9	36.1	64.9	32.1	<0.0001
rs2369146	3' flanking	G/A	25.0	37.3	20.9	14.9	28.0	<0.0001
rs1470515	3' flanking	G/A	37.5	28.8	40.8	62.6	36.8	<0.0001

[a] From the NCBI dbSNP.

[b] Reference alleles are indicated in parentheses.

[c] P values were estimated by a χ^2 test (df=3) for genotype distribution across the 4 ethnic groups.

SupplementaryTable 2a. The multivariable-adjusted odds ratio (95% CI) for diabetes risk associated with genetic variants using the additive effect model among whites.

		Adjusted OR (95% CI)[a]				
Gene	**SNP ID**	**Model 1[c]**	**Model 2[d]**	**Model 3[e]**	**Model 4[f]**	**Model 5[g]**
CRP	rs4275453	0.88(0.76-1.01)	0.89(0.76-1.04)	0.88(0.72-1.07)	0.89(0.75-1.05)	0.87(0.71- 1.07)
(931/929)[b]	rs2808634	0.87(0.75-1.01)	0.88(0.74-1.04)	0.87(0.70-1.07)	0.84(0.70-1.00)	0.82(0.66 – 1.03)
	rs3093059	0.90(0.69-1.18)	0.96(0.72-1.28)	0.82(0.57-1.18)	0.96(0.71-1.30)	0.86(0.59 - 1.25)
	rs2794521	0.86(0.74-1.00)	0.88(0.74-1.04)	0.86(0.69-1.06)	0.85(0.70-1.02)	0.82(0.65 – 1.02)
	rs1417938	1.14(0.98-1.31)	1.15(0.98-1.34)	1.09(0.89-1.33)	1.10(0.93-1.30)	1.03(0.84 - 1.28)
	rs1800947	1.14(0.87-1.49)	1.18(0.88-1.59)	1.17(0.81-1.70)	1.23(0.90-1.68)	1.25(0.85 - 1.84)
	rs1130864	**1.18(1.02-1.36)**	**1.20(1.02-1.40)**	1.18(0.96-1.45)	1.13(0.95-1.34)	1.09(0.88 - 1.35)
	rs1205	1.00(0.87-1.16)	0.96(0.82-1.13)	1.03(0.84-1.26)	1.02(0.86-1.21)	1.10(0.89 - 1.37)
	rs3093075	0.96(0.73-1.26)	1.00(0.75-1.34)	0.85(0.59-1.24)	1.04(0.76-1.41)	0.90(0.61 - 1.32)
	rs3093068	0.83(0.66-1.06)	0.87(0.68-1.13)	0.76(0.56-1.04)	0.93(0.71-1.22)	0.81(0.59 - 1.13)
	rs2808629	1.02(0.88-1.17)	0.99(0.84-1.15)	1.05(0.86-1.28)	1.06(0.89-1.25)	1.15(0.93 - 1.43)
	rs2369146	0.88(0.75-1.03)	0.85(0.72-1.02)	0.84(0.67-1.05)	**0.81(0.66-0.98)**	0.79(0.62 – 1.00)
	rs1470515	1.04(0.91-1.19)	1.05(0.90-1.22)	1.12(0.93-1.36)	1.10(0.94-1.30)	1.21(1.99 - 1.48)
TNF-α	rs2239705	0.97(0.81-1.15)	0.96(0.79-1.16)	0.99(0.77-1.26)	0.99(0.81-1.21)	1.02(0.79-1.31)
(929/952)	rs2230365	1.08(0.90-1.30)	1.03(0.84-1.26)	1.16(0.89-1.51)	0.98(0.79-1.22)	1.06(0.81-1.40)
	rs2009658	0.96(0.80-1.16)	0.92(0.75-1.13)	0.99(0.76-1.30)	0.93(0.74-1.16)	0.96(0.73-1.27)
	rs2239704	1.02(0.90-1.17)	1.02(0.89-1.18)	1.04(0.87-1.24)	1.03(0.89-1.20)	1.09(0.90-1.31)
	rs909253	0.96(0.83-1.10)	0.99(0.85-1.15)	0.99(0.82-1.20)	1.00(0.85-1.18)	0.99(0.81-1.21)
	rs2857713	0.97(0.83-1.12)	0.93(0.79-1.09)	0.91(0.74-1.12)	0.89(0.75-1.07)	0.84(0.67-1.05)
	rs1041981	0.97(0.84-1.12)	1.00(0.86-1.17)	1.00(0.83-1.21)	1.02(0.87-1.20)	1.02(0.83-1.24)
	rs1799964	0.97(0.82-1.14)	0.92(0.77-1.11)	0.96(0.76-1.20)	0.90(0.74-1.10)	0.89(0.70-1.13)
	rs1799724	0.94(0.75-1.17)	0.97(0.76-1.23)	0.96(0.71-1.30)	1.04(0.80-1.35)	1.04(0.76-1.43)
	rs1800750	0.94(0.56-1.58)	1.00(0.57-1.73)	0.73(0.36-1.47)	0.81(0.43-1.52)	0.51(0.24-1.09)
	rs1800629	0.91(0.76-1.10)	0.88(0.71-1.07)	0.87(0.68-1.12)	0.89(0.72-1.11)	0.91(0.70-1.18)
	rs361525	1.00(0.76-1.32)	1.00(0.74-1.35)	0.89(0.61-1.29)	0.91(0.65-1.27)	0.74(0.49-1.11)
	rs3093661	1.06(0.77-1.46)	1.00(0.71-1.42)	0.92(0.60-1.42)	0.94(0.64-1.37)	0.79(0.50-1.26)
	rs3093662	1.06(0.84-1.33)	1.08(0.84-1.40)	0.99(0.72-1.37)	1.01(0.77-1.33)	0.87(0.62-1.23)
	rs769178	0.96(0.77-1.21)	0.99(0.77-1.27)	0.99(0.73-1.35)	1.07(0.82-1.40)	1.06(0.77-1.47)
	rs2256974	1.05(0.89-1.26)	1.12(0.93-1.36)	1.15(0.90-1.46)	1.12(0.91-1.37)	1.10(0.85-1.42)
IL-6	rs1880242	0.96(0.88-1.03)	0.93(0.85-1.02)	0.99(0.88-1.10)	0.95(0.87-1.04)	1.01 (0.90-1.13)

(946/960)	rs10499563	0.97(0.91-1.04)	0.97(0.90-1.04)	0.98(0.89-1.09)	0.96(0.89-1.04)	0.97 (0.88-1.08)
	rs2069824	**0.87(0.79-0.96)**	**0.88(0.79-0.98)**	0.89(0.77-1.02)	0.93(0.83-1.04)	0.93 (0.81-1.08)
	rs1800797	0.96(0.84-1.10)	1.01(0.87-1.17)	1.07(0.89-1.29)	1.03(0.88-1.21)	1.10 (0.90-1.34)
	rs1800796	1.05(0.79-1.41)	1.01(0.74-1.38)	1.14(0.77-1.68)	1.06(0.76-1.48)	1.18 (0.79-1.76)
	rs1800795	0.96(0.84-1.11)	1.00(0.86-1.16)	1.05(0.87-1.27)	1.01(0.86-1.18)	1.06 (0.87-1.29)
	rs2069830	0.96(0.06-15.49)	0.92(0.06-15.47)	1.40(0.08-25.7)	1.38(0.08-22.54)	1.81 (0.09-38.7)
	rs2069838	0.79(0.21-2.94)	0.68(0.18-2.63)	0.68(0.13-3.44)	0.69(0.17-2.88)	0.76 (0.15-3.88)
	rs2069840	0.96(0.83-1.11)	1.04(0.88-1.22)	1.09(0.89-1.34)	1.10(0.93-1.31)	1.16 (0.93-1.43)
	rs2069842	0.75(0.17-3.33)	0.72(0.15-3.42)	0.87(0.10-7.67)	1.09(0.21-5.56)	1.07 (0.11-10.4)
	rs2069845	1.02(0.89-1.17)	0.98(0.84-1.13)	0.92(0.76-1.11)	0.97(0.82-1.13)	0.91 (0.75-1.11)
	rs2069861	1.05(0.83-1.33)	1.02(0.78-1.32)	0.82(0.59-1.14)	1.15(0.87-1.53)	0.88 (0.62-1.25)
	rs1524106	0.87(0.75-1.00)	0.89(0.76-1.04)	0.87(0.72-1.06)	0.88(0.74-1.04)	0.86 (0.70-1.06)
	rs1524103	1.11(0.95-1.30)	1.10(0.93-1.30)	1.16(0.93-1.44)	1.10(0.92-1.32)	1.17 (0.93-1.47)

[a] Odds Ratio (OR) of each single SNP was calculated for additive genetic effect model and were estimated using conditional logistic regression models.

[b] Sample size for each ethnic group was shown for each gene in the format of (cases/controls).

[c] Model 1 was adjusted for matching factors (age, clinical center, and time of blood draw).

[d] Model 2 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy and physical activity.

[e] Model 3 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and BMI.

[f] Model 4 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and family history.

[g] Model 5 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity, BMI and family history.

Supplementary Table 2b. The multivariable-adjusted odds ratio (95% CI) for diabetes risk associated with genetic variants using the additive effect model among blacks.

		Adjusted OR (95% CI)[a]				
Gene	**SNP ID**	**Model 1[c]**	**Model 2[d]**	**Model 3[e]**	**Model 4[f]**	**Model 5[g]**
CRP (362/743)[b]	rs4275453	0.95(0.79-1.13)	0.97(0.80-1.17)	0.90(0.73-1.11)	0.95(0.76-1.18)	0.85(0.67 - 1.09)
	rs2808634	**0.72(0.56-0.93)***	**0.74(0.56-0.97)**	**0.68(0.51-0.91)***	0.82(0.59-1.14)	0.74(0.51 - 1.07)
	rs3093059	1.08(0.88-1.33)	1.01(0.81-1.26)	1.08(0.85-1.38)	1.09(0.84-1.41)	1.17(0.88 - 1.56)
	rs2794521	**0.74(0.57-0.96)**	0.76(0.57-1.00)	**0.72(0.53-0.98)**	0.83(0.59-1.17)	0.78(0.54 - 1.14)
	rs1417938	1.16(0.88-1.52)	1.22(0.91-1.64)	1.38(1.00-1.90)	1.19(0.84-1.70)	1.36(0.93 – 2.00)
	rs1800947	0.47(0.19-1.18)	0.42(0.15-1.17)	0.49(0.17-1.43)	0.46(0.14-1.45)	0.52(0.15 - 1.81)
	rs1130864	1.20(0.95-1.51)	1.27(0.99-1.63)	**1.41(1.07-1.86)***	1.08(0.80-1.46)	1.18(0.86 - 1.63)
	rs1205	**0.78(0.62-0.99)**	**0.77(0.60-0.98)**	0.77(0.59-1.00)	0.78(0.58-1.04)	0.80(0.59 - 1.10)
	rs3093075	1.08(0.87-1.32)	1.00(0.80-1.25)	1.06(0.83-1.35)	1.09(0.84-1.41)	1.17(0.88 - 1.55)
	rs3093068	**1.44(1.18-1.77)****	**1.46(1.16-1.82)***	**1.42(1.12-1.81)***	**1.37(1.05-1.78)**	1.30(0.97 - 1.73)
	rs2808629	**0.74(0.59-0.93)***	**0.71(0.56-0.91)***	**0.71(0.54-0.92)***	**0.72(0.54-0.95)**	**0.73(0.54 – 0.99)**
	rs2369146	0.89(0.74-1.08)	0.86(0.70-1.06)	0.86(0.69-1.07)	0.95(0.75-1.20)	0.96(0.74 - 1.24)
	rs1470515	**0.77(0.62-0.96)**	**0.75(0.59-0.95)**	**0.73(0.57-0.95)***	**0.72(0.54-0.95)**	**0.70(0.52 - 0.94)**
TNF-α (364/746)	rs2239705	0.86(0.64-1.15)	0.80(0.58-1.09)	0.76(0.54-1.06)	0.80(0.56-1.15)	0.74(0.50-1.09)
	rs2230365	0.89(0.60-1.31)	0.93(0.62-1.40)	0.93(0.60-1.45)	0.96(0.59-1.56)	0.91(0.53-1.54)
	rs2009658	0.92(0.70-1.23)	0.96(0.71-1.30)	0.93(0.67-1.30)	1.00(0.70-1.44)	0.90(0.60-1.33)
	rs2239704	0.88(0.71-1.09)	0.81(0.64-1.02)	0.84(0.65-1.08)	**0.76(0.58-1.00)**	0.79(0.59-1.07)
	rs909253	1.19(0.98-1.45)	**1.28(1.03-1.58)**	**1.33(1.06-1.68)**	1.28(1.00-1.64)	1.40(1.06-1.85)
	rs2857713	0.91(0.73-1.13)	0.92(0.73-1.16)	0.86(0.67-1.12)	1.04(0.79-1.36)	0.94(0.70-1.27)
	rs1041981	1.16(0.95-1.40)	1.22(0.99-1.50)	**1.28(1.02-1.61)**	1.26(0.99-1.61)	1.39(1.06-1.83)
	rs1799964	0.90(0.71-1.15)	0.93(0.72-1.21)	0.87(0.66-1.16)	0.97(0.72-1.32)	0.86(0.61-1.20)
	rs1799724	0.72(0.43-1.23)	0.78(0.44-1.36)	0.65(0.34-1.21)	0.76(0.40-1.46)	0.62(0.30-1.28)
	rs1800750	1.21(0.67-2.17)	1.06(0.57-1.99)	0.96(0.48-1.92)	1.23(0.59-2.59)	1.22(0.55-2.69)
	rs1800629	1.08(0.82-1.41)	1.09(0.81-1.46)	1.10(0.81-1.51)	1.19(0.85-1.67)	1.24(0.86-1.78)
	rs361525	1.01(0.66-1.56)	1.06(0.67-1.67)	1.01(0.61-1.68)	1.09(0.63-1.86)	1.04(0.57-1.87)

	rs3093661	0.82(0.50-1.36)	0.80(0.47-1.37)	0.70(0.39-1.25)	0.76(0.40-1.42)	0.66(0.33-1.32)
	rs3093662	0.80(0.56-1.14)	0.78(0.53-1.13)	**0.65(0.43-0.99)**	0.85(0.56-1.30)	0.70(0.43-1.12)
	rs769178	0.64(0.37-1.10)	0.69(0.39-1.22)	0.60(0.32-1.13)	0.67(0.34-1.31)	0.58(0.28-1.20)
	rs2256974	1.17(0.96-1.43)	**1.26(1.01-1.56)**	**1.34(1.05-1.70)**	1.15(0.89-1.47)	1.26(0.96-1.67)
IL-6	rs1880242	1.12(0.99-1.27)	1.12(0.98-1.28)	1.11(0.96-1.28)	1.07(0.93-1.23)	1.04 (0.89-1.22)
(365/753)	rs10499563	0.96(0.85-1.08)	0.95(0.84-1.08)	0.95(0.84-1.08)	0.98(0.85-1.12)	0.96 (0.83-1.11)
	rs2069824	1.02(0.88-1.18)	1.01(0.86-1.17)	1.01(0.86-1.19)	1.04(0.88-1.24)	1.06 (0.89-1.27)
	rs1800797	0.90(0.68-1.20)	0.86(0.63-1.17)	0.83(0.59-1.16)	0.92(0.64-1.32)	0.84 (0.57-1.26)
	rs1800796	1.25(0.91-1.74)	1.38(0.97-1.97)	1.30(0.90-1.89)	1.42(0.95-2.13)	1.34 (0.88-2.04)
	rs1800795	1.02(0.76-1.38)	0.98(0.71-1.36)	0.95(0.67-1.35)	1.04(0.70-1.52)	0.93 (0.61-1.41)
	rs2069830	0.94(0.68-1.30)	0.97(0.68-1.39)	0.98(0.67-1.44)	1.04(0.69-1.58)	0.99 (0.62-1.56)
	rs2069838	1.06(0.73-1.52)	1.11(0.76-1.63)	1.13(0.75-1.72)	1.21(0.78-1.89)	1.22 (0.76-1.97)
	rs2069840	0.89(0.70-1.13)	0.88(0.68-1.13)	0.87(0.66-1.14)	0.86(0.64-1.17)	0.89 (0.65-1.23)
	rs2069842	1.15(0.81-1.64)	1.28(0.88-1.88)	1.26(0.84-1.90)	1.53(0.96-2.44)	1.43 (0.87-2.37)
	rs2069845	1.03(0.85-1.25)	1.12(0.91-1.38)	1.08(0.86-1.36)	1.12(0.87-1.43)	1.13 (0.87-1.48)
	rs2069861	0.97(0.50-1.89)	1.03(0.52-2.07)	1.08(0.52-2.28)	0.98(0.43-2.23)	1.12 (0.43-2.91)
	rs1524106	1.03(0.86-1.24)	1.09(0.90-1.33)	1.05(0.86-1.30)	1.09(0.87-1.38)	1.13 (0.88-1.44)
	rs1524103	1.02(0.85-1.22)	0.96(0.78-1.17)	0.99(0.80-1.22)	0.98(0.78-1.23)	0.94 (0.73-1.20)

[a] Odds Ratio (OR) of each single SNP was calculated for additive genetic effect model and were estimated using conditional logistic regression models.

[b] Sample size for each ethnic group was shown for each gene in the format of (cases/controls).

[c] Model 1 was adjusted for matching factors (age, clinical center, and time of blood draw).

[d] Model 2 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy and physical activity.

[e] Model 3 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and BMI.

[f] Model 4 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and family history.

[g] Model 5 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity, BMI and family history.

* Adjusted q-value<0.05 after FDR

** Adjusted q-value<0.01 after FDR

Supplementary Table 2c. The multivariable-adjusted odds ratio (95% CI) for diabetes risk associated with genetic variants using the additive effect model among Hispanics.

		Adjusted OR (95% CI)[a]				
Gene	**SNP ID**	**Model 1[c]**	**Model 2[d]**	**Model 3[e]**	**Model 4[f]**	**Model 5[g]**
CRP	rs4275453	0.95(0.66-1.35)	0.98(0.66-1.44)	0.91(0.60-1.38)	0.96(0.60-1.51)	0.93(0.57 - 1.52)
(139/278)[b]	rs2808634	0.80(0.52-1.22)	0.83(0.53-1.32)	0.81(0.50-1.34)	0.84(0.49-1.42)	0.84(0.48 - 1.49)
	rs3093059	0.79(0.46-1.36)	0.64(0.35-1.18)	0.63(0.33-1.20)	0.66(0.34-1.30)	0.62(0.30 - 1.25)
	rs2794521	0.76(0.49-1.18)	0.81(0.51-1.30)	0.75(0.45-1.25)	0.79(0.45-1.37)	0.76(0.41 - 1.38)
	rs1417938	1.04(0.74-1.46)	1.15(0.79-1.68)	1.18(0.77-1.80)	1.06(0.69-1.62)	1.07(0.65 - 1.76)
	rs1800947	0.90(0.34-2.38)	1.09(0.38-3.08)	0.87(0.29-2.58)	1.62(0.49-5.39)	1.46(0.40 - 5.31)
	rs1130864	0.94(0.67-1.33)	1.01(0.70-1.48)	0.97(0.64-1.49)	0.97(0.64-1.48)	0.91(0.56 - 1.48)
	rs1205	1.02(0.74-1.40)	0.97(0.68-1.38)	1.02(0.70-1.50)	1.09(0.71-1.66)	1.19(0.74 - 1.92)
	rs3093075	0.80(0.46-1.37)	0.69(0.38-1.27)	0.70(0.37-1.32)	0.62(0.32-1.21)	0.57(0.28- 1.16)
	rs3093068	0.96(0.56-1.65)	0.95(0.53-1.69)	0.86(0.47-1.59)	0.95(0.50-1.82)	0.84(0.42 - 1.66)
	rs2808629	0.99(0.73-1.36)	0.95(0.67-1.34)	1.07(0.73-1.57)	1.08(0.70-1.65)	1.30(0.81 - 2.10)
	rs2369146	1.19(0.83-1.71)	1.17(0.79-1.72)	1.19(0.79-1.79)	1.22(0.77-1.93)	1.23(0.76 – 2.00)
	rs1470515	0.92(0.68-1.25)	0.85(0.61-1.19)	0.87(0.61-1.25)	0.92(0.61-1.39)	0.97(0.6 2- 1.53)
TNF-α	rs2239705	0.90(0.65-1.26)	0.98(0.68-1.43)	0.94(0.62-1.40)	1.00(0.65-1.54)	0.96(0.61-1.53)
(140/277)	rs2230365	0.76(0.50-1.17)	0.75(0.46-1.23)	0.82(0.48-1.40)	0.72(0.41-1.25)	0.77(0.42-1.44)
	rs2009658	0.84(0.56-1.27)	0.90(0.58-1.42)	0.90(0.55-1.46)	0.77(0.46-1.32)	0.75(0.42-1.33)
	rs2239704	1.14(0.86-1.50)	1.23(0.91-1.66)	1.13(0.83-1.56)	1.38(0.98-1.95)	1.33(0.92-1.92)
	rs909253	0.94(0.69-1.27)	0.79(0.56-1.12)	0.89(0.61-1.28)	0.75(0.50-1.14)	0.83(0.54-1.28)
	rs2857713	0.91(0.64-1.28)	1.01(0.69-1.48)	0.96(0.64-1.44)	0.93(0.60-1.46)	0.90(0.55-1.46)
	rs1041981	0.95(0.71-1.29)	0.80(0.57-1.12)	0.89(0.62-1.28)	0.76(0.51-1.14)	0.84(0.55-1.28)
	rs1799964	1.03(0.70-1.51)	1.14(0.75-1.74)	1.08(0.69-1.69)	1.07(0.66-1.74)	1.01(0.60-1.72)
	rs1799724	1.02(0.69-1.50)	1.13(0.74-1.73)	1.02(0.65-1.60)	1.19(0.73-1.95)	1.14(0.67-1.94)
	rs1800750	2.08(0.67-6.47)	2.13(0.66-6.91)	1.41(0.38-5.22)	3.42(0.88-13.37)	2.58(0.55-12.2)
	rs1800629	1.10(0.66-1.84)	1.01(0.57-1.79)	1.17(0.63-2.18)	1.09(0.56-2.11)	1.19(0.56-2.52)
	rs361525	1.78(0.84-3.79)	2.06(0.88-4.82)	1.67(0.67-4.19)	2.56(0.95-6.93)	2.23(0.77-6.46)
	rs3093661	1.35(0.54-3.33)	1.53(0.54-4.34)	1.36(0.44-4.23)	1.52(0.46-5.06)	1.56(0.44-5.50)
	rs3093662	1.16(0.63-2.14)	1.21(0.61-2.40)	1.03(0.50-2.12)	1.22(0.56-2.65)	1.09(0.47-2.52)
	rs769178	0.96(0.65-1.41)	1.07(0.70-1.66)	0.97(0.61-1.54)	1.16(0.70-1.91)	1.09(0.64-1.86)
	rs2256974	0.93(0.65-1.32)	0.76(0.51-1.12)	0.77(0.51-1.19)	0.75(0.48-1.17)	0.80(0.49-1.30)
IL-6	rs1880242	1.19(0.94-1.52)	1.09(0.85-1.40)	1.05(0.78-1.41)	1.12(0.79-1.58)	1.14 (0.74-1.76)

(140/279)	rs10499563	1.01(0.85-1.21)	1.10(0.90-1.35)	1.06(0.83-1.35)	1.01(0.81-1.25)	0.97 (0.76-1.23)
	rs2069824	0.98(0.75-1.26)	0.95(0.72-1.25)	0.89(0.66-1.21)	0.77(0.54-1.11)	**0.66 (0.45-0.98)**
	rs1800797	1.22(0.83-1.81)	1.20(0.77-1.86)	1.10(0.68-1.79)	1.14(0.66-1.96)	1.12 (0.62-2.04)
	rs1800796	0.98(0.67-1.42)	0.94(0.62-1.43)	0.99(0.63-1.55)	1.04(0.64-1.70)	1.02 (0.60-1.73)
	rs1800795	1.25(0.84-1.87)	1.23(0.78-1.92)	1.19(0.73-1.96)	1.26(0.74-2.13)	1.35 (0.76-2.41)
	rs2069830	4.61(0.46-45.87)	1.06(0.06-17.81)	2.11(0.10-46.40)	---[h]	---[h]
	rs2069838	---[h]	---[h]	---[h]	---[h]	---[h]
	rs2069840	1.08(0.78-1.49)	1.04(0.73-1.48)	1.12(0.77-1.62)	0.88(0.58-1.32)	0.94 (0.61-1.44)
	rs2069842	0.56(0.16-1.98)	0.98(0.23-4.22)	0.93(0.20-4.34)	0.94(0.15-5.98)	1.07 (0.14-8.55)
	rs2069845	0.80(0.57-1.12)	0.77(0.53-1.13)	0.77(0.52-1.16)	0.87(0.56-1.35)	0.82 (0.51-1.32)
	rs2069861	1.22(0.58-2.58)	1.21(0.53-2.75)	1.22(0.52-2.87)	1.15(0.45-2.94)	1.16 (0.43-3.16)
	rs1524106	1.00(0.73-1.36)	0.98(0.69-1.37)	1.01(0.71-1.45)	0.92(0.63-1.34)	0.87 (0.58-1.31)
	rs1524103	0.95(0.67-1.34)	0.95(0.66-1.37)	0.97(0.66-1.43)	1.07(0.70-1.65)	1.16 (0.73-1.85)

[a] Odds Ratio (OR) of each single SNP was calculated for additive genetic effect model and were estimated using conditional logistic regression models.

[b] Sample size for each ethnic group was shown for each gene in the format of (cases/controls).

[c] Model 1 was adjusted for matching factors (age, clinical center, and time of blood draw).

[d] Model 2 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy and physical activity.

[e] Model 3 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and BMI.

[f] Model 4 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and family history.

[g] Model 5 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity, BMI and family history.

[h] Result is difficult to interpret because of small sample size within strata.

Supplementary 2d. The multivariable-adjusted odds ratio (95% CI) for diabetes risk associated with genetic variants using the additive effect model among Asians/Pacific Islanders.

		Adjusted OR (95% CI)[a]				
Gene	**SNP ID**	**Model 1[c]**	**Model 2[d]**	**Model 3[e]**	**Model 4[f]**	**Model 5[g]**
CRP	rs4275453	0.88(0.53-1.47)	0.92(0.55-1.55)	0.96(0.55-1.70)	1.01(0.54-1.90)	1.04(0.51 - 2.15)
(77/163)[b]	rs2808634	1.05(0.53-2.07)	1.06(0.53-2.13)	1.09(0.49-2.40)	1.74(0.72-4.19)	1.93(0.73 - 5.11)
	rs3093059	0.90(0.48-1.68)	0.94(0.50-1.78)	0.99(0.49-1.97)	0.82(0.38-1.74)	0.76(0.32 - 1.80)
	rs2794521	1.05(0.52-2.12)	1.02(0.49-2.11)	1.19(0.53-2.68)	1.32(0.55-3.20)	1.66(0.66 - 4.19)
	rs1417938	0.84(0.36-1.96)	0.90(0.38-2.15)	0.79(0.29-2.16)	0.86(0.33-2.27)	0.75(0.23 - 2.47)
	rs1800947	0.77(0.36-1.68)	0.74(0.34-1.63)	0.61(0.26-1.43)	0.94(0.34-2.60)	1.00(0.33 – 2.96)
	rs1130864	0.96(0.44-2.08)	1.04(0.47-2.32)	1.02(0.41-2.53)	1.12(0.46-2.75)	1.15(0.40 - 3.31)
	rs1205	0.93(0.58-1.47)	0.87(0.54-1.40)	0.88(0.52-1.49)	0.83(0.47-1.48)	0.82(0.43 - 1.56)
	rs3093075	1.13(0.63-2.05)	1.15(0.63-2.12)	1.24(0.63-2.43)	1.10(0.53-2.27)	1.09(0.47 - 2.51)
	rs3093068	1.16(0.64-2.10)	1.18(0.64-2.17)	1.30(0.65-2.60)	1.02(0.49-2.13)	1.11(0.47 - 2.61)
	rs2808629	0.92(0.59-1.43)	0.88(0.56-1.38)	0.89(0.54-1.47)	0.78(0.45-1.34)	0.72(0.39 - 1.35)
	rs2369146	1.09(0.56-2.15)	1.10(0.55-2.18)	1.17(0.54-2.56)	1.86(0.75-4.59)	2.19(0.80 – 5.97)
	rs1470515	0.99(0.64-1.55)	0.96(0.61-1.53)	1.00(0.60-1.67)	0.95(0.54-1.67)	0.95(0.50 - 1.83)
TNF-α	rs2239705	0.60(0.33-1.09)	0.60(0.33-1.10)	0.61(0.30-1.23)	0.65(0.32-1.35)	0.75(0.31-1.79)
(79/165)	rs2230365	1.09(0.72-1.66)	1.04(0.68-1.60)	1.02(0.63-1.64)	0.98(0.59-1.63)	0.86(0.47-1.55)
	rs2009658	1.12(0.65-1.93)	1.02(0.59-1.79)	0.87(0.47-1.62)	1.11(0.58-2.11)	0.92(0.44-1.94)
	rs2239704	0.77(0.51-1.16)	0.79(0.52-1.20)	0.73(0.45-1.19)	0.76(0.46-1.26)	0.66(0.35-1.23)
	rs909253	1.36(0.91-2.04)	1.42(0.94-2.15)	**1.71(1.04-2.79)**	1.29(0.77-2.15)	1.77(0.92-3.38)
	rs2857713	1.00(0.61-1.65)	0.92(0.55-1.54)	0.81(0.45-1.45)	1.13(0.64-1.99)	0.97(0.50-1.91)
	rs1041981	1.37(0.92-2.06)	1.43(0.94-2.16)	**1.70(1.04-2.77)**	1.30(0.78-2.19)	1.79(0.93-3.43)
	rs1799964	1.14(0.68-1.91)	1.03(0.60-1.76)	0.87(0.47-1.62)	1.21(0.67-2.17)	1.01(0.50-2.03)
	rs1799724	0.67(0.37-1.20)	0.65(0.35-1.19)	0.70(0.34-1.43)	0.67(0.32-1.41)	0.80(0.34-1.92)
	rs1800750	---[h]	---[h]	---[h]	---[h]	---[h]
	rs1800629	**5.58(1.55-20.1)**	**6.28(1.73-22.80)**	**6.12(1.54-24.36)**	**6.50(1.40-30.28)**	6.29(1.14-34.8)[h]
	rs361525	1.17(0.42-3.25)	1.09(0.38-3.07)	1.16(0.36-3.74)	1.55(0.47-5.14)	1.56(0.39-6.17)
	rs3093661	0.69(0.21-2.22)	0.67(0.20-2.24)	0.54(0.14-2.12)	0.88(0.21-3.64)	0.48(0.08-2.92)
	rs3093662	1.07(0.39-2.90)	1.00(0.36-2.75)	0.95(0.31-2.90)	1.46(0.46-4.68)	1.24(0.33-4.63)
	rs769178	0.73(0.40-1.32)	0.73(0.40-1.35)	0.82(0.40-1.66)	0.74(0.35-1.57)	0.94(0.39-2.27)

	rs2256974	1.03(0.67-1.60)	1.03(0.66-1.60)	1.30(0.76-2.22)	1.02(0.59-1.75)	1.49(0.76-2.93)
IL-6	rs1880242	1.25(0.95-1.66)	1.27(0.96-1.68)	1.21(0.88-1.67)	1.26(0.92-1.71)	1.14 (0.81-1.60)
(77/165)	rs10499563	0.96(0.78-1.18)	0.97(0.79-1.19)	0.93(0.74-1.17)	0.87(0.66-1.16)	0.80 (0.58-1.11)
	rs2069824	---[h]	---[h]	---[h]	---[h]	---[h]
	rs1800797	0.92(0.36-2.38)	0.79(0.30-2.08)	1.00(0.33-3.04)	0.46(0.11-1.94)	0.66 (0.09-4.77)
	rs1800796	1.18(0.77-1.81)	1.19(0.77-1.82)	1.04(0.65-1.65)	1.11(0.69-1.81)	0.90 (0.52-1.55)
	rs1800795	0.84(0.20-3.58)	0.79(0.18-3.40)	1.45(0.27-7.77)	0.17(0.01-1.99)	0.31 (0.01-6.92)
	rs2069830	---[h]	---[h]	---[h]	---[h]	---[h]
	rs2069838	---[h]	---[h]	---[h]	---[h]	---[h]
	rs2069840	1.08(0.51-2.29)	1.07(0.50-2.28)	0.80(0.35-1.82)	1.03(0.44-2.42)	0.64 (0.24-1.69)
	rs2069842	---[h]	---[h]	---[h]	---[h]	---[h]
	rs2069845	---[h]	---[h]	---[h]	---[h]	---[h]
	rs2069861	---[h]	---[h]	---[h]	---[h]	---[h]
	rs1524106	1.23(0.63-2.41)	1.24(0.62-2.46)	1.00(0.47-2.13)	1.51(0.68-3.36)	0.97 (0.40-2.34)
	rs1524103	0.91(0.54-1.52)	0.92(0.55-1.55)	0.86(0.49-1.52)	0.86(0.48-1.55)	0.75 (0.38-1.47)

[a] Odds Ratio (OR) of each single SNP was calculated for additive genetic effect model and were estimated using conditional logistic regression models.

[b] Sample size for each ethnic group was shown for each gene in the format of (cases/controls).

[c] Model 1 was adjusted for matching factors (age, clinical center, and time of blood draw).

[d] Model 2 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy and physical activity.

[e] Model 3 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and BMI.

[f] Model 4 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity and family history.

[g] Model 5 was adjusted for matching factors , cigarette smoking, alcohol intake, hormone replacement therapy , physical activity, BMI and family history.

[h] Result is difficult to interpret because of small sample size within strata.

2B.6 References

1. Szanto A, Nagy L. The many faces of PPARgamma: anti-inflammatory by any means? Immunobiology 2008;213:789-803.

2. Ahmed W, Ziouzenkova O, Brown J, et al. PPARs and their metabolic modulation: new mechanisms for transcriptional regulation? J Intern Med 2007;262:184-98.

3. Blackburn GL. From bench to bedside: novel mechanisms and therapeutic advances through the development of selective peroxisome proliferator-activated receptor gamma modulators. Am J Clin Nutr;91:251S-3S.

4. Cheatham WW. Peroxisome proliferator-activated receptor translational research and clinical experience. Am J Clin Nutr;91:262S-6S.

5. Ryan KK, Li B, Grayson BE, Matter EK, Woods SC, Seeley RJ. A role for central nervous system PPAR-gamma in the regulation of energy balance. Nat Med;17:623-6.

6. Kahn BB, McGraw TE. Rosiglitazone, PPARgamma, and type 2 diabetes. N Engl J Med 2010;363:2667-9.

7. Gaulton KJ, Willer CJ, Li Y, et al. Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes. Diabetes 2008;57:3136-44.

8. Gouda HN, Sagoo GS, Harding AH, Yates J, Sandhu MS, Higgins JP. The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. Am J Epidemiol;171:645-55.

9. Herder C, Rathmann W, Strassburger K, et al. Variants of the PPARG, IGF2BP2, CDKAL1, HHEX, and TCF7L2 genes confer risk of type 2 diabetes independently of BMI in the German KORA studies. Horm Metab Res 2008;40:722-6.

10. Hu C, Zhang R, Wang C, et al. PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. PLoS One 2009;4:e7643.

11. Lin Y, Li P, Cai L, et al. Association study of genetic variants in eight genes/loci with type 2 diabetes in a Han Chinese population. BMC Med Genet;11:97.

12. Lyssenko V, Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 2008;359:2220-32.

13. Meigs JB, Manning AK, Fox CS, et al. Genome-wide association with diabetes-related traits in the Framingham Heart Study. BMC Med Genet 2007;8 Suppl 1:S16.

14. Nemoto M, Sasaki T, Deeb SS, Fujimoto WY, Tajima N. Differential effect of PPARgamma2 variants in the development of type 2 diabetes between native Japanese and Japanese Americans. Diabetes Res Clin Pract 2002;57:131-7.

15. Sanghera DK, Ortega L, Han S, et al. 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 2008;9:59.

16. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-5.

17. Webster RJ, Warrington NM, Beilby JP, Frayling TM, Palmer LJ. The longitudinal association of common susceptibility variants for type 2 diabetes and obesity with fasting glucose level and BMI. BMC Med Genet;11:140.

18. Weedon MN, McCarthy MI, Hitman G, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3:e374.

19. Wen J, Ronn T, Olsson A, et al. Investigation of type 2 diabetes risk alleles support CDKN2A/B, CDKAL1, and TCF7L2 as susceptibility genes in a Han Chinese cohort. PLoS One;5:e9153.

20. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 1998;19:61-109.

21. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol 2003;13:S5-17.

22. Liu S, Tinker L, Song Y, et al. A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. Arch Intern Med 2007;167:1676-85.

23. Hao K, Liu S, Niu T. A sparse marker extension tree algorithm for selecting the best set of haplotype tagging single nucleotide polymorphisms. Genet Epidemiol 2005;29:336-52.

24. Hsu YH, Niu T, Song Y, Tinker L, Kuller LH, Liu S. Genetic variants in the UCP2-UCP3 gene cluster and risk of diabetes in the Women's Health Initiative Observational Study. Diabetes 2008;57:1101-7.

25. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.

26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57:289-300.

27. Hansen SK, Nielsen EM, Ek J, et al. Analysis of separate and combined effects of common variation in KCNJ11 and PPARG on risk of type 2 diabetes. J Clin Endocrinol Metab 2005;90:3629-37.

28. He Y, Li C, Amos CI, Xiong M, Ling H, Jin L. Accelerating haplotype-based genome-wide association study using perfect phylogeny and phase-known reference data. PLoS One;6:e22097.

29. Tzeng JY, Lu W, Farmen MW, Liu Y, Sullivan PF. Haplotype-based pharmacogenetic analysis for longitudinal quantitative traits in the presence of dropout. J Biopharm Stat;20:334-50.

30. Heude B, Pelloux V, Forhan A, et al. Association of the Pro12Ala and C1431T variants of PPARgamma and their haplotypes with susceptibility to gestational diabetes. J Clin Endocrinol Metab 2011;96:E1656-60.

31. Lorenz AJ, Hamblin MT, Jannink JL. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS One 2010;5:e14079.

32. Feng T, Zhu X. Genome-wide searching of rare genetic variants in WTCCC data. Hum Genet;128:269-80.

33. Parikh H, Lyssenko V, Groop LC. Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. BMC Med Genomics 2009;2:72.

34. Perry JR, McCarthy MI, Hattersley AT, et al. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. Diabetes 2009;58:1463-7.

35. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet;86:581-91.

36. Small KS, Hedman AK, Grundberg E, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet;43:561-4.

37. Luan J, Browne PO, Harding AH, et al. Evidence for gene-nutrient interaction at the PPARgamma locus. Diabetes 2001;50:686-9.

38. Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet;43:513-8.

Chapter 2C:

Common Genetic Variants in Peroxisome Proliferator-activated Receptor γ (*PPARG*) and Clinical Diabetes Risk among Women's Health Initiative Postmenopausal Women

2C.1 Introduction

The peroxisome proliferator-activated receptor γ (PPARG) is a ligand-activated transcription factor that plays an essential role in the regulation of lipid uptake, adipocyte differentiation, and energy balance[1]. It further acts as an anti-inflammatory molecular by hindering inflammatory reactions that are critical in the pathogenesis of type 2 diabetes (T2D)[2]. Thus, PPARG agonist is also the target of the thiazolidinedione (TZD) class of insulin-sensitizing drugs for glycemic control[3-5].

PPARG activates the expression of genes involved in glucose and lipid metabolism, which converts nutritional signals into metabolic consequences[5]. PPARG is also a nuclear receptor and transcription factor that controls the expression of many genes and plays a vital role in adipocyte differentiation[6]. The *PPARG Pro12Ala* (rs1801282) has been the most vastly investigated single nucleotide polymorphism (SNP), which was believed to alter transcriptional activity as a result of its location in the functional binding domain that has been associated with risk of diabetes and its intermediate traits[7-19].

Therefore, we aimed to comprehensively assess all common variants in the *PPARG* gene in relation to T2D risk in two independent studies in the Women's Health Initiative. We examined these genetic associations in a nested case-control study of postmenopausal women participated in the Women's Health Initiative Observation Study (WHI-OS) and the WHI SNP Health Association Resource (SHARe) cohort respectively.

2C.2 Research Design and Methods

2C.2.1 Study Participants

Details regarding the design of our case-control study nested in the WHI-OS have been described in the Section 2.2.1 in Chapter 2 (P.12-13)[20-22].

2C.2.2 Haplotype-tagging single nucleotide polymorphism selection and genotyping methods

Details regarding the SNP selection and genotyping methods are provided in the Section 2.2.3 in Chapter 2A (P.13-15) [23,24].

2C.2.3 Statistical Analysis

We first estimated the minor allele frequency (MAF) of the 24 tagSNPs among the controls in each ethnic group. We tested for heterogeneity of genotype distributions across ethnicities using the χ^2 test.

In both single-SNP and haplotype-based analyses, we employed multivariable logistic regression to calculate odds ratios (ORs) and 95% confidence intervals (95% CIs) for each genetic variant with diabetes risk. We made adjustments for matching factors (age, clinical center, time of blood draw, and ethnicity) and other covariates [body mass index, fasting glucose and insulin levels in logarithmic scale, cigarette smoking (never, past, and current), alcohol intake (never, past, and current), hormone-replacement therapy use (never, past, and current), family history of diabetes, and values of the total metabolic equivalent (MET) value from the individual's recreational physical activity per week at baseline.

In single-SNP analyses, each SNP was coded as an additive, dominant, or recessive genetic model. In haplotype-based analyses, only haplotypes with estimated frequencies ≥ 5% in the combined cases and controls were included for analyses. To increase the genomic coverage,

we utilized a sliding-window (window width = 3 SNPs) haplotype-based analysis. For each window, we used an omnibus likelihood ratio test (LRT), which was a χ^2 test (degrees of freedom = number of haplotypes in a particular window – 1). The test used a measure derived based on the difference of the logarithmic likelihood of two logistic regression models: (1) the full model that contains the haplotype covariates, and (ii) the reduced model that does not contain the haplotype covariates. A $-\log_{10}P > 2.64$ (P value < 0.0023) was used as the global significance threshold using Bonferroni correction for 22 window frames. By selecting those haplotypes that were significantly associated with diabetes risk in the combined population, we then evaluated the association between the resulting haplotype/haplotype combinations and diabetes risk.

Additional single-SNP analyses were conducted to validate results from the WHI-OS in a larger cohort, the WHI-SHARe. The WHI-SHARe cohort included 8,421 African American and 3,587 Hispanic American postmenopausal women with raw genotyping data available (909, 622 genotypes were produced by the Affymetrix Genome-wide Human SNP Array 6.0, Santa Clara, CA). Eight out of 24 tagSNPs were included in the WHI-SHARe data. After removing 234 related individuals and 56 individuals with discordant race, we used multivariate logistic regression to calculate the estimates of OR and 95% CI per each additional copy of the reference allele of each SNP under additive genetic model (R, version 2.13), with adjustment for the set of covariates listed above as well as global ancestry using 3 principal components (PCs) computed with EIGENSTRAT[25]. We estimated ethnic interaction by fitting a model with race*SNP interaction term. The likelihood ratio test was used to compare model with SNP versus model without SNP. We also performed a statistical analysis of pathways using the Gene Set Enrichment Algorithm (GSEA) with the 871, 309 SNPs remained after genotype cleaning

provided by the GenGen suite. More details will be provided in Section 3.2.3 in Chapter 3 (P.108 – P.109).

To account for potential false positives due to multiple comparisons, we calculated the false discovery rate (FDR) statistics with q values by incorporating all P values from multiple tests performed for the association of SNPs and diabetes risk using the method of Benjamini and Hochberg[26]. The FDR statistics with $q < 0.05$ were considered significant.

2C.3 Results

2C.3.1 Estimation of allele frequencies

24 tagSNPs were genotyped in 1,543 diabetes cases and 2,170 matched controls. **Figure 1** showed the characteristics of the 24 tagSNPs. A total of 19 SNPs were located within the *PPARG* gene (147kb) with two SNPs (rs1801282 [*Pro12Ala*] and rs3856806 [*His477His*]) resided in exon 4 and 12 respectively. The estimated MAFs in controls stratified by ethnicity were shown in **Table 1**. Except for rs10510411, rs12629293, and rs12636454 (with P values of 0.10, 0.06, and 0.32), the genotype distributions of all other tagSNPs varied significantly across the four ethnic groups (P values ≤ 0.0001).

2C.3.2 Single-SNP analyses

The association between each tagSNP with diabetes risk in each ethnic group, as well as in the combined population, was assessed under additive, dominant, and recessive genetic models. The results under the additive genetic model were presented in Table 2. After adjusting for matching factors and risk factors for T2D, the *Pro12Ala* (rs1801282) SNP was associated with T2D risk among Hispanic women (nominal P value = 0.04). Several other SNPs

(rs6809631, rs9817428, rs10510411, rs12629293, and rs12636454) were also found to have association with T2D risk among the combined group (all nominal P values < 0.05). After adjusting for multiple comparisons, no significant associations between any of the SNPs and diabetes risk were obtained (all q values > 0.05). The results under the dominant and recessive genetic model are shown in **Supplementary Table 1** and **2**. As shown in **Table 3**, we further investigated the association between the non-synonymous SNPs (rs1801282) by genotype with T2D risk. Among Hispanic, individuals with CG genotype appeared to be protective to T2D risk (OR= 0.17, 95% CI= 0.04-0.77, P value=0.02) compared to individuals having CC genotype at the rs1801282 SNP.

2C.3.3 Haplotype-based analyses

Figure 2 shows the results of the sliding-window (with width = 3) haplotype-based analyses. There were a total of 22 window frames for the 24 tagSNPs. Using the omnibus LRT for testing each window frame in the combined population, rs1175540 (SNP20), rs1175544 (SNP21), and rs1797912 (SNP22) gave rise to the most significant P value [P value for LRT test = 5.9×10^{-4}, $-\log_{10}$(P value for LRT test)=2.97]. An adjacent haplotype consisting of rs709157 (SNP19), rs1175540 (SNP20), and rs1175544 (SNP21) also resulted in P value smaller than the Bonferroni correction threshold of 0.0023.

As shown in **Table 4**, there are three common haplotypes (i.e. haplotype frequency ≥ 1%) formed by rs1175540(C/A)-rs1175544(C/T)-rs1797912(A/C): h000 (51.7% in controls and 53.7% in cases), h111 (22.9% in controls and 25.8% in cases), and h100 (18.2% in controls and 12.8% in cases), where “0” and “1” denoted the major and minor alleles at each SNP locus. Using the most common haplotype h000 as the referent group, h100 is found to have significant

association with diabetes risk (OR = 1.02, 95% CI = 1.00-1.03, P value 0.002). For another haplotype that is formed by rs1175540(C/A), rs1175544(C/T), and rs1797912(A/C), h010 is significantly associated with diabetes risk (OR = 1.45, 95% CI = 1.18-1.77, P value 0.0003) using h000 as the referent group. By combining four adjacent windows that are near to the *Pro12Ala* (rs1801282) SNP, a combined haplotype was formed, i.e. rs12629293(A/G)-rs12636454(T/C)-rs4518111(C/A)-rs10510418(A/C)-rs1801282(C/G)-rs1373640(C/T). Using h001000 as the reference group, h000000 was significantly associated with diabetes risk (OR = 1.47, 95% CI = 1.16-1.86, P value = 0.001).

2C.3.4 Validation analysis in WHI-SHARe population

Table 5 showed the additional analyses conducted in the WHI-SHARe to validate the single-SNP findings. Among Hispanic in the WHI-SHARe population, the rs1801282 SNP was also found to have association with T2D risk (nominal P value = 0.03). With adjustment for multiple comparisons, none of the eight tagSNPs were found to have significant association with diabetes risk using the additive genetic model (all q values > 0.05). The results under the dominant and recessive genetic model are shown in **Supplementary Table 3** and **4**.

In the GSEA, the PPAR signaling pathway was among the top ten pathways (nominal p-value = 0.002) among 5,729 African American women in the WHI-SHARe cohort.

2C.4 Discussion

We replicated the association between the well-studied *Pro12Ala* (rs1801282) SNP and diabetes risk among the WHI-OS Hispanic population. However, none of our 24 tagSNPs in the *PPARG* gene were significantly associated with diabetes in the WHI-OS and WHI-SHARe

cohorts of postmenopausal women. Using a sliding-window width of 3, 2 haplotypes were found to have significant associations with diabetes risk after adjusting for multiple comparisons. We also observed the PPAR signaling pathway ranked among the top ten pathways in the GSEA among the WHI-SHARe African American population.

In the literature, the association between the *PPARG* locus and diabetes risk has been investigated in various ethnic populations. The *PPARG Pro12Ala* (rs1801282) was found to have significant association with diabetes risk among Swedish[12], Finnish[7,16], German[9], North Indian[15], and Chinese populations[10,19]. In our study, the rs1801282 SNP appeared to be associated with diabetes risk among Hispanic women in both WHI-OS and WHI-SHARe populations (with nominal P values <0.05). Our null findings after multiple comparisons adjustment for a relationship between genetic variants of the *PPARG* gene and diabetes risk were consistent with results from previous studies of Denmark and European American[13,27] as well as Chinese population[11]. This is likely due to limited power, especially after stratifying according to ethnic group.

Because haplotype-based analysis is arguably more powerful than single-marker analysis[28,29], our comprehensive assessment of genetic variants in the *PPARG* genomic region allowed us to perform the global haplotype-based LRT as the main haplotype test, and also haplotype-specific analyses using a sliding-window approach. Two haplotypes (rs709157(G/A)-rs1175540(C/A)-rs1175544(C/T) and rs1175540(C/A)-rs1175544(C/T)-rs1797912(A/C)) appeared to be associated with diabetes risk among women in the WHI-OS cohort. For example, individuals with the h010 haplotype combination, i.e. rs709157-G, rs1175540-A, and rs1175544-C, appeared to have higher diabetes risk (OR=1.45, 95%CI=1.18-1.77, P value=0.0003) comparing to those individuals having h000 combination, i.e. rs709157-G, rs1175540-C, and rs1175544-C.

Haplotype-based analysis allows the examination of the combined effect of several SNPs[30]. This analytical method offers the following advantages: it 1) captures epistastic interactions between SNPs at a locus; 2) allows testing between haplotype alleles by encapsulating information from evolutionary history; 3) reduces the number of tests as well as the type 1 error rate and therefore provide more power than single-SNP analyses[31].

Thus far, several genome-wide scans for diabetes have been reported. For example, Wellcome Trust Case Control Consortium (WTCCC)[32-35], Diabetes Genetics Initiative (DGI) scan[33,34], Finland-United States Investigation of NIDDM Genetics (FUSION) scan[35,36], and the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) study. Two Genome-wide Association Study (GWAS) consortium diabetes and/or diabetes related conditions have been recently formed- Meta-analysis of Type 2 Diabetes in African Americans Consortium (MEDIA), and the Genomics and Randomized Trials Network, or Genome-wide Association Research Network into Effects of Treatments (GARNET). Recently, a HuGE review and meta-analysis showed a moderate level of heterogeneity attributable to genuine variation in gene effect size for the *PPARG Pro12Ala* genetic variant, which reflected the variation observed between ethnic populations as well as differences in body mass index[8]. This may be explained by the role played by ethnicity and differences in dietary and habits[37]. Our multiethnic population, which comprised of European, African, Hispanic, and Asian Americans with comprehensive assessment of demographic and lifestyle variables, may provide meaningful findings to the literature regarding the association between common variants of the *PPARG* gene and diabetes risk.

Candidate gene approach, particularly at common low penetrance susceptibility loci in complex diseases may become a major component of following up the genes emerging from GWAS to establish a functional rationale underlying the importance of allelic variation[38].

Further, a candidate gene study takes advantage of both increased statistical efficiency of association analysis of complex diseases together with clinical and biological understanding of the phenotype. Because we employed a prospective design of a well-characterized source population, our study thus minimized selection biases and confounding.

In conclusion, our multiethnic case-control study of postmenopausal women replicated the association between the *PPARG Pro12Ala* genetic variant with diabetes risk and also found that haplotype-based analysis is more powerful than single-SNP analysis for identifying genetic variants in the *PPARG* gene with diabetes risk.

Table 1. Minor allele frequencies (MAFs) of the 24 tagSNPs in the *PPARG* gene.

dbSNP ID	SNP ID	Genome Coordinate[a]	Major/ Minor Allele	MAF (%)[b]					*P* for heterogeneity[c]
				European American	African American	Hispanic American	Asian American	Combined	
				(n=942)	(n=744)	(n=276)	(n=162)	(n=2,124)	
rs9878908	*SNP1*	12242462	*T/C*	0.2022	0.1122	0.1178	0.3302	0.1695	<.0001
rs6798713	*SNP2*	12245587	*T/C*	0.1812	0.1938	0.1137	0.3282	0.1881	<.0001
rs6809631	*SNP3*	12275647	*A/T*	0.2607	0.2889	0.3073	0.4785	0.2934	<.0001
rs9817428	*SNP4*	12280267	*C/A*	0.2593	0.3616	0.3147	0.4697	0.3186	<.0001
rs10510411	*SNP5*	12286849	*G/A*	0.2603	0.3038	0.3105	0.2638	0.2822	0.1039
rs12629293	*SNP6*	12291746	*A/G*	0.2641	0.2293	0.3011	0.2813	0.2581	0.0577
rs12636454	*SNP7*	12300214	*T/C*	0.2611	0.2969	0.3032	0.2719	0.2799	0.3227
rs4518111	*SNP8*	12317344	*C/A*	0.4033	0.1835	0.3859	0.5123	0.3328	<.0001
rs10510418	*SNP9*	12328563	*A/C*	0.3259	0.111	0.2744	0.2284	0.2369	<.0001
rs1801282	*SNP10*	12333125	*C/G*	0.1192	0.0255	0.0655	0.0494	0.0742	<.0001
rs1373640	*SNP12*	12342601	*C/T*	0.339	0.1198	0.2591	0.1883	0.2403	<.0001
rs2972162	*SNP13*	12364793	*C/T*	0.5207	0.3351	0.5345	0.4286	0.4505	<.0001
rs10510419	*SNP14*	12366936	*G/T*	0.1457	0.1159	0.2527	0.0123	0.1391	<.0001
rs2959272	*SNP16*	12382833	*C/A*	0.5277	0.4158	0.5469	0.4294	0.4837	<.0001
rs709150	*SNP18*	12391337	*C/G*	0.4721	0.2255	0.4063	0.5648	0.3845	<.0001
rs709157	*SNP19*	12402024	*G/A*	0.3047	0.0875	0.2364	0.0245	0.1983	<.0001
rs1175540	*SNP20*	12405243	*C/A*	0.3607	0.6514	0.317	0.4146	0.4607	<.0001
rs1175544	*SNP21*	12407044	*C/T*	0.3231	0.1292	0.2473	0.3742	0.2495	<.0001
rs1797912	*SNP22*	12410239	*A/C*	0.3672	0.1523	0.2726	0.4146	0.2846	<.0001
rs1152002	*SNP23*	12411871	*G/A*	0.4856	0.4298	0.3727	0.4479	0.4486	0.0001
rs3856806	*SNP24*	12415557	*C/T*	0.1303	0.0643	0.074	0.1933	0.1047	<.0001
rs1152003	*SNP25*	12417055	*C/G*	0.3383	0.6171	0.5163	0.5185	0.4725	<.0001
rs1152007	*SNP26*	12427547	*C/G*	0.3455	0.1808	0.3327	0.4634	0.2955	<.0001
rs709167	*SNP28*	12442955	*A/C*	0.4601	0.2917	0.4076	0.8025	0.4204	<.0001

[a]: Genome coordinate was according to chromosome 3 genomic contig (reference assembly) NT_022517.18.

[b]: MAF was estimated in the controls only (The minor allele was defined based on the entire control population).

[c]: The *P* value was estimated based on a χ^2 test (d.f. = 3) for genotype distribution across the four ethnicities.

Table 2. Single-SNP association studies of the 24 tagSNPs in the *PPARG* gene with diabetes risk under additive model[a].

SNP ID	White	Black	Hispanic	Asian	Combined	*P* value
	(855/872)[b]	(301/634)	(113/241)	(67/151)	(1336/1898)	
rs9878908	0.85(0.55-1.32)	0.75(0.35-1.61)	1.18(0.45-3.06)	1.00(0.62-1.60)	0.89(0.65-1.21)	0.46
rs6798713	0.75(0.47-1.17)	0.98(0.57-1.67)	1.40(0.52-3.74)	1.06(0.67-1.67)	0.89(0.66-1.20)	0.44
rs6809631	0.71(0.46-1.10)	0.78(0.50-1.23)	0.66(0.32-1.38)	0.80(0.48-1.34)	**0.76(0.58-0.99)**	**0.04[e]**
rs9817428	0.76(0.49-1.17)	0.80(0.53-1.21)	0.66(0.32-1.37)	0.78(0.47-1.30)	**0.78(0.60-1.00)**	**0.05[f]**
rs10510411	0.68(0.44-1.06)	0.64(0.37-1.11)	0.78(0.38-1.60)	1.06(0.66-1.71)	**0.68(0.52-0.91)**	**0.01[g]**
rs12629293	0.66(0.43-1.04)	0.78(0.49-1.24)	0.78(0.38-1.58)	1.05(0.65-1.70)	**0.72(0.55-0.94)**	**0.02[h]**
rs12636454	0.68(0.44-1.06)	0.79(0.52-1.21)	0.76(0.37-1.54)	1.10(0.69-1.76)	**0.73(0.57-0.95)**	**0.02[i]**
rs4518111	1.36(0.91-2.03)	1.06(0.65-1.75)	1.40(0.70-2.82)	1.18(0.74-1.89)	1.20(0.94-1.54)	0.15
rs10510418	1.12(0.74-1.69)	0.75(0.43-1.33)	0.91(0.40-2.06)	0.69(0.38-1.27)	1.03(0.79-1.36)	0.81
rs1801282	0.62(0.35-1.12)	0.59(0.15-2.31)	**0.27(0.08-0.93)[c]**	0.21(0.04-1.04)	**0.55(0.35-0.86)**	**0.01[j]**
rs1373640	1.17(0.79-1.75)	0.86(0.49-1.49)	1.14(0.53-2.46)	0.56(0.27-1.12)	1.11(0.84-1.45)	0.46
rs2972162	1.23(0.82-1.84)	1.19(0.79-1.80)	1.02(0.51-2.04)	0.98(0.58-1.67)	1.17(0.92-1.48)	0.20
rs10510419	1.45(0.83-2.54)	1.25(0.69-2.26)	0.98(0.39-2.45)	1.49(0.19-11.51)	1.22(0.86-1.72)	0.26
rs2959272	1.22(0.82-1.83)	1.12(0.75-1.68)	0.99(0.49-2.00)	1.08(0.64-1.80)	1.15(0.91-1.46)	0.24
rs709150	0.67(0.44-1.03)	0.80(0.48-1.31)	0.99(0.50-1.96)	0.94(0.56-1.57)	0.80(0.62-1.03)	0.08
rs709157	1.32(0.88-1.98)	0.67(0.34-1.31)	0.96(0.44-2.06)	---[d]	1.08(0.81-1.43)	0.61
rs1175540	1.12(0.76-1.65)	1.04(0.68-1.60)	0.97(0.46-2.03)	1.18(0.69-2.02)	1.08(0.85-1.37)	0.53
rs1175544	1.21(0.82-1.79)	0.81(0.45-1.47)	0.99(0.45-2.17)	1.25(0.74-2.10)	1.11(0.85-1.44)	0.44
rs1797912	1.18(0.80-1.74)	0.91(0.56-1.49)	1.09(0.49-2.45)	1.00(0.60-1.65)	1.12(0.86-1.44)	0.40
rs1152002	0.93(0.64-1.37)	0.94(0.62-1.43)	2.01(0.92-4.41)	0.84(0.51-1.37)	1.06(0.83-1.35)	0.65
rs3856806	0.93(0.50-1.71)	0.52(0.22-1.22)	0.31(0.09-1.07)	0.83(0.47-1.48)	0.71(0.48-1.07)	0.10
rs1152003	1.09(0.73-1.63)	0.96(0.65-1.41)	1.72(0.84-3.50)	1.01(0.62-1.64)	1.07(0.85-1.36)	0.56
rs1152007	1.15(0.74-1.81)	0.74(0.45-1.24)	1.10(0.46-2.64)	0.88(0.56-1.37)	0.92(0.70-1.21)	0.55
rs709167	1.06(0.73-1.55)	1.09(0.66-1.79)	1.12(0.49-2.54)	0.86(0.47-1.57)	1.06(0.82-1.36)	0.68

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the additive genetic model, using conditional logistic regression models with adjustments for age, ethnicity (combined analysis only), body mass index (BMI), ln(fasting insulin), ln(fasting glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline. Due to the small Asian population size, BMI, ln(fasting insulin), ln(fasting glucose) were excluded to cause the model to converge.

[b]: Sample size is presented as cases/controls.

[c]: P-value = 0.04, with q = 0.65 after FDR using the method of Benjamini and Hochberg.

[d]: Result is difficult to interpret because of small sample size within strata.

[e]: P-value = 0.04, with q = 0.19 after FDR using the method of Benjamini and Hochberg.
[f]: P-value = 0.05, with q = 0.21 after FDR using the method of Benjamini and Hochberg.
[g]: P-value = 0.01, with q = 0.11 after FDR using the method of Benjamini and Hochberg.
[h]: P-value = 0.02, with q = 0.12 after FDR using the method of Benjamini and Hochberg.
[i]: P-value = 0.02, with q = 0.12 after FDR using the method of Benjamini and Hochberg.
[j]: P-value = 0.01, with q = 0.11 after FDR using the method of Benjamini and Hochberg.

Table 3. Association between the non-synonymous SNP (rs1801282) and diabetes risk by genotype[a].

Genotype	White (855/872)[b]	Black (301/634)	Hispanic (113/241)	Asian (67/151)	Combined (1336/1898)
rs1801282					
CC	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
CG	0.60(0.32-1.14)	0.59(0.15-2.31)	**0.17(0.04-0.77)[d]**	0.13(0.01-1.27)	**0.51(0.31-0.83)[e]**
GG	0.58(0.03-10.8)	---[c]	---[e]	---[c]	0.77(0.07-8.02)

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the dominant genetic model, using conditional logistic regression models with adjustments for age, ethnicity (combined analysis only), body mass index (BMI), ln(fasting insulin), ln(fasting glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline. Due to the small Asian population size, BMI, ln(fasting insulin), ln(fasting glucose) were excluded to cause the model to converge.
[b]: Sample size is presented as cases/controls.
[c]: Participants do not possess this genotype.
[d]: P value = 0.02.
[e]: P value = 0.006.

Table 4. Haplotype-specific analyses for the *PPARG* haplotypes and diabetes risk in the combined population[a].

Haplotype[b]	Frequency in controls (%)	Frequency in cases (%)	All	*P* value
rs709157(G/A)-rs1175540(C/A)-rs1175544(C/T)				
h000	53.6	55.7	1.00 (ref)	---
h111	18.8	22.7	0.95(0.78-1.16)	0.61
h010	20.4	14.9	1.45(1.18-1.77)	0.0003
h011	5.67	4.80	0.97(0.67-1.42)	0.88
rs1175540(C/A)-rs1175544(C/T)-rs1797912(A/C)				
h000	51.7	53.7	1.00 (ref)	---
h111	22.9	25.8	1.40 (1.13-1.72)	0.27
h100	18.2	12.8	1.02(1.00-1.03)	0.002
rs12629293(A/G)-rs12636454(T/C)-rs4518111(C/A)-rs10510418(A/C)-rs1801282(C/G)-rs1373640(C/T)				
h001000	31.2	33.3	1.00 (ref)	---
h000101	21.5	24.0	1.03(0.84-1.26)	0.77
h110000	17.7	16.6	1.14(0.91-1.44)	0.25
h000000	15.9	11.5	1.47(1.16-1.86)	0.001
h110010	6.55	7.45	1.21(0.87-1.68)	0.25

[a]: Adjusted OR (95% CI) for each haplotype was estimated under the additive genetic model, using logistic regression models with adjustments for matching factors (age, ethnicity, clinical center, time of blood draw), body mass index (BMI), ln(fasting insulin), ln(fasting glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline.

[b]: At each SNP locus, "0" and "1" denote the major and minor alleles, respectively. Only those haplotypes **with frequencies ≥ 5%** are reported.

Table 5. Single-SNP association studies of the 8 tagSNPs (captured in WHI-SHARe) in the *PPARG* gene with diabetes risk under additive genetic model[a] (n=161/921)[b].

SNP ID	SNP name	Reference allele	Black (103/592)[b]	Hispanic (58/329)[b]	Combined	P-value for ethnic interaction[c]	P-value for likelihood ratio test[d]
rs9878908	SNP_A-1875778	C	0.84(0.52-1.38)	0.90(0.43-1.88)	0.90(0.61-1.34)	0.74	0.64
rs9817428	SNP_A-1949196	A	0.99(0.73-1.34)	0.57(0.32-1.02)	0.89(0.68-1.16)	0.14	0.45
rs10510418	SNP_A-1971789	C	0.98(0.61-1.57)	1.04(0.56-1.91)	0.94(0.65-1.35)	0.96	0.70
rs1801282	SNP_A-1971790	G	0.83(0.28-2.41)	**0.36(0.14-0.91)[e]**	0.60(0.31-1.16)	0.46	0.16
rs2972162	SNP_A-1946610	T	0.95(0.69-1.31)	1.06(0.64-1.76)	1.01(0.77-1.32)	0.57	0.94
rs10510419	SNP_A-4209319	T	0.90(0.56-1.46)	1.64(0.85-3.17)	1.16(0.79-1.68)	0.14	0.53
rs1175544	SNP_A-8304334	T	1.06(0.68-1.65)	1.37(0.74-2.55)	1.09(0.77-1.55)	0.74	0.69
rs1152003	SNP_A-2140799	G	1.14(0.83-1.56)	0.76(0.47-1.23)	1.04(0.80-1.35)	0.22	0.78

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the additive genetic model, using logistic regression adjustments for global ancestry (3 PCs), age, ethnicity (combined analysis only), body mass index (BMI), ln(insulin), ln(glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline.
[b]: Sample sizes for each ethnic group are presented as cases/controls.
[c]: Ethnic interaction was estimated by fitting a model with race*SNP interaction term and adjusting for global ancestry using 3 PCs.
[d]: The likelihood ratio test compared model with SNP versus model without SNP.
[e]: P-value = 0.03, with q = 0.23 after FDR using the method of Benjamini and Hochberg.

Supplementary Table 1. Single-SNP association studies of the 24 tagSNPs in the *PPARG* gene with diabetes risk under dominant genetic model[a].

SNP ID	White	Black	Hispanic	Asian	Combined	*P* value
	(855/872)[b]	(301/634)	(113/241)	(67/151)	(1336/1898)	
rs9878908	0.89(0.51-1.54)	0.78(0.35-1.76)	1.13(0.41-3.10)	1.04(0.53-2.06)	0.90(0.62-1.31)	0.58
rs6798713	0.74(0.43-1.28)	0.98(0.53-1.80)	1.37(0.49-3.82)	1.10(0.55-2.20)	0.90(0.64-1.27)	0.55
rs6809631	0.71(0.40-1.23)	0.67(0.36-1.22)	0.54(0.21-1.35)	0.57(0.25-1.30)	0.67(0.48-0.95)	0.03
rs9817428	0.76(0.44-1.31)	0.61(0.34-1.12)	0.53(0.21-1.34)	0.55(0.24-1.24)	0.68(0.48-0.96)	0.03
rs10510411	0.69(0.40-1.20)	0.63(0.33-1.22)	0.69(0.28-1.72)	0.98(0.51-1.88)	0.65(0.46-0.93)	0.02
rs12629293	0.69(0.39-1.21)	0.78(0.45-1.36)	0.68(0.28-1.65)	0.97(0.51-1.85)	0.70(0.50-0.98)	0.04
rs12636454	0.72(0.42-1.24)	0.74(0.43-1.29)	0.65(0.27-1.56)	1.08(0.57-2.04)	0.71(0.51-0.99)	0.05
rs4518111	1.41(0.78-2.53)	1.15(0.64-2.09)	1.73(0.61-4.87)	1.07(0.50-2.27)	1.34(0.94-1.90)	0.10
rs10510418	1.00(0.60-1.68)	0.68(0.35-1.32)	1.23(0.42-3.56)	0.83(0.41-1.64)	1.00(0.71-1.40)	0.99
rs1801282	0.60(0.32-1.12)	0.59(0.15-2.31)	**0.19(0.04-0.81)[d]**	0.21(0.04-1.04)	**0.51(0.32-0.83)**	**0.01[h]**
rs1373640	1.07(0.65-1.77)	0.83(0.43-1.60)	1.77(0.63-4.99)	0.59(0.27-1.28)	1.10(0.78-1.54)	0.58
rs2972162	0.97(0.51-1.84)	1.09(0.64-1.87)	2.02(0.61-6.71)	0.99(0.45-2.16)	1.17(0.82-1.68)	0.38
rs10510419	1.28(0.68-2.39)	1.13(0.58-2.21)	0.87(0.30-2.53)	1.49(0.19-11.5)	1.11(0.75-1.64)	0.60
rs2959272	0.85(0.44-1.65)	1.03(0.56-1.88)	1.62(0.50-5.20)	1.20(0.54-2.66)	1.10(0.75-1.60)	0.63
rs709150	**0.43(0.22-0.85)[c]**	0.78(0.43-1.43)	1.94(0.58-6.52)	0.93(0.38-2.29)	0.75(0.52-1.09)	0.13
rs709157	1.29(0.78-2.14)	0.67(0.33-1.39)	1.18(0.45-3.06)	---[g]	1.04(0.74-1.48)	0.81
rs1175540	1.03(0.61-1.73)	0.84(0.33-2.14)	1.42(0.55-3.65)	1.17(0.53-2.59)	1.12(0.78-1.62)	0.54
rs1175544	1.18(0.70-2.00)	0.79(0.40-1.55)	1.42(0.53-3.79)	1.20(0.58-2.51)	1.18(0.83-1.66)	0.36
rs1797912	1.13(0.67-1.92)	0.84(0.45-1.60)	1.57(0.56-4.34)	0.96(0.46-1.97)	1.16(0.82-1.63)	0.41
rs1152002	0.84(0.45-1.57)	0.88(0.49-1.61)	1.57(0.56-4.34)	0.79(0.39-1.56)	1.12(0.78-1.60)	0.55
rs3856806	0.94(0.50-1.77)	0.52(0.22-1.22)	**4.51(1.28-15.9)[e]**	0.79(0.39-1.59)	0.70(0.46-1.07)	0.10
rs1152003	1.01(0.59-1.72)	0.75(0.36-1.58)	**0.22(0.05-0.97)[f]**	1.00(0.43-2.32)	1.01(0.70-1.45)	0.97
rs1152007	1.14(0.64-2.04)	0.76(0.41-1.42)	1.77(0.59-5.33)	0.94(0.46-1.93)	0.94(0.66-1.35)	0.74
rs709167	1.24(0.70-2.21)	1.28(0.70-2.36)	0.99(0.32-3.09)	0.75(0.17-3.27)	1.23(0.85-1.77)	0.27

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the dominant genetic model, using conditional logistic regression models with adjustments for age, ethnicity (combined analysis only), body mass index (BMI), ln(fasting insulin), ln(fasting glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline. Due to the small Asian population size, BMI, ln(fasting insulin), ln(fasting glucose) were excluded to cause the model to converge.

[b]: Sample size is presented as cases/controls.

[c]: P-value = 0.02, with q = 0.37 after FDR using the method of Benjamini and Hochberg.
[d]: P-value = 0.03, with q = 0.30 after FDR using the method of Benjamini and Hochberg.
[e]: P-value = 0.02, with q = 0.30 after FDR using the method of Benjamini and Hochberg.
[e]: P-value = 0.05, with q = 0.37 after FDR using the method of Benjamini and Hochberg.
[g]: Result is difficult to interpret because of small sample size within strata.
[h]: P-value = 0.01, with q = 0.16 after FDR using the method of Benjamini and Hochberg.

Supplementary Table 2. Single-SNP association studies of the 24 tagSNPs in the *PPARG* gene with diabetes risk under recessive genetic model[a].

SNP ID	White	Black	Hispanic	Asian	Combined	*P* value
	(855/872)[b]	(301/634)	(113/241)	(67/151)	(1336/1898)	
rs9878908	0.59(0.20-1.73)	---[d]	---[d]	0.91(0.35-2.36)	0.71(0.31-1.66)	0.43
rs6798713	0.50(0.14-1.79)	0.91(0.17-5.00)	---[d]	1.07(0.44-2.62)	0.72(0.30-1.71)	0.45
rs6809631	0.48(0.17-1.34)	0.93(0.33-2.64)	0.91(0.18-4.77)	0.98(0.45-2.16)	0.79(0.44-1.43)	0.44
rs9817428	0.54(0.18-1.57)	1.05(0.46-2.39)	0.89(0.18-4.32)	0.96(0.44-2.08)	0.85(0.49-1.47)	0.56
rs10510411	0.41(0.14-1.21)	0.49(0.14-1.76)	0.91(0.19-4.42)	1.43(0.48-4.24)	0.55(0.29-1.04)	0.07
rs12629293	0.33(0.11-1.01)	0.54(0.14-2.00)	0.95(0.19-4.84)	1.39(0.47-4.12)	**0.51(0.26-0.99)**	**0.05**[e]
rs12636454	0.33(0.11-1.02)	0.70(0.26-1.88)	1.00(0.20-5.09)	1.32(0.46-3.82)	0.55(0.30-1.03)	0.06
rs4518111	1.66(0.79-3.48)	0.71(0.16-3.16)	1.32(0.36-4.87)	1.49(0.68-3.29)	1.15(0.70-1.87)	0.59
rs10510418	1.83(0.70-4.81)	1.08(0.15-7.63)	0.38(0.06-2.33)	---[d]	1.23(0.63-2.41)	0.55
rs1801282	0.60(0.03-11.3)	0.82(0.15-4.56)	---[d]	---[d]	0.79(0.08-8.37)	0.85
rs1373640	1.87(0.73-4.76)	1.81(0.75-4.40)	0.38(0.06-2.30)	---[d]	1.27(0.66-2.42)	0.48
rs2972162	1.76(0.92-3.39)	4.32(0.58-32.0)	0.53(0.16-1.78)	0.96(0.36-2.57)	1.30(0.85-1.99)	0.22
rs10510419	**14.4(1.71-121)**[c]	1.40(0.67-2.94)	1.80(0.16-20.4)	---[d]	**3.83(1.15-12.82)**	**0.03**[f]
rs2959272	1.90(1.00-3.62)	0.65(0.17-2.45)	0.58(0.17-2.00)	1.00(0.40-2.50)	1.35(0.90-2.00)	0.14
rs709150	0.87(0.45-1.67)	0.30(0.02-5.56)	0.49(0.14-1.67)	0.92(0.42-2.03)	0.74(0.47-1.17)	0.20
rs709157	1.98(0.73-5.43)	1.13(0.65-1.97)	0.44(0.07-2.85)	---[d]	1.34(0.65-2.75)	0.43
rs1175540	1.64(0.67-4.02)	0.78(0.12-4.97)	0.24(0.04-1.55)	1.34(0.52-3.45)	1.09(0.71-1.65)	0.70
rs1175544	1.67(0.68-4.10)	1.07(0.30-3.84)	0.28(0.04-1.94)	1.58(0.61-4.12)	1.05(0.58-1.90)	0.88
rs1797912	1.59(0.69-3.69)	1.00(0.47-2.16)	0.33(0.05-2.37)	1.07(0.42-2.76)	1.16(0.66-2.04)	0.61
rs1152002	0.99(0.54-1.83)	1.08(0.62-1.87)	0.83(0.18-3.78)	0.83(0.32-2.11)	1.02(0.67-1.54)	0.94
rs3856806	0.46(0.01-23.8)	0.42(0.10-1.73)	---[d]	0.83(0.17-4.08)	0.62(0.06-6.13)	0.68
rs1152003	1.44(0.63-3.34)	0.64(0.20-2.04)	2.43(0.73-8.07)	1.02(0.49-2.14)	1.21(0.81-1.80)	0.35
rs1152007	1.37(0.51-3.65)	---[d]	1.45(0.28-7.48)	0.71(0.32-1.60)	0.79(0.43-1.45)	0.45
rs709167	0.90(0.47-1.71)	0.91(0.17-5.00)	0.51(0.11-2.31)	0.85(0.39-1.83)	0.86(0.55-1.36)	0.53

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the recessive genetic model, using conditional logistic regression models with adjustments for age, ethnicity (combined analysis only), body mass index (BMI), ln(fasting insulin), ln(fasting glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline. Due to the small Asian population size, BMI, ln(fasting insulin), ln(fasting glucose) were excluded to cause the model to converge.

[b]: Sample size is presented as cases/controls.

[c]: P-value = 0.01, with q = 0.32 after FDR using the method of Benjamini and Hochberg.
[d]: Result is difficult to interpret because of small sample size within strata.
[e]: P-value = 0.05, with q = 0.40 after FDR using the method of Benjamini and Hochberg.
[f]: P-value = 0.03, with q = 0.40 after FDR using the method of Benjamini and Hochberg.

Supplementary Table 3. Single-SNP association studies of the 8 tagSNPs (captured in WHI-SHARe) in the *PPARG* gene with diabetes risk under dominant genetic model[a] (n=161/921)[b].

SNP ID	SNP name	Reference allele	Black (103/592)[b]	Hispanic (58/329)[b]	Combined	P-value for ethnic interaction[c]	P-value for likelihood ratio test[d]
rs9878908	SNP_A-1875778	C	0.87(0.50-1.49)	0.82(0.37-1.81)	0.90(0.58-1.40)	0.92	0.67
rs9817428	SNP_A-1949196	A	1.08(0.70-1.66)	0.61(0.30-1.28)	0.98(0.68-1.41)	0.24	0.98
rs10510418	SNP_A-1971789	C	1.03(0.61-1.75)	1.03(0.50-2.15)	0.98(0.65-1.49)	0.89	0.91
rs1801282	SNP_A-1971790	G	0.84(0.28-2.48)	**0.30(0.11-0.82)**[e]	0.57(0.29-1.14)	0.38	0.14
rs2972162	SNP_A-1946610	T	0.98(0.50-1.92)	1.11(0.47-2.64)	1.07(0.64-1.80)	0.91	0.78
rs10510419	SNP_A-4209319	T	10.8(0.54-213)	---[f]	---[f]	---[f]	---[f]
rs1175544	SNP_A-8304334	T	1.16(0.71-1.92)	1.87(0.88-3.95)	1.28(0.86-1.92)	0.50	0.27
rs1152003	SNP_A-2140799	G	1.01(0.57-1.78)	0.64(0.30-1.40)	0.90(0.57-1.41)	0.39	0.63

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the dominant genetic model, using logistic regression adjustments for global ancestry (3 PCs), age, ethnicity (combined analysis only), body mass index (BMI), ln(insulin), ln(glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline.
[b]: Sample sizes for each ethnic group are presented as cases/controls.
[c]: Ethnic interaction was estimated by fitting a model with race*SNP interaction term and adjusting for global ancestry using 3 PCs.
[d]: The likelihood ratio test compared model with SNP versus model without SNP.

[e]: P-value = 0.02, with q = 0.16 after FDR using the method of Benjamini and Hochberg.
[f]: Result is difficult to interpret because of small sample size within strata

Supplementary Table 4. Single-SNP association studies of the 8 tagpSNPs (captured in WHI-SHARe) in the *PPARG* gene with diabetes under recessive model[a] (n=161/921)[b].

SNP ID	SNP name	Reference allele	Black (103/592)[b]	Hispanic (58/329)[b]	Combined	P-value for ethnic interaction[c]	P-value for likelihood ratio test[d]
rs9878908	SNP_A-1875778	C	0.47(0.07-3.33)	2.48(0.21-29.0)	0.77(0.16-3.61)	0.34	0.78
rs9817428	SNP_A-1949196	A	0.80(0.42-1.52)	0.30(0.01-1.10)	0.64(0.36-1.12)	0.22	0.13
rs10510418	SNP_A-1971789	C	0.58(0.10-3.18)	1.09(0.22-5.44)	0.61(0.18-2.03)	0.68	0.37
rs1801282	SNP_A-1971790	G	---[e]	---[e]	1.31(0.004-43.2)	0.98	0.88
rs2972162	SNP_A-1946610	T	0.93(0.60-1.42)	1.05(0.48-2.31)	0.98(0.68-1.43)	0.68	0.92
rs10510419	SNP_A-4209319	T	0.83(0.51-1.37)	1.51(0.74-3.09)	1.05(0.70-1.56)	0.16	0.92
rs1175544	SNP_A-8304334	T	0.51(0.10-2.51)	0.28(0.002-3.16)	0.36(0.10-1.30)	0.68	0.08
rs1152003	SNP_A-2140799	G	1.30(0.83-2.03)	0.75(0.34-1.66)	1.18(0.80-1.73)	0.32	0.41

[a]: Adjusted OR (95% CI) for each tagSNP was estimated under the recessive genetic model, using logistic regression adjustments for global ancestry (3 PCs), age, ethnicity (combined analysis only), body mass index (BMI)), ln(insulin), ln(glucose), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), diabetes family history (presence/absence), and physical activity per week at baseline.

[b] Sample sizes for each ethnic group are presented as cases/controls.

[c] Ethnic interaction was estimated by fitting a model with race*SNP interaction term and adjusting for global ancestry using 3 PCs.

[d] The likelihood ratio test compared model with SNP versus model without SNP.
[e] Result is difficult to interpret because of small sample size within strata.

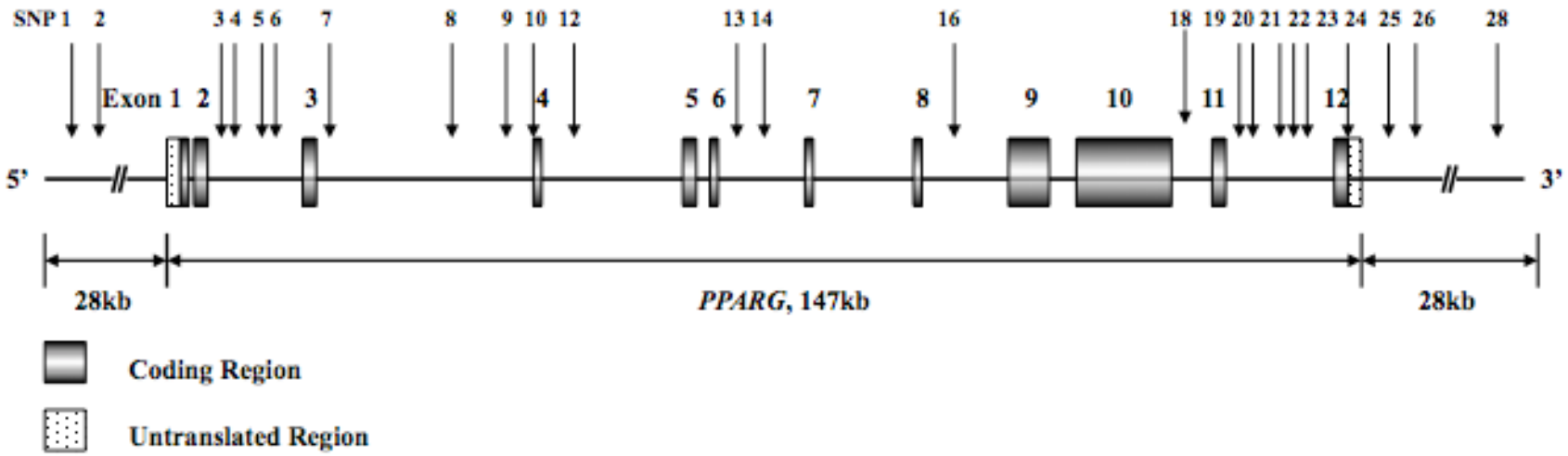


Figure 1. The schematic presentation showed the position of 24 tagSNPs that spanned the peroxisome proliferator-activated receptor γ (*PPARG*) genomic region.

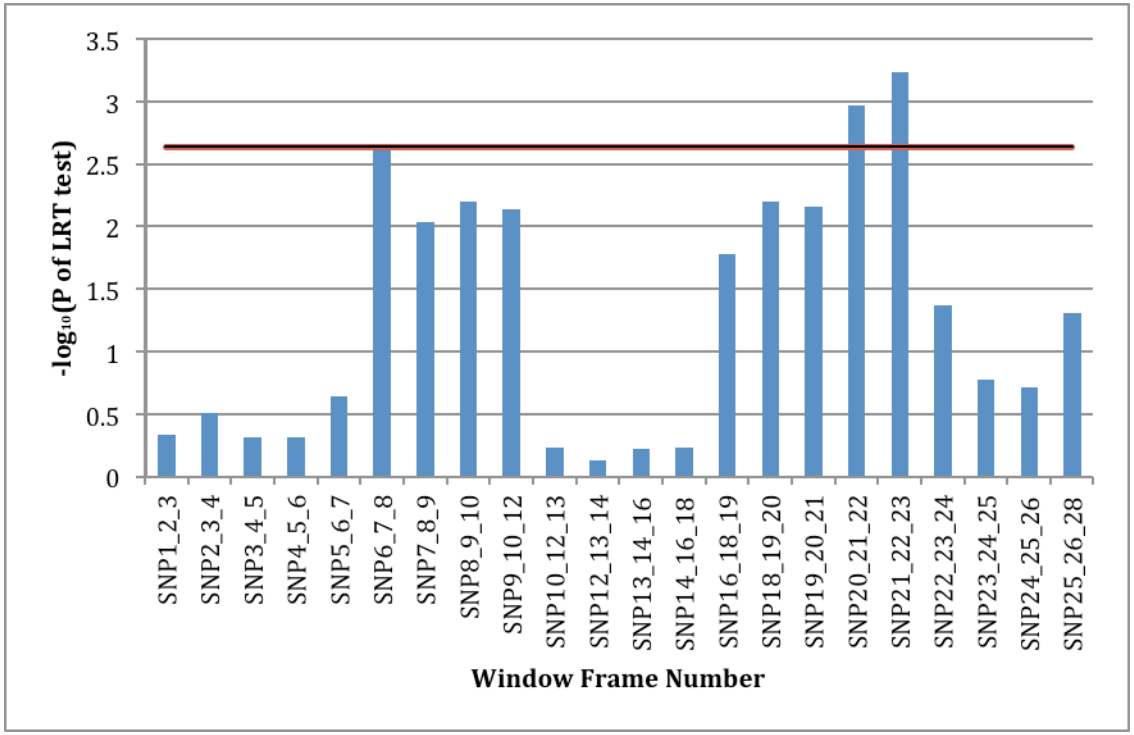


Figure 2. Sliding-window (window width = 3) haplotype-based analysis of 24 *PPARG* tagSNPs using additive genetic model. Haplotype effects were estimated for the combined multiethnic population, using an omnibus likelihood ratio test. The x-axis denoted the sliding window frames, and the y-axis denoted the $-\log_{10}$(P value). A $-\log_{10}$(P value) > 2.64 was employed as the global significance threshold using Bonferroni correction for 22 window frames (red line).

2C.6 References

1. Szanto A, Nagy L. The many faces of PPARgamma: anti-inflammatory by any means? Immunobiology 2008;213:789-803.

2. Ahmed W, Ziouzenkova O, Brown J, et al. PPARs and their metabolic modulation: new mechanisms for transcriptional regulation? J Intern Med 2007;262:184-98.

3. Blackburn GL. From bench to bedside: novel mechanisms and therapeutic advances through the development of selective peroxisome proliferator-activated receptor gamma modulators. Am J Clin Nutr;91:251S-3S.

4. Cheatham WW. Peroxisome proliferator-activated receptor translational research and clinical experience. Am J Clin Nutr;91:262S-6S.

5. Ryan KK, Li B, Grayson BE, Matter EK, Woods SC, Seeley RJ. A role for central nervous system PPAR-gamma in the regulation of energy balance. Nat Med;17:623-6.

6. Kahn BB, McGraw TE. Rosiglitazone, PPARgamma, and type 2 diabetes. N Engl J Med 2010;363:2667-9.

7. Gaulton KJ, Willer CJ, Li Y, et al. Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes. Diabetes 2008;57:3136-44.

8. Gouda HN, Sagoo GS, Harding AH, Yates J, Sandhu MS, Higgins JP. The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. Am J Epidemiol;171:645-55.

9. Herder C, Rathmann W, Strassburger K, et al. Variants of the PPARG, IGF2BP2, CDKAL1, HHEX, and TCF7L2 genes confer risk of type 2 diabetes independently of BMI in the German KORA studies. Horm Metab Res 2008;40:722-6.

10. Hu C, Zhang R, Wang C, et al. PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. PLoS One 2009;4:e7643.

11. Lin Y, Li P, Cai L, et al. Association study of genetic variants in eight genes/loci with type 2 diabetes in a Han Chinese population. BMC Med Genet;11:97.

12. Lyssenko V, Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 2008;359:2220-32.

13. Meigs JB, Manning AK, Fox CS, et al. Genome-wide association with diabetes-related traits in the Framingham Heart Study. BMC Med Genet 2007;8 Suppl 1:S16.

14. Nemoto M, Sasaki T, Deeb SS, Fujimoto WY, Tajima N. Differential effect of PPARgamma2 variants in the development of type 2 diabetes between native Japanese and Japanese Americans. Diabetes Res Clin Pract 2002;57:131-7.

15. Sanghera DK, Ortega L, Han S, et al. 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 2008;9:59.

16. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-5.

17. Webster RJ, Warrington NM, Beilby JP, Frayling TM, Palmer LJ. The longitudinal association of common susceptibility variants for type 2 diabetes and obesity with fasting glucose level and BMI. BMC Med Genet;11:140.

18. Weedon MN, McCarthy MI, Hitman G, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3:e374.

19. Wen J, Ronn T, Olsson A, et al. Investigation of type 2 diabetes risk alleles support CDKN2A/B, CDKAL1, and TCF7L2 as susceptibility genes in a Han Chinese cohort. PLoS One;5:e9153.

20. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 1998;19:61-109.

21. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol 2003;13:S5-17.

22. Liu S, Tinker L, Song Y, et al. A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. Arch Intern Med 2007;167:1676-85.

23. Hao K, Liu S, Niu T. A sparse marker extension tree algorithm for selecting the best set of haplotype tagging single nucleotide polymorphisms. Genet Epidemiol 2005;29:336-52.

24. Hsu YH, Niu T, Song Y, Tinker L, Kuller LH, Liu S. Genetic variants in the UCP2-UCP3 gene cluster and risk of diabetes in the Women's Health Initiative Observational Study. Diabetes 2008;57:1101-7.

25. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.

26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57:289-300.

27. Hansen SK, Nielsen EM, Ek J, et al. Analysis of separate and combined effects of common variation in KCNJ11 and PPARG on risk of type 2 diabetes. J Clin Endocrinol Metab 2005;90:3629-37.

28. He Y, Li C, Amos CI, Xiong M, Ling H, Jin L. Accelerating haplotype-based genome-wide association study using perfect phylogeny and phase-known reference data. PLoS One;6:e22097.

29. Tzeng JY, Lu W, Farmen MW, Liu Y, Sullivan PF. Haplotype-based pharmacogenetic analysis for longitudinal quantitative traits in the presence of dropout. J Biopharm Stat;20:334-50.

30. Heude B, Pelloux V, Forhan A, et al. Association of the Pro12Ala and C1431T variants of PPARgamma and their haplotypes with susceptibility to gestational diabetes. J Clin Endocrinol Metab 2011;96:E1656-60.

31. Lorenz AJ, Hamblin MT, Jannink JL. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS One 2010;5:e14079.

32. Feng T, Zhu X. Genome-wide searching of rare genetic variants in WTCCC data. Hum Genet;128:269-80.

33. Parikh H, Lyssenko V, Groop LC. Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. BMC Med Genomics 2009;2:72.

34. Perry JR, McCarthy MI, Hattersley AT, et al. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. Diabetes 2009;58:1463-7.

35. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet;86:581-91.

36. Small KS, Hedman AK, Grundberg E, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet;43:561-4.

37. Luan J, Browne PO, Harding AH, et al. Evidence for gene-nutrient interaction at the PPARgamma locus. Diabetes 2001;50:686-9.

38. Freedman ML, Monteiro AN, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet;43:513-8.

Chapter 3:

Genome-wide Association Study for Vascular Disease Utilizing a Pathway-based Approach among African and Hispanic Postmenopausal Women

3.1 Introduction

Genome wide association studies (GWAS) have been a commonly used method for disease gene discovery in recent decades. GWAS have also facilitated the detection of biological contributions to complex traits and diseases; for example, the principal pattern appears to be of many loci but with small effects individually on phenotype[1]. Additionally, GWAS findings provide critical information about the role of common genetic variants to disease traits. However, simply investigating single nucleotide polymorphism (SNP) with the trait of interest is likely to be inadequate to dissect the complex genetic architecture of many common diseases[2-11]. GWAS can provide primary genetic information that can be followed up by additional analysis via statistical procedures to accumulate evidence. These procedures may potentially provide helpful information to prioritize the most important group of results before further functional validation can proceed[10,12,13].

Recently, there is an increasing interest in utilizing pathways-based analysis to excavate the breadth of GWAS signals that may be collectively clustered in pathways of significance even though individually they may not reach genome-wide significance[14]. Biological pathways encapsulate molecular and biological processes and can be detected as clusters of genes that are related in a functional manner. In most pathway-based GWAS studies, a pathway usually refers to a group of biologically linked genes found in single or multiple databases[15]. GWAS pathway analysis (GWASPA) allows for assimilating GWAS signals in some known genetic and molecular pathways to investigate whether a cluster of related genes in the same functional pathway are associated conjointly with the specific disease trait of interest[9]. For example, in studies of Crohn's disease[16-19], age-related macular degeneration[20-22], Parkinson's disease[23], amyotrophic lateral sclerosis[24], neuropsychiatric disorders[25-28], and rheumatoid arthritis[29,30],

pathway-based approaches have provided additional informative findings beyond traditional single-SNP analysis in GWAS. More recently, the Wellcome Trust Case Control Consortium (WTCCC) Type 2 Diabetes (T2D) study comprised of 4,862 Caucasian individuals reported that biological pathways are jointly associated with T2D and WNT signaling was the top pathway, filling in some portion of the 'missing heritability'[9,31-33]. However, the literature lacks a comprehensive report of T2D GWASPA study in minority populations, especially among African and Hispanic American women.

To take advantage of these recent developments, we proposed to characterize the SNPs effect using a pathway-based approach and investigate its role in the development of vascular disease (VD) in a prospective cohort of 12,008 African American and Hispanic American women enrolled in the Women's Health Initiative SNP Health Association Resource (WHI-SHARe).

3.2 Research Design and Methods

3.2.1 Study participants

WHI-SHARe participants are women enrolled in the Women's Health Initiative (WHI) whose self-reported ethnicity was African or Hispanic American. Please refer to the Section 2.2.1 in Chapter 2 (P.12-13)[34,35]. The WHI-SHARe cohort included African (n=12,151) and Hispanic American (n=5,469) postmenopausal women. Among the eligible participants with available DNA, 8,515 African Americans and 3,642 Hispanic Americans were randomly chosen.

3.2.2 Vascular disease definition

The incident cases of CVD were classified based on any event of myocardial infarction (MI), stroke, deep vein thrombosis, and pulmonary embolism during the follow-up. The incident cases of T2D were also identified on the basis of those clinical cases that had no prior history of T2D at baseline and diagnosed during the follow-up period. Those women who were free of T2D and CVD were used as controls. We defined VD by diagnosis of cardiovascular (CVD) and/or T2D because the two traits have shared risk factors. Obesity, insulin resistance, endothelial dysfunction, dyslipidemia, proinflammatory and prothrombotic factors are common risk factors for both CVD and T2D. However, etiologic mechanisms underlying these factors are poorly understood[36,37]. We integrated information on genotypes using a pathway-based approach to construct association networks for VD risk.

3.2.3 Genotyping, bioinformatics assessment of pathway databases, and statistical analysis

Genotyping was conducted on the Affymetrix Genome-wide Human 6.0 array (Affymetrix ®, Santa Clara, CA). We adopted a gene-set enrichment analysis (GSEA) provided by the GenGen suite (http://www.openbioinformatics.org/gengen/)[38]. We formatted the genotyping data files using SNP and Gene mappings from the UCSC Genome Browser annotation (genome build: hg18). The human biological pathways were defined using Gene Ontology (GO)[39], BioCarta, and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[40] databases. We utilized the most updated versions downloaded from the GenGen's website at the time of conducting the analysis.

In brief, our GSEA involves the following five main steps. First, we conducted association analyses to generate P values for all 871, 309 SNPs that passed quality control

including relatedness and admixture checking. Second, we mapped specific SNPs to specific genes using those with the highest χ^2 statistics within a 500 kb flanking window on both side of the gene to denote the overall test statistic. Thirdly, we ranked each gene and aggregate them into functional groups (537 and 475 pathways met the criteria of containing between 20 and 200 GWAS-captured genes for the African and Hispanic American women respectively). Fourth, we identified over-representation of significant SNPs in clusters using a Kolmogorov-Smirnov-like running sum statistic with normalization to take into account different gene sizes. Finally, we conducted analyses to detect pathways associated with VD risk[32,38].

3.3 Results

A total of 45 pathways reached a nominal P value < 0.05 among 5,729 African American women; while 13 pathways reached a nominal P value < 0.05 in 2,869 Hispanic Americans. **Table 1** and **2** shows the top 10 pathways associated with incident VD (T2D and/or CVD) for African and Hispanic women, respectively. No pathways were associated with VD after adjusting for multiple comparisons (false discovery rate q value < 0.05) in either African or Hispanic American women.

Among African American women, the top five pathways, ranked using nominal P values, were glycerolipid metabolism (hsa00561), regulation of ion transport (GO0043269), urea cycle and metabolism of amino groups (hsa00220), propanoate metabolism (hsa00640), and autophagy (GO0006914).

For Hispanic group, the top ten pathways were mostly related to biological processes, including positive regulation of mononuclear cell proliferation (GO0032946), positive regulation of lymphocyte activation (GO0051251), regulation of mononuclear cell proliferation

(GO0032944), notch signaling pathway (hsa04330), and regulation of cell migration (GO0030334). We plan to further examine the association between the genes among the top pathways with VD risk in future.

Figure 1 and **2** are quantile-quantile (QQ) plots of the 537 and 475 pathways for both ethnic groups. There were modest deviations away from the null distribution, slightly outside the 95% confidence intervals. This deviation may have been due to residual confounding because we were not able to adjust for covariates, including global ancestry, due to the limitation of the GenGen program, which used chi-squared test statistics and could not handle covariates[9]. We plan to further investigate other analytical options to overcome this limitation. However, we did partially account for population stratification by stratifying the analysis by ethnicity.

3.4 Discussion

The major finding from our study is that vascular diseases genes intersect in several pathways. This is consistent with a recent study that showed manifold T2D-related loci fall in differentiated biological pathways[32]. We still need to replicate this study with another independent study population.

The top pathway (i.e. with the smallest nominal P value) among the African American group was glycerolipid metabolism (hsa00561). Glycerolipid acyltransfereases was known to play a critical part in pathophysiological processes of triglyceride (TAG) metabolism and energy balance. Monoacylglycerol (MGAT2) and diacylglycerol (DGAT1) acyltransferases are important enzymes linked with intestinal triglyceride absorption. These enzymes were shown to be involved in TAG metabolism and whole body energy homeostasis. The study also suggested

that inhibition of these enzymes may offer therapeutic benefits for metabolic diseases including T2D[41].

The PPAR signaling pathway was also among the ten top pathways. According to the KEGG pathway database, PPARs are nuclear hormone receptors being activated by fatty acids and their by-products. PPAR-α, PPAR-β, and PPAR-γ are the three subtypes of PPAR and they show various expression patterns in vertebrates. PPAR-α is mainly involved in the clearance of cellular or circulating lipids through regulating gene expression included in lipid metabolism in both skeletal muscle and liver. PPAR-β plays a role in cell proliferation and lipid oxidation. PPAR-γ disseminates adipocyte differentiation to augment blood glucose uptake[42].

Even though GWASPA is attractive, its analytical methods are still at an early development stage, and additional factors regarding statistics need to be addressed. For example, some biases related to gene-chip coverage, gene and pathway size, linkage disequilibrium (LD) pattern among SNPs, and adjustment for covariates such as population stratification remain to be solved[43]. First, the analytical approach we used (the GenGen program) was to choose a single SNP with the most significant association with VD from each gene among all the SNPs residing in that gene. This method may not be ideal because it does not sufficiently capture the gene effect with a single SNP. Instead, SNPs that are responsible for functionality should be used to represent a gene. Therefore, it would be better if a gene score can be defined by multiple SNPs that can characterize a gene rather than only using a SNP with the most significant statistics with the disease trait[33]. We plan to further investigate this analytical approach. Second, the significance of pathways that contain several genes with a few independent association signals may disappear. This may be part of the reason that our study did not yield any pathways that reach the FDR significance threshold of 0.05. It may be preferable if the pathways with genes of

multiple independent associated SNPs are weighted according to the specific effect sizes. Third, we used all available SNPs to recapitulate information from a pathway. This can affect the power for pathway-based analysis because it includes some SNPs that are unrelated to VD. Instead, using a SNP screening step combining with some dimension reduction techniques may remove some of these irrelevant SNPs and thus improve the signal dilution artifact in a more efficient way[13,44]. Other approaches such as SNP/Variant set Enrichment Analysis have also been proposed to choose various SNPs to characterize each gene using some adaptive truncated product statistic[10,33]. Fourth, when several SNPs are involved in the enrichment score calculation such as in our case, LD patterns among these SNPs may affect the quality of the score. Wang et al. provided a partial solution to control for this bias in the GenGen program by introducing a normalization step in the calculation of the gene-set enrichment scores such as scaling using the mean score estimated from permutation tests[33,38]. With the improvements mentioned above, we may be able to identify biological pathways that are significantly associated with VD risk using the GWASPA.

In spite of gene sets and pathways, the identified variants associated with VD risk in our analysis still explain only a small fraction of the heritability. Whether the so-called "missing heritability" will encompass rare variants, structural variants, epigenetic effects or some other unknown mechanisms remains to be seen[14].

To conclude, using WHI-SHARe GWAS data and a pathway-based analytical approach, our study observed that SNPs associated with vascular disease cluster into multiple biological pathways. We are also in the process of replicating this finding in another independent study population.

Table 1. Top 10 pathways associated with incident vascular disease (T2D and/or CVD) among African American women in the WHI-SHARe cohort (n=5729).

Database	Pathway	Gene Size	Nominal P-value
KEGG	Glycerolipid metabolism (hsa00561)	45	0.001
GO	Regulation of ion transport (GO0043269)	21	0.001
KEGG	Urea cycle and metabolism of amino groups (hsa00220)	21	0.001
KEGG	Propanoate metabolism (hsa00640)	33	0.002
GO	Autophagy (GO0006914)	21	0.002
KEGG	PPAR signaling pathway (hsa03320)	65	0.002
GO	Nuclear transport (GO0051169)	135	0.003
KEGG	Nucleotide excision repair (hsa03420)	38	0.005
GO	Segmentation (GO0035282)	31	0.006
KEGG	Mismatch repair (hsa03430)	22	0.006

*Pathways are ranked by nominal p-value.

Table 2. Top 10 pathways associated with incident vascular disease (T2D and/or CVD) among Hispanic American women in the WHI-SHARe cohort. (n=2869)

Database	Pathway	Gene Size	Nominal P-value
GO	Positive regulation of mononuclear cell proliferation (GO0032946)	31	0.008
GO	Positive regulation of lymphocyte activation (GO0051251)	59	0.01
GO	Regulation of mononuclear cell proliferation (GO0032944)	41	0.02
KEGG	Notch signaling pathway (hsa04330)	35	0.02
GO	Regulation of cell migration (GO0030334)	40	0.02
BioCarta	Cell cycle: g2/m checkpoint (g2Pathway)	21	0.02
GO	Calcium-dependent cell-cell adhesion (GO0016339)	24	0.02
GO	Limb development (GO0060173)	42	0.02
GO	Phospholipid transport (GO0015914)	21	0.04
GO	Oxidoreductase activity (GO0016712)	24	0.04

*Pathways are ranked by nominal p-value.

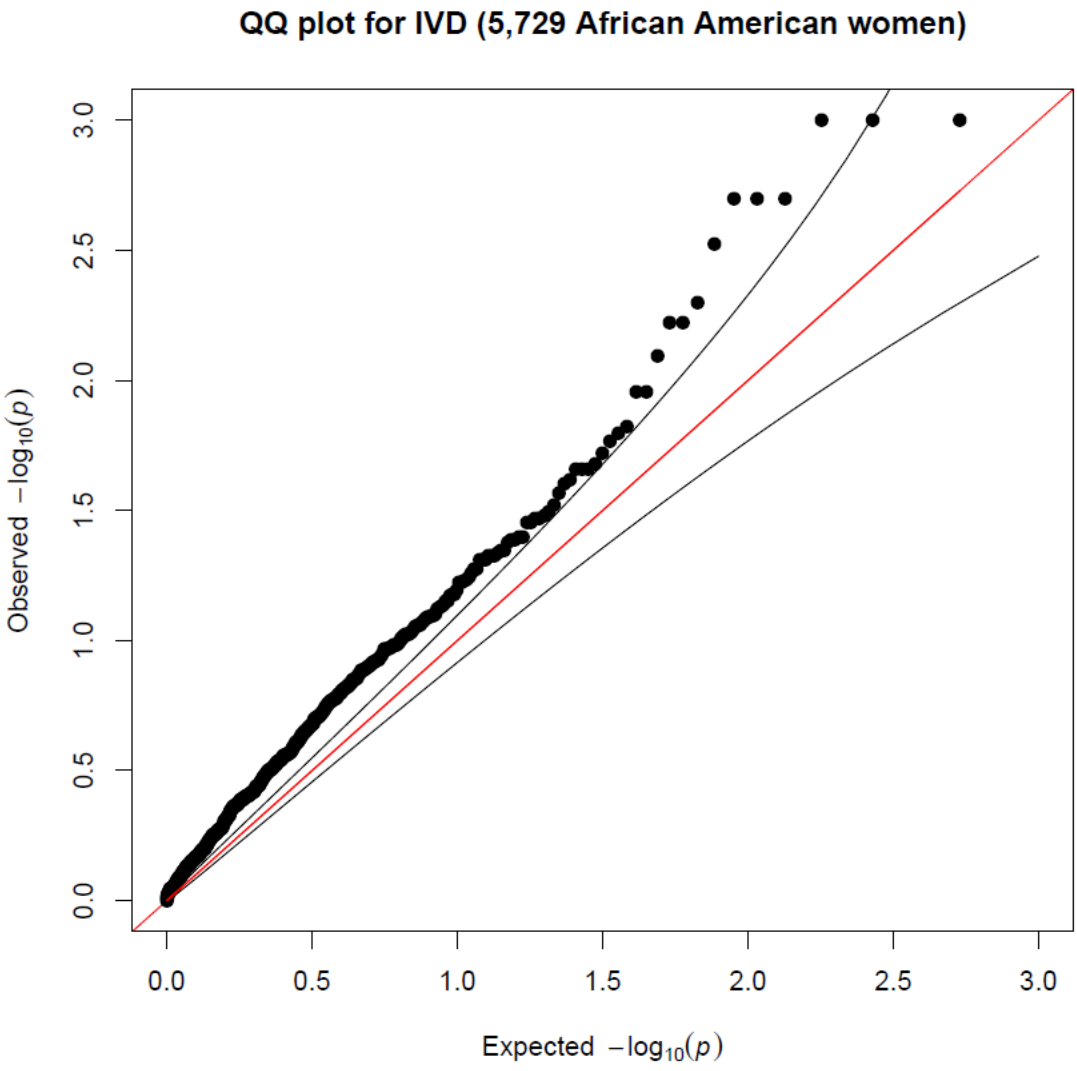


Figure 1. Q-Q plot that shows the P values based on the WHI-SHARe African American women data. Dashed lines denote 95% confidence interval.

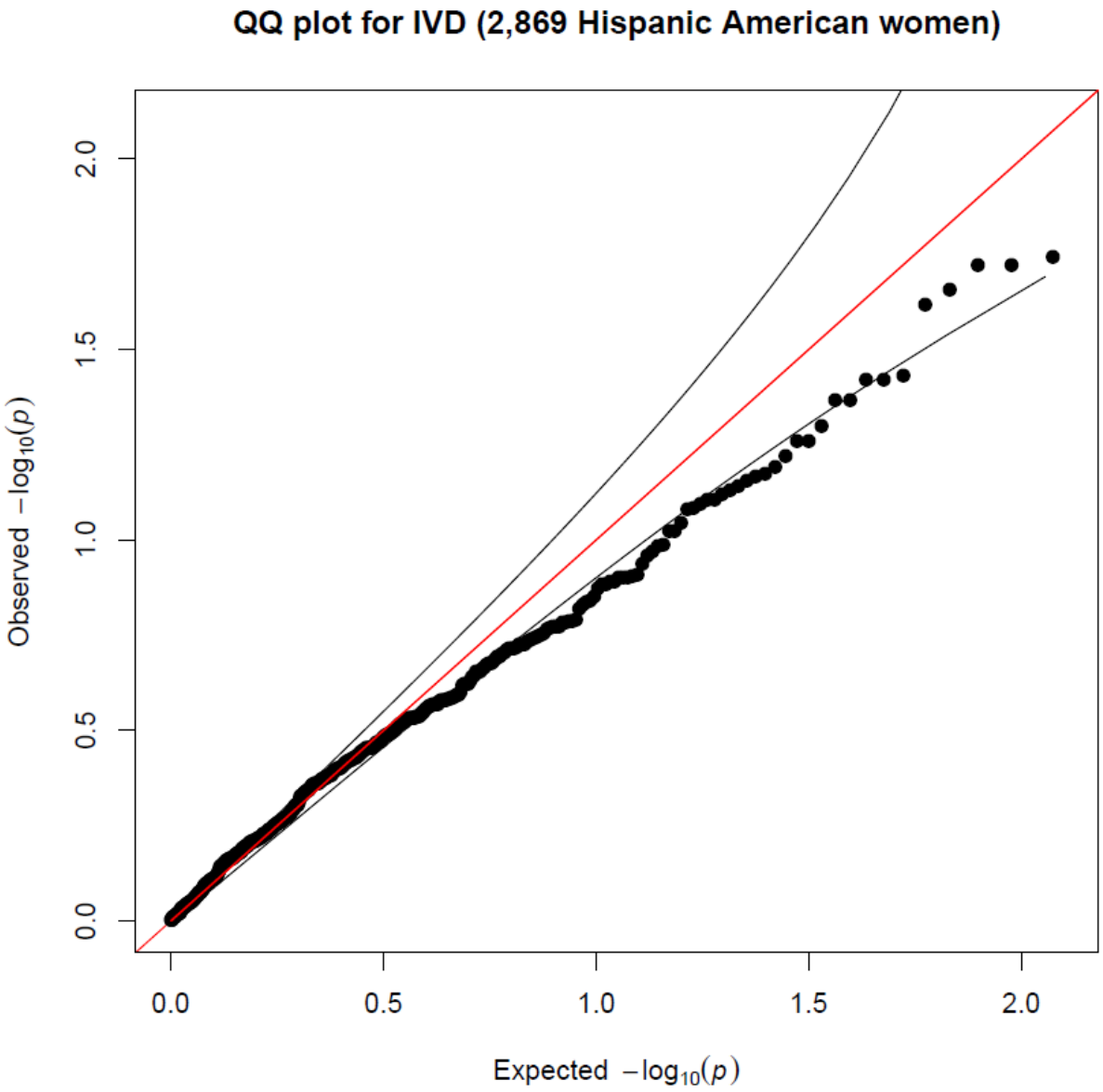


Figure 2. Q-Q plot that shows the P values based on the WHI-SHARe Hispanic women data. Dashed lines denote 95% confidence interval.

3.6 References

1. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 2011;187:367-83.

2. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631-7.

3. Peng G, Luo L, Siu H, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. Eur J Hum Genet 2010;18:111-7.

4. Salonen JT, Uimari P, Aalto JM, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. Am J Hum Genet 2007;81:338-45.

5. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-6.

6. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-5.

7. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881-5.

8. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770-5.

9. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 2010;11:843-54.

10. Weng L, Macciardi F, Subramanian A, et al. SNP-based pathway enrichment analysis for genome-wide association studies. BMC Bioinformatics;12:99.

11. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336-41.

12. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet 2010;86:6-22.

13. Chen X, Wang L, Hu B, Guo M, Barnard J, Zhu X. Pathway-based analysis for genome-wide association studies using supervised principal components. Genet Epidemiol 2010;34:716-24.

14. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health 2010;31:21-36.

15. Yaspan BL, Veatch OJ. Strategies for pathway analysis from GWAS data. Curr Protoc Hum Genet;Chapter 1:Unit1 20.

16. Abraham C, Cho J. Interleukin-23/Th17 pathways and inflammatory bowel disease. Inflamm Bowel Dis 2009;15:1090-100.

17. Abraham C, Cho JH. IL-23 and autoimmunity: new insights into the pathogenesis of inflammatory bowel disease. Annu Rev Med 2009;60:97-110.

18. Dong C. TH17 cells in development: an updated view of their molecular identity and genetic programming. Nat Rev Immunol 2008;8:337-48.

19. Yoshida H, Nakaya M, Miyazaki Y. Interleukin 27: a double-edged sword for offense and defense. J Leukoc Biol 2009;86:1295-303.

20. Edwards AO, Ritter R, 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science 2005;308:421-4.

21. Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science 2005;308:419-21.

22. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385-9.

23. Lesnick TG, Papapetropoulos S, Mash DC, et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS Genet 2007;3:e98.

24. Lesnick TG, Sorenson EJ, Ahlskog JE, et al. Beyond Parkinson disease: amyotrophic lateral sclerosis and the axon guidance pathway. PLoS One 2008;3:e1449.

25. Askland K, Read C, Moore J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. Hum Genet 2009;125:63-79.

26. Holmans P, Green EK, Pahwa JS, et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 2009;85:13-24.

27. O'Dushlaine C, Kenny E, Heron E, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Mol Psychiatry 2011;16:286-92.

28. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009;459:528-33.

29. Ballard DH, Aporntewan C, Lee JY, Lee JS, Wu Z, Zhao H. A pathway analysis applied to Genetic Analysis Workshop 16 genome-wide rheumatoid arthritis data. BMC Proc 2009;3 Suppl 7:S91.

30. Tintle NL, Borchers B, Brown M, Bekmetjev A. Comparing gene set analysis methods on single-nucleotide polymorphism data from Genetic Analysis Workshop 16. BMC Proc 2009;3 Suppl 7:S96.

31. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.

32. Perry JR, McCarthy MI, Hattersley AT, et al. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. Diabetes 2009;58:1463-7.

33. Yang W, de las Fuentes L, Davila-Roman VG, Charles Gu C. Variable set enrichment analysis in genome-wide association studies. Eur J Hum Genet;19:893-900.

34. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials 1998;19:61-109.

35. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol 2003;13:S5-17.

36. Manolio TA. Cohort studies and the genetics of complex disease. Nat Genet 2009;41:5-6.

37. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 2009;41:35-46.

38. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 2007;81:1278-83.

39. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.

40. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.

41. Cao G, Konrad RJ, Li SD, Hammond C. Glycerolipid acyltransferases in triglyceride metabolism and energy homeostasis-potential as drug targets. Endocr Metab Immune Disord Drug Targets 2012.

42. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109-14.

43. Yaspan BL, Bush WS, Torstenson ES, et al. Genetic analysis of biological pathway data through genomic randomization. Hum Genet 2011;129:563-71.

44. Zhao J, Gupta S, Seielstad M, Liu J, Thalamuthu A. Pathway-based analysis using reduced gene subsets in genome-wide association studies. BMC Bioinformatics 2011;12:17.

Chapter 4:

Genome-wide Association Study for Body Mass Index and Glycated Hemoglobin Levels in the Framingham Heart Study 500K Project

4.1 Introduction

Obesity is a well-known predictor for metabolic diseases such as diabetes. Obesity may influence diabetes through pleiotropic genetic effects of obesity loci such as the well known *FTO* and *TMEM18*, or as an environmental effect by impairing insulin resistance and beta-cell dysfunction through secretion of free fatty acids and adipokines[1]. Apart from diet and lifestyle determinants, most of the population variation in body mass index (BMI), a commonly used measure of obesity, is governed by genetic factors[2-4]. Several loci such as *INSIG2*, *FTO*, *MC4R*, *TMEM18*, *GNPDA2*, *BDNF*, *NEGR1*, *SH2B1*, *ETV5*, *MTCH2*, *KCTD15*, *LCT*, and *A2BP1,* have been shown to be associated with BMI in multiple populations using a genome-wide approach[3-9]. However, some of these genes were not replicated in other independent studies.

Glycated hemoglobin (HbA1C) results from the nonenzymatic glycation of hemoglobin molecules. Due to the fact that the glycation process is irreversible and is directly proportional to intracellular glucose concentrations, HbA1C has become an index of the mean glycemic measure over the average life span of erythrocytes. Therefore, the American Diabetes Association has proposed HbA1C as a diagnostic criterion for diabetes[10,11]. HbA1C levels have also been reported to be a strong risk factor of diabetes development[12,13]. Possibly through the modulation of hematologic parameters or blood glucose, HbA1C levels are likely to be genetically determined because the heritability of HbA1C levels are comparatively high (47-59%)[14,15]. In a previous study, *GCK*, *SLC30A8*, *HK1* and *G6PC2* were the genetic loci identified to be associated with HbA1C levels in non-diabetic individuals. *CAPN10*, *RETN*, *ADIPOQ*, TCF7L2, *SORCS1*, *FN3K, HFE*, *TMPRSS6*, *ATP11A/TUBGCP3*, *ANK1*, *SPTA1*, *BNC1*, *GSC*, and *WDR72* have also been described with glycated hemoglobin in candidate gene studies, genome-wide

association studies (GWAS) and meta-analysis of GWAS[11-13,15-17]. However, relatively little is known about the role of genetic variations in the regulation of glucose concentrations.

In this study, we utilized data from the Framingham Heart Study (FHS) 500K Project to further investigate the association of two diabetes-related quantitative traits, i.e. BMI and HbA1C levels, using a new analytic method called Pedigree-based GWAS (unpublished) implemented in the Mendel software package[18]. We also planned to validate the SNPs among several genes that were known to be related to BMI in previous studies.

4.2 Research Design and Methods

4.2.1 Study Populations

The FHS is a joint project of the National Heart, Lung and Blood Institute and Boston University. In 1948, the researchers recruited 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, and began the first round of extensive physical examinations and lifestyle interviews that they would later analyze for common patterns related to CVD development. Since 1948, the subjects have returned to the study every two years for an examination consisting of a detailed medical history, physical examination, and laboratory tests. In 1971, the study enrolled a second generation cohort, which consists of 5,124 of the original participants' adult children and their spouses, to participate in similar examinations. The second examination of the Offspring cohort occurred eight years after the first examination, and subsequent examinations have occurred approximately every four years thereafter. In April 2002 the Study entered a new phase: the enrollment of a third generation of participants, the grandchildren of the original cohort. The first examination of the third generation study was

completed in July 2005 and involved 4,095 participants. Thus, the FHS has evolved into a prospective, community-based, three-generation family study.

Participants from these three generations underwent genotyping with the Affymetrix 500K GeneChip. Participants (n =6,852) from the three generations were genotyped for the Affymetrix GeneChip Human Mapping 500K SNP set. As shown in **Table 1**, there were 357 participants from the original cohort, 2,584 participants from the second generation, 96 spouses of second generation including men and women, and 3,815 participants from the third generation. A total of 6,752 white participants (3,072 males and 3,680 females) had genotyping and BMI values available; whereas 2,533 participants, with 1,099 males and 1,434 females, had genotyping and HbA1C values available.

4.2.2 Mean BMI and HbA1C levels definition

Body weight and height were measured from examination cycle 1 to examination cycle 28 for the original cohort, examination cycle 1 to 8 for the second generation cohort, and examination 1 in the third generation cohort. BMI was then calculated using this equation: {Weight in pound/[(height in inch) x (height in inch)]} x 703. Mean BMI was then calculated by taking average of all available examination cycles. Among 6,752 participants without missing mean BMI values, the average of the mean BMI was 26.8 (range: 15.6-60.5) with a standard deviation of 5.07, as shown in **Table 1**.

HbA1C levels were measured in examination cycles 19/20 and 22 in the original cohort and examination cycle 7 in the second generation cohort. The mean HbA1C levels were obtained by taking the average of all available measurements. Among 2,533 participants without missing mean HbA1C levels, the average of the mean HbA1C levels was 5.6 (range: 3.58-13.2) with a

standard deviation of 0.88, as shown in **Table 1**. Age was also averaged over the examination cycles.

4.2.3. Genotyping and Quality Control

Framingham participants were genotyped using the Affymetrix (Santa Clara, CA) GeneChip Human Mapping 500K Array Set[19], which produced approximately 500,568 SNPs. 370,563 SNPs remained after using a genotyping success rate per SNP of 0.90 and minor allele frequencies for founder > 0.05. 6,748 and 2,529 individuals remained for the BMI and HbA1C traits respectively after using a genotyping success rate per person of 0.90.

4.2.4 Statistical Analysis

The association analyses for both BMI and HbA1C traits were performed using the pedigree-based GWAS option in Mendel 12 beta version (unpublished). This new method utilizes score tests instead of likelihood ratio tests to overcome the current computational bottleneck for dense marker mapping in pedigrees. Statistical models were adjusted for gender and age. We did not adjust for population stratification because our study participants are all whites. To account for multiple comparisons, we applied a Bonferroni threshold of 1.35E-07 (i.e. 0.05/370563) and a false discovery rate (FDR) q value threshold of 0.05.

To validate this new analytical method, we tested the association between SNPs from six previously published genes (*INSIG1, INSIG2, PPARG, ADIPOQ, ESR1* and *LEP)* that showed association with BMI[6] using this new method. We first extracted 494 SNPs among these six genes and investigated their association with the BMI trait.

4.3 Results

4.3.1 GWAS on mean BMI

Table 2 presents the top fifteen signals for the BMI trait, which includes relevant examination cycles and adjustment for gender and age. The top SNP rs17627690 is 3.02 kb upstream from the polycystic kidney disease 1 like 1 (*PKD1L1*) gene. From Refseq, this gene encodes a member of the polycystin protein family that contains 11 transmembrane domains, a receptor for egg jelly (REJ) domain, and a polycystin-1, lipoxygenase, alpha-toxin (PLAT) domain. The encoded protein may be involved in the male reproductive system. Alternative splice variants have been described, however their biological nature has not been determined. Among these signals, three SNPs (rs13373826, rs11163494, and rs211787) are located in a known gene, solute carrier family 44member 5 gene (*SLC44A5*), which is also known as *CTL5*. This gene is protein coding, however, not much has been published about this gene in the literature. **Figure 1** shows the manhattan plot for all 370,563 SNPs that passed the quality control criteria. By employing FDR, the top SNP also reached a significant level of q value <0.05.

We also ran the same analysis on BMI using the latest available measurements with the last observable age as covariate. The top 15 signals are shown in **Table 3** and **Figure 2**. A top hit passed the Bonferroni threshold with a P value of 1.09E-07. However, this genetic variant has no known function. The other results are similar to what was obtained using the average BMI values, except that one top SNP (rs17629371) located in the EGF-like repeats and discoidin I-like domains 3 (*EDIL3*) genomic region. The protein encoded by this gene is known to be an integrin ligand, which plays a role in angiogenesis. A deleted locus of this gene was also found in childhood obesity cases[20]. The top six SNPs passed the FDR threshold of q value < 0.05.

4.3.2 GWAS on mean HbA1C levels

The 15 top signals for the HbA1C trait after adjusting for gender and age are presented in **Table 4**. One of the top SNPs rs7540760, residing in the *LAX1* gene (also known as *LAX* gene), lied in the promoter or regulatory region. The *LAX1* gene is known to be protein coding. The manhattan plot for the mean HbA1C levels is shown in **Figure 3.**

4.3.3 Validation results for known genes associated with BMI

In addition, we identified 494 SNPs in six genes (*INSIG1, INSIG2, PPARG, ADIPOQ, ESR1* and *LEP)* that were previously shown to have association with BMI. Except for the *ADIPOQ* gene, 30 SNPs residing in or near to the *INSIG1, INSIG2, PPARG, ESR1* and *LEP* genomic regions passed the nominal P value threshold of 0.05. **Table 5** shows the association estimates and other relevant information for these 30 SNPs. We confirmed the association between five of the six genes that have previously been identified in the obesity field[6] using the new analytical method, pedigree-based GWAS, implemented in Mendel.

4.4 Discussion

In our analysis of the two diabetes-related quantitative traits, we found strong associations between SNPs near the *LOC100507205* locus and BMI trait among 6,752 participants from the original, the second generation, and the third generation cohorts in the Framingham Heart Study. We also replicated five well-validated genes that have been previously reported to be significantly associated with the BMI trait.

We also examined the genetic association with the most recent observation of BMI and adjusted for the most recent age observation (instead of means). Using a FDR q value threshold

of 0.05, we observed that six SNPs inside one single gene (*LOC100507205*) were significantly associated with BMI, a finding that is worth pursuing. We will look into it further for this potential significant finding.

One of the top SNPs for the HbA1C trait resides in the hedgehog acyltransferase (*HHAT*) gene. From RefSeq, this gene encodes an enzyme, which acts within the secretory pathway to catalyze amino-terminal palmitoylation of 'hedgehog' (OMIM, 2002). The other top SNP rs1081487 lies in the GLIS family zinc finger 3 (*GLIS3*) gene, which encodes a nuclear protein with five C2H2-type finger domains. Its protein functions as both an activator and repressor of transcription and is involved in the development of pancreatic beta cells, thyroid, kidney, liver and eye. Mutations in this gene have been associated with neonatal diabetes[21,22]. Another top SNP rs1869699 is in the genomic region of ribosomal protein S6 kinase, 90k Da, polypeptide (*RPS6KA2*). This gene encodes a member of the ribosomal S6 kinase family of serine/threonine kinases. The activity of this protein has been associated with control of cell growth and differentiation. Genetic variants in this gene have been previously associated with risk of rectal cancer[23]. The associations with HbA1C levels might probably reach genome-wide significance if more participants had available HbA1C measurements.

New genetic loci have been identified in previous studies for the BMI and HbA1C traits[3-9,11-13,15-17]. However, it is perplexing that some associations appear to replicate in one independent cohort but not the other[10], although this may be due to the different genetic coverage between studies. For example, Patterson et al. discovered the rs1358030 in the *SORCS1* gene, but this genetic variant was not replicated in the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC). The question whether this genetic variant actually represents the causal SNP or is simply linked with an ungenotyped causal variant in the region warrants further

detailed fine-mapping and functional studies[13]. The same procedure should also be conducted for all other "novel" genetic signals detected from recent GWAS, particularly for both the BMI and HbA1C traits.

With the availability of large collections of linkage data such as the Framingham Heart Study, the use of family-based design has become feasible[1,24-27]. Using families is an improved strategy for both genotype and phenotype data due to the reasons that will be discussed in the following. Family data can also offer other related information, for example, estimates of locus-specific heritability and candidate regions based on linkage analysis[26,28]. Further, family-based GWAS presented several advantages over association testing among unrelated individuals. On one hand, family-based GWAS provides better genotype quality control because it allows genotyping errors to be identified by taking into account inconsistencies between a parent and his/ her child/children's genotype, which estimates genotyping error rate in a direct manner. On the other hand, family-based designs are more robust to population stratification[24,26]. Additionally, family-based designs provide different genetic analyses that cannot be conducted using a sample of unrelated individuals such as testing the effect of imprinted genes on phenotypes[29].

To date, several genome-wide linkage or association studies have adopted the pedigree structure of study subjects in the analytical approaches[5,6,30,31] such as what was used in this study. Apart from the family-based genome-wide linkage analysis using the LOD scores, generalized estimating equations (GEE) and family-based association testing (FBAT) are two commonly used methods[32]. New improvements have been made to the family-based analytical design in the genome-wide scan setting. First, Naylor et al. has adopted a Bayesian approach in family-based GWAS. This method constructed a Bayes factor that was conditional on the parental genotype

and offspring phenotype data to inform the prior odds for each genetic marker. Association testing for each marker was obtained by assessing its genetic effect size through fitting the conditional mean model[33]. Second, two-stage testing strategies have been proposed in family-based GWAS[34,35] such as using the Van Steen algorithm implemented in a PBAT software package[34]. The first (screening) step included assessing the association evidence at a population-based level. The second step then prioritized the SNPs based on the results from the screening step for testing. This strategy claimed to achieve the statistical power level of population-based studies. To date, no methodological approach has adopted score tests in place of likelihood ratio tests to map dense markers in pedigrees like the pedigree-based GWAS option in Mendel. This score-test method reduced the arithmetic calculation in terms of the mean and variance component of the statistical model. This method provided several advantages: 1) it works for pedigree data, random sample data, or a mixture of both; 2) it allows correction for population stratification and covariates adjustment; 3) it accommodates both univariate and multivariate traits; and 4) it fosters both score and likelihood ratio tests.

To conclude, we found strong associations between SNPs near the *LOC100507205* locus and BMI among participants from the original, the second generation, and the third generation cohorts in the Framingham Heart Study. We also replicated five well-validated genes that have been previously reported to be significantly associated with the BMI trait. HbA1C levels are potentially associated with SNPs on the Affymetrix 500K SNP GeneChip. These data can serve as a replication resource as more genes become detected with BMI and HbA1C levels. Further studies in other populations are warranted to investigate the association between these possible genetic variants with both BMI and HbA1C levels.

Table 1. Characteristics of the three-generation FHS participants (n=6,842)[a].

	Original Cohort	Second Generation	Spouses	Third Generation	All
N	305	2,412	96	3,812	6,852
Gender					
Male	30.5	45.6	46.9	46.7	45.5
Female	69.5	54.4	53.1	53.3	54.5
Age					
Mean measurement	305 (61.7±3.73)	2,412 (53.0±9.17)	96 (65.1±9.78)	3,812 (41.6±9.09)	6,852 (47.3±11.2)
Most recent measurement	305 (88.4±3.68)	2,412 (66.6±9.25)	96 (65.1±9.78)	3,812 (40.2±8.87)	6,852 (53.0±17.5)
BMI (kg/m^2)					
Mean measurement	305 (25.8±3.56)	2,412 (26.8±4.48)	---[b]	3,808 (26.9±5.56)	6,752 (26.8±5.07)
Most recent measurement	305 (21.6±7.45)	2,412 (28.3±5.42)	---[b]	3,808 (26.9±5.56)	6,752 (27.1±5.80)
HbA1C (%)	286 (5.56±0.66)	2,055 (5.60±0.91)	---[b]	---[b]	2,533 (5.60±0.88)
Menopause					
Yes	---[b]	42.8	---[b]	52.4	45.4
No	---[b]	11.6	---[b]	21.2	16.1
Missing or not relevant	---[b]	45.6	---[b]	26.4	38.5
Current Smoking					
Yes	5.25	9.45	---[b]	16.7	13.2
No	94.8	90.5	---[b]	83.3	85.3
Missing	0	0	---[b]	0.18	1.50

[a]:Data are n (means ± SD) or %.
[b]:Data are not available.

Table 2. Top 15 association results for mean BMI with adjustment for gender and age. (n=6,748)

SNP	Chromosome	Physical Position (bp)[a]	Regression estimate (95%CI)	P value	Gene (Gene Alias)
rs17627690	7	47777759	-0.70 (-0.95 - 0.44)	1.38E-07*	3.02 kb from *PKD1L1*
rs1601333	11	41771130	0.45 (0.27 - 0.63)	1.28E-06	395 kb from *LOC100507205*
rs404089	11	41794057	0.44 (0.26 - 0.62)	1.88E-06	---[b]
rs5016183	11	41789771	0.44 (0.26 - 0.62)	2.28E-06	376 kb from *LOC100507205*
rs4533028	11	41773933	0.43 (0.25 - 0.61)	3.11E-06	392 kb from *LOC100507205*
rs437023	11	41794483	0.43 (0.25 - 0.61)	3.47E-06	371 kb from *LOC100507205*
rs1843246	11	41771426	0.43 (0.25 - 0.61)	3.50E-06	394 kb from *LOC100507205*
rs13373826	1	75743383	-0.56 (-0.81 - -0.32)	6.63E-06	*SLC44A5* (*CTL5*)
rs12807116	11	41790350	0.42 (0.23 - 0.60)	7.74E-06	376 kb from *LOC100507205*
rs2862384	11	41788094	0.42 (0.23 - 0.60)	8.50E-06	---[b]
rs435075	11	41794856	0.40 (0.22 - 0.57)	9.95E-06	371 kb from *LOC100507205*
rs391960	11	41792409	0.41 (0.23 - 0.59)	1.07E-05	373 kb from *LOC100507205*
rs11163494	1	75731941	-0.55 (-0.79 - 0.31)	1.07E-05	*SLC44A5* (*CTL5*)
rs410148	11	41793041	0.42 (0.23 - 0.61)	1.19E-05	373 kb from *LOC100507205*
rs211787	1	75765050	-0.54 (-0.78 - -0.29)	1.43E-05	*SLC44A5* (*CTL5*)

[a] Information is based on Genome Build 36.2.
[b] No mapped gene can be found.
* FDR q value < 0.05.

Table 3. Top 15 association results for BMI using the latest available measurement with adjustment for gender and age. (n=6,748)

SNP	Chromosome	Physical Position (bp)[a]	Regression estimate (95%CI)	P value	Gene (Gene Alias)
rs1601333	11	41771130	0.56 (0.36 – 0.77)	1.09E-07*	395 kb from *LOC100507205*
rs4533028	11	41773933	0.55 (0.35 – 0.76)	1.85E-07*	392 kb from *LOC100507205*
rs1843246	11	41771426	0.55 (0.34 – 0.75)	2.56E-07*	394 kb from *LOC100507205*
rs404089	11	41794057	0.55 (0.34 – 0.75)	2.87E-07*	372 kb from *LOC100507205*
rs5016183	11	41789771	0.54 (0.33 – 0.75)	4.44E-07*	376 kb from *LOC100507205*
rs437023	11	41794483	0.53 (0.32 – 0.74)	6.54E-07*	371 kb from *LOC100507205*
rs2862384	11	41788094	0.52 (0.31 – 0.74)	1.29E-06	378 kb from *LOC100507205*
rs13373826	1	75743383	-0.70 (-0.98 - -0.41)	1.31E-06	*SLC44A5*
rs17627690	7	47777759	-0.74 (-1.04 - -0.44)	1.32E-06	3.02 kb from *PKD1L1*
rs12807116	11	41790350	0.52 (0.31 – 0.73)	1.33E-06	376 kb from *LOC100507205*
rs1412715	1	75758255	-0.68 (-0.96 - -0.40)	2.07E-06	*SLC44A5*
rs11163494	1	75731941	-0.68 (-0.96 - -0.40)	2.31E-06	*SLC44A5*
rs17629371	5	83394558	0.56 (0.33 – 0.80)	3.10E-06	*EDIL3*
rs391960	11	41792409	0.50 (0.29 – 0.70)	3.17E-06	373 kb from *LOC100507205*
rs211787	1	75765050	-0.66 (-0.94 - -0.38)	3.60E-06	*SLC44A5*

[a] Information is based on Genome Build 36.2.

[b] No mapped gene can be found.

* FDR q value < 0.05.

Table 4. Top 15 association results[a] for mean HbA1C levels with adjustment for gender and age. (n=2,529n=)

SNP	Chromosome	Physical Position (bp)[b]	Regression estimate (95%CI)	P value	Gene
rs6128254	20	56030413	-0.12 (-0.17 - -0.06)	1.14E-05	127 kb from *MIR4532*
rs11061417	12	130314013	-0.14 (-0.20 - -0.08)	1.18E-05	50.6 kb from *LOC116437*
rs7540760	1	202001608	-0.13 (-0.19 - -0.07)	1.18E-05	*LAX1*
rs2182766	1	81569056	-0.24 (-0.35 - -0.14)	1.27E-05	470 kb from *LPHN2*
rs4504959	1	208868502	-0.18 (-0.26 - -0.10)	1.36E-05	*HHAT*
rs10814874	9	4184762	0.11 (0.06 – 0.16)	1.49E-05	*GLIS3*
rs847386	7	16941578	-0.10 (-0.15 - -0.06)	2.31E-05	53.4 kb from *AGR3*
rs7543636	1	81602058	-0.20 (-0.30 - -0.11)	2.80E-05	437 kb from *LPHN2*
rs7662934	4	13375175	-0.10 (-0.15 - -0.05)	2.87E-05	137kb from *BOD1L*
rs10847096	12	125189182	-0.13 (-0.19 - -0.07)	3.01E-05	---[b]
rs2242575	6	167159750	-0.15 (-0.22 - -0.08)	4.79E-05	*RPS6KA2*
rs1869699	11	11687578	-0.15 (-0.22 - -0.08)	4.81E-05	52.7 kb from *MIR4299*
rs9355742	6	159997163	-0.11 (-0.17 - -0.06)	5.13E-05	23.0 kb from *SOD2*
rs927030	20	20844942	-0.15 (-0.23 - -0.08)	6.26E-05	204 kb from *RALGAPA2*
rs699618	12	61665150	-0.16 (-0.24 - -0.08)	7.18E-05	50.2 kb from *PPM1H*

[a] Information is based on Genome Build 36.2.

[b] No mapped gene can be found.

Table 5. Associations[a] of mean BMI with all SNPs in or near genes (up to 500kb away) for a well replicated genes (*INSIG1, INSIG2, PPARG, ADIPOQ, ESR1* and *LEP*)[b] in the published literature with a P value < 0.05. (n=6,750)

Gene	SNP	Chromosome	Physical Position (bp)[c]	Regression estimate (95%CI)	P value
ESR1	rs851981	6	152068767	-0.42 (-0.62 - -0.22)	3.18E-05
ESR1	rs851983	6	152066108	-0.33 (-0.50 - -0.16)	0.0002
ESR1	rs7772579	6	152084195	-0.35 (-0.54 - -0.16)	0.0004
ESR1	rs2982571	6	152054432	-0.30 (-0.47 - -0.13)	0.0006
ESR1	rs2982554	6	152099703	0.29 (0.12 – 0.46)	0.0007
ESR1	rs3020348	6	152099607	0.29 (0.12 – 0.45)	0.0007
ESR1	rs2982565	6	152093547	-0.39 (-0.61 - -0.16)	0.0008
ESR1	rs2982561	6	152094345	0.28 (0.12 – 0.45)	0.0008
1.10 kb from ESR1	rs2982573	6	152052227	-0.29 (-0.46 - -0.12)	0.0009
ESR1	rs3020306	6	152107579	0.28 (0.11 – 0.45)	0.001
ESR1	rs1856057	6	152109562	0.26 (0.09 – 0.43)	0.003
ESR1	rs851985	6	152062083	-0.26 (-0.44 - -0.09)	0.003
ESR1	rs6899458	6	152093736	-0.31 (-0.51 - -0.11)	0.003
LEP	rs10244329	7	127675925	0.24 (0.08 – 0.41)	0.004
383 kb from INSIG2	rs7570731	2	118967367	-0.38 (-0.65 - -0.12)	0.005
351 kb from INSIG2	rs34271671	2	118935045	-0.29 (-0.50 - -0.09)	0.006
30.8 kb from PPARG	rs11917039	3	12481675	-0.30 (-0.52 - -0.09)	0.006
7.88 kb from LEP	rs10954175	7	127692793	0.22 (0.05 – 0.39)	0.01
350 kb from INSIG2	rs41515249	2	118934464	-0.29 (-0.52 - -0.06)	0.01
ESR1	rs2347759	6	152137153	0.26 (0.05 – 0.47)	0.01
413 kb from INSIG2	rs17825729	2	118996974	-0.35 (-0.63 - -0.07)	0.01
ESR1	rs9371552	6	152132228	0.29 (0.06 – 0.52)	0.01
352 kb from INSIG2	rs41388348	2	118936266	-0.26 (-0.46 - -0.05)	0.01
ESR1	rs6916835	6	152131717	0.25 (0.04 – 0.46)	0.02
ESR1	rs6939257	6	152131738	0.25 (0.04 – 0.46)	0.02
1.45 kb from LEP	rs2060715	7	127686365	0.20 (0.03 – 0.37)	0.02
ESR1	rs6930355	6	152413084	-0.30 (-0.57 - -0.03)	0.03
1.12 kb from LEP	rs12537573	7	127686036	0.22 (0.02 – 0.43)	0.03
419 kb from INSIG2	rs17008564	2	119002661	-0.29 (-0.56 - -0.02)	0.04
ESR1	rs3778089	6	152435454	-0.27 (-0.53 - -0.001)	0.05

[a] Models are adjusted for sex and age.
[b] No SNPs in or near *ADIPOQ* gene had a P value < 0.05.
[c] Information is based on Genome Build 36.2.

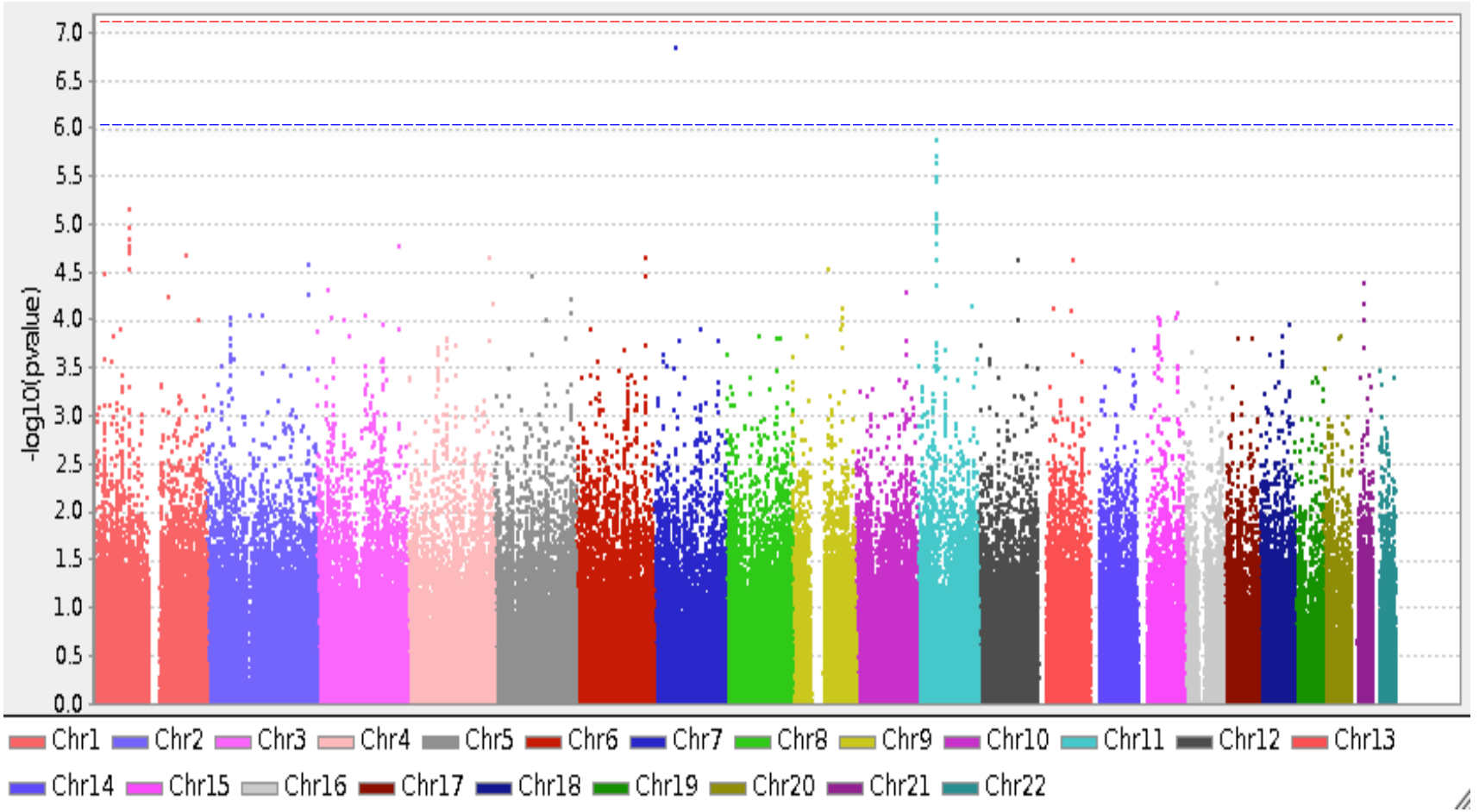


Figure 1. Genome-wide association study to mean BMI in the Framingham Heart Study 500K Project with adjustment for gender and age. (n=6,748) Red dashed line indicates the Bonferroni threshold, whereas blue dashed line denotes FDR threshold (q value < 0.05).

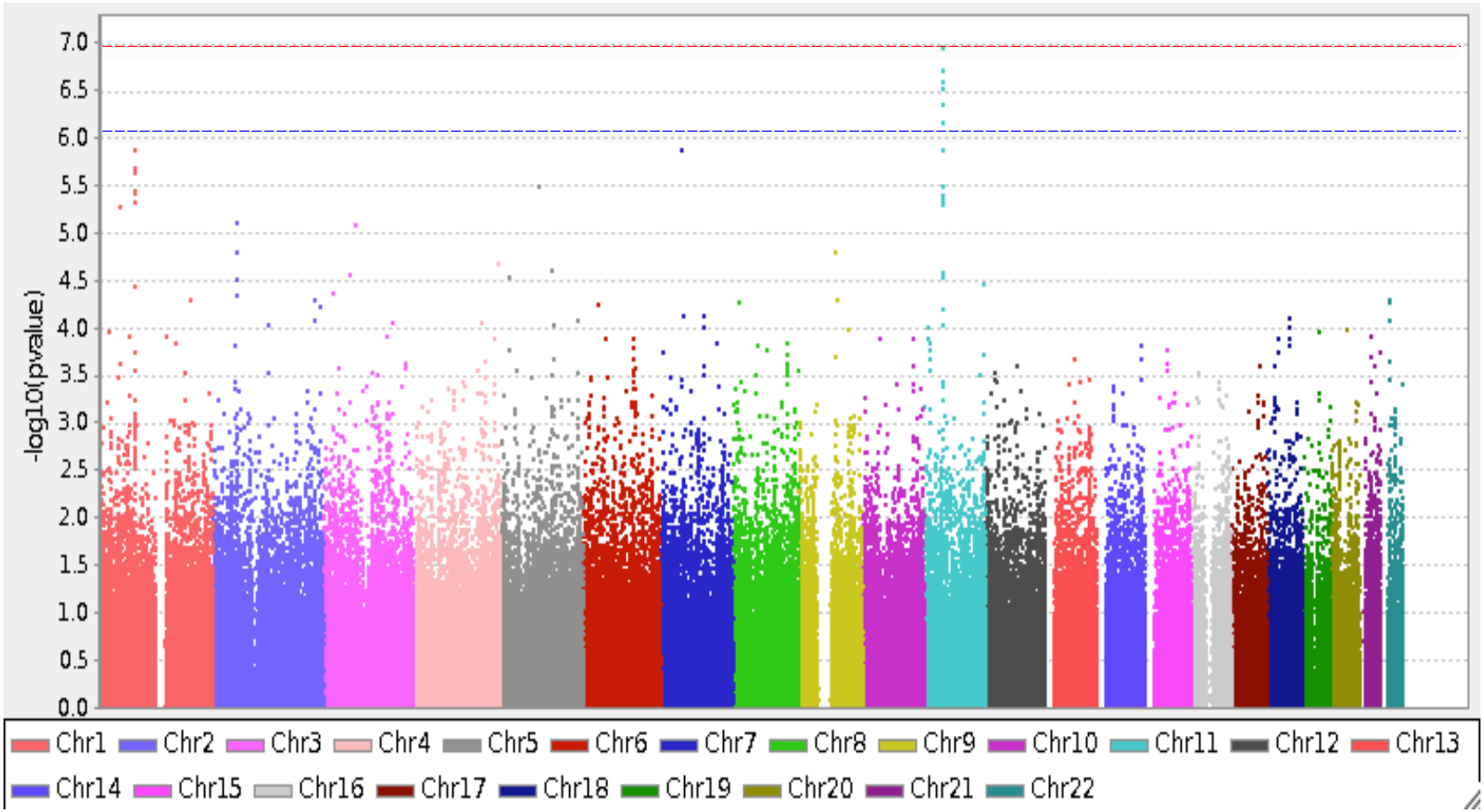


Figure 2. Genome-wide association study to BMI using the latest measurement in the Framingham Heart Study 500K Project with adjustment for gender and age. (n=6,748) Red dashed line indicates the Bonferroni threshold, whereas blue dashed line denotes FDR threshold (q value < 0.05).

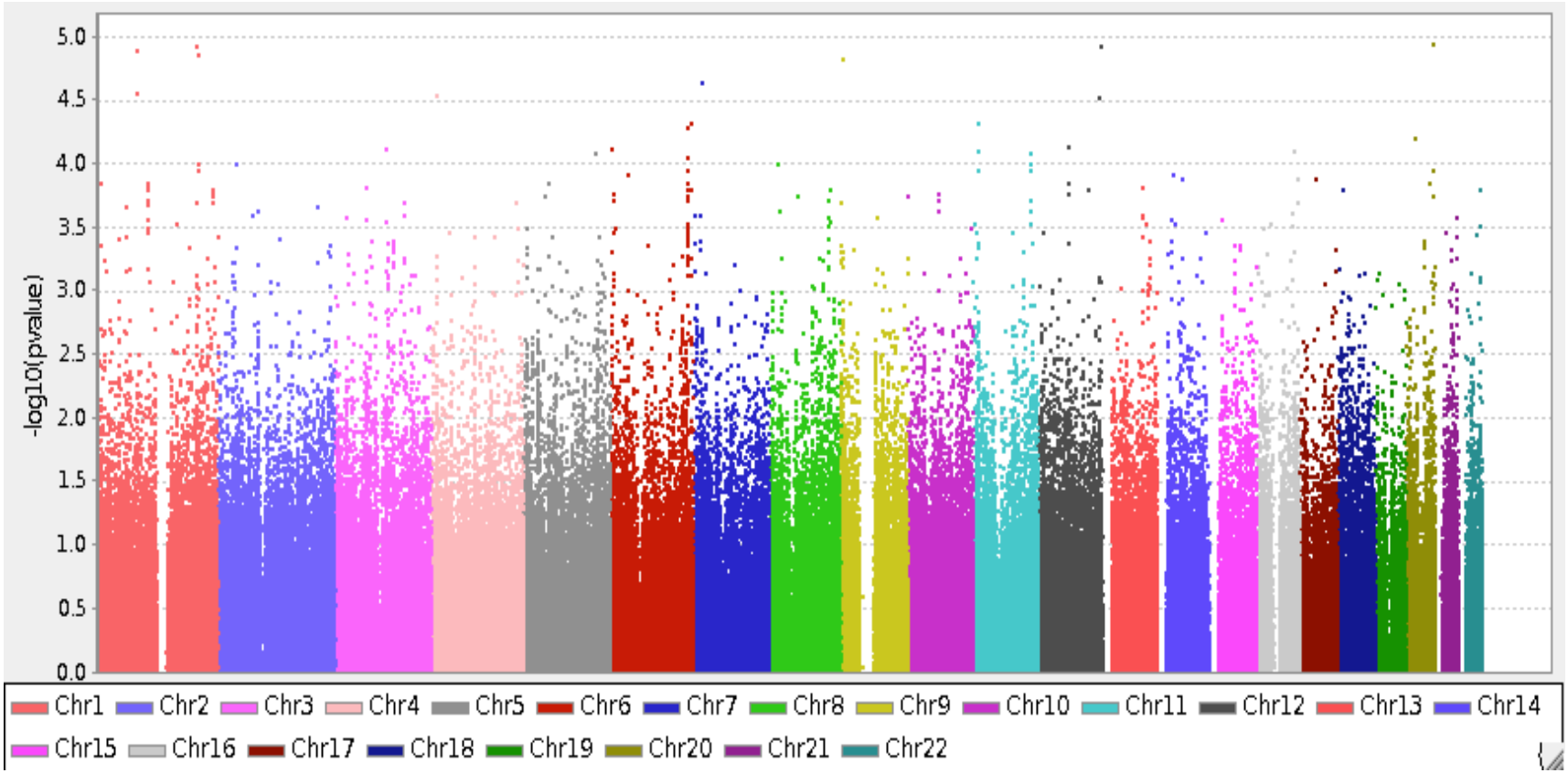


Figure 3. Genome-wide association study to mean HbA1C in the Framingham Heart Study 500K Project with adjustment for gender and age. (n=2,529)

4.6 References

1. Meigs JB, Manning AK, Dupuis J, Liu C, Florez JC, Cupples LA. Ordered stratification to reduce heterogeneity in linkage to diabetes-related quantitative traits. Obesity (Silver Spring) 2008;16:2314-22.

2. Hinney A, Vogel CI, Hebebrand J. From monogenic to polygenic obesity: recent advances. Eur Child Adolesc Psychiatry;19:297-310.

3. Lyon HN, Emilsson V, Hinney A, et al. The association of a SNP upstream of INSIG2 with body mass index is reproduced in several but not all cohorts. PLoS Genet 2007;3:e61.

4. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937-48.

5. Feitosa MF, North KE, Myers RH, Pankow JS, Borecki IB. Evidence for three novel QTLs for adiposity on chromosome 2 with epistatic interactions: the NHLBI Family Heart Study. Obesity (Silver Spring) 2009;17:2190-5.

6. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. BMC Med Genet 2007;8 Suppl 1:S18.

7. Karns R, Zhang G, Jeran N, et al. Replication of genetic variants from genome-wide association studies with metabolic traits in an island population of the Adriatic coast of Croatia. Eur J Hum Genet 2011;19:341-6.

8. Kettunen J, Silander K, Saarela O, et al. European lactase persistence genotype shows evidence of association with increase in body mass index. Hum Mol Genet;19:1129-36.

9. Ma L, Hanson RL, Traurig MT, et al. Evaluation of A2BP1 as an obesity gene. Diabetes;59:2837-45.

10. Florez JC. A genome-wide association study of treated A1C: a genetic needle in an environmental haystack? Diabetes 2010;59:332-4.

11. Pare G, Chasman DI, Parker AN, et al. Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. PLoS Genet 2008;4:e1000312.

12. Franklin CS, Aulchenko YS, Huffman JE, et al. The TCF7L2 diabetes risk variant is associated with HbA(C) levels: a genome-wide association meta-analysis. Ann Hum Genet 2010;74:471-8.

13. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190.

14. Bonnefond A, Vaxillaire M, Labrune Y, et al. Genetic variant in HK1 is associated with a proanemic state and A1C but not other glycemic control-related traits. Diabetes 2009;58:2687-97.

15. Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A(C) levels via glycemic and nonglycemic pathways. Diabetes 2010;59:3229-39.

16. Gjesing AP, Nielsen AA, Brandslund I, et al. Studies of a genetic variant in HK1 in relation to quantitative metabolic traits and to the prevalence of type 2 diabetes. BMC Med Genet 2011;12:99.

17. Hertel JK, Johansson S, Raeder H, et al. Evaluation of four novel genetic variants affecting hemoglobin A1c levels in a population-based type 2 diabetes cohort (the HUNT2 study). BMC Med Genet 2011;12:20.

18. Lange K, Sinsheimer JS, Sobel E. Association testing with Mendel. Genet Epidemiol 2005;29:36-50.

19. Wilk JB, Chen TH, Gottlieb DJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 2009;5:e1000429.

20. Glessner JT, Bradfield JP, Wang K, et al. A genome-wide study reveals copy number variants exclusive to childhood obesity cases. Am J Hum Genet;87:661-6.

21. Liu C, Li H, Qi L, et al. Variants in GLIS3 and CRY2 are associated with type 2 diabetes and impaired fasting glucose in Chinese Hans. PLoS One;6:e21464.

22. Rees SD, Hydrie MZ, O'Hare JP, et al. Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes. PLoS One;6:e24710.

23. Slattery ML, Lundgreen A, Herrick JS, Wolff RK. Genetic variation in RPS6KA1, RPS6KA2, RPS6KB1, RPS6KB2, and PDK1 and risk of colon or rectal cancer. Mutat Res;706:13-20.

24. Benyamin B, Visscher PM, McRae AF. Family-based genome-wide association studies. Pharmacogenomics 2009;10:181-90.

25. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. Am J Hum Genet 2007;81:913-26.

26. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. Nat Rev Genet;12:465-74.

27. Won S, Wilk JB, Mathias RA, et al. On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies. PLoS Genet 2009;5:e1000741.

28. Wilson AF, Ziegler A. Lessons learned from Genetic Analysis Workshop 17: transitioning from genome-wide association studies to whole-genome statistical genetic analysis. Genet Epidemiol 2011;35 Suppl 1:S107-14.

29. Becker T, Baur MP, Knapp M. Detection of parent-of-origin effects in nuclear families using haplotype analysis. Hum Hered 2006;62:64-76.

30. Sammalisto S, Hiekkalinna T, Schwander K, et al. Genome-wide linkage screen for stature and body mass index in 3.032 families: evidence for sex- and population-specific genetic effects. Eur J Hum Genet 2009;17:258-66.

31. Wilk JB, Laramie JM, Latourelle JC, et al. NYD-SP18 is associated with obesity in the NHLBI Family Heart Study. Int J Obes (Lond) 2008;32:930-5.

32. Florez JC, Manning AK, Dupuis J, et al. A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. Diabetes 2007;56:3063-74.

33. Naylor MG, Weiss ST, Lange C. A Bayesian approach to genetic association studies with family-based designs. Genet Epidemiol;34:569-74.

34. Murphy A, S TW, Lange C. Two-stage testing strategies for genome-wide association studies in family-based designs. Methods Mol Biol;620:485-96.

35. Sha Q, Zhang Z, Zhang S. Joint analysis for genome-wide association studies in family-based designs. PLoS One;6:e21957.

Chapter 5:

Conclusions and Future Research Directions

5.1 Summary and Conclusions

In the past several decades, much effort has been dedicated to the identification of disease genes through molecular genetic analysis for metabolic diseases such as type 2 diabetes (T2D)[1-3], cardiovascular diseases (CVD) [4,5] as well as other diabetes-related traits including body mass index (BMI)[6] and glycated hemoglobin (HbA1C) levels[7]. However, the pathogenesis of metabolic diseases remains incompletely understood, hindering advancement of more effective diagnosis, treatment and prevention strategies[8].

In this dissertation, we have sought to determine the association of several candidate gene loci related to adiposity, inflammation, and lipid storage, with diabetes risk using a candidate gene approach. We also aimed to discover biological pathways that may lead to the development of vascular disease (T2D and/or CVD) as well as detect genetic loci related to BMI and HbA1C levels using a genome-wide approach. In our candidate gene-based analyses, we observed a significant association between *FABP4* genotype and reduced VCAM-1 levels among African American women, a finding that warrants validation in future research. However our findings did not confirm the notion that common genetic variants in the *FABP4* gene may be associated with risk of T2D in a multiethnic cohort of postmenopausal women, We also observed modest associations between *TNF* genetic variants and circulating concentrations of TNF-α-R2, although common variants of *CRP*, *TNF*, and *IL6* genes were not significantly associated with T2D risk in postmenopausal women. Further, we replicated the association between the *PPARG Pro12Ala* genetic variant with diabetes risk and found that haplotype-based analysis is more powerful than single-SNP analysis for identifying genetic variants in the *PPARG* gene with risk of T2D.

Second, using a pathway-based analytical approach and genome-wide scan data collected among African and Hispanic American postmenopausal women in the WHI-SNP Health

Association Resource (SHARe) cohort, we observed that SNPs associated with vascular disease cluster into several biological pathways including those related to glycerolipid metabolism and PPAR signaling pathways. Third, we also observed that diabetes-related quantitative traits, i.e. BMI and HbA1C levels, were potentially associated with SNPs in the family-based Framingham Heart Study cohort using a genome-wide scan with about 500,000 SNP probes. We found strong associations between SNPs near the *LOC100507205* locus and BMI among the FHS. We also replicated five well-validated genes that have been previously reported to be significantly associated with the BMI trait.

Despite the limitations of the candidate gene approach including the limited number of variants assayed, and associations that may be hard to replicate[9], this approach can be regarded as a helpful first step in discovering potential biological pathways between genetic determinants and complex diseases such as metabolic diseases. If significant findings arise, these results may be useful in indicating biological mechanisms and suggest further experiments to test their functional roles in biochemical processes and metabolic diseases pathogenesis[10,11]. The selection of SNPs for genotyping in candidate gene study can be a challenging task during the design phase. It is important to focus on SNPs that are more prone to affect metabolic diseases risk by evaluating the linkage disequilibrium and potential haplotypes among the SNPs, as well as considering the potential function and location of each SNP[11] such as using our haplotype-tagging SNPs selection criteria: (i) cSNPs > ssSNPs > 5'-upstream SNPs > 3'-downstream SNPs > intronic SNPs; (ii) minor allele frequency (MAF) ≥ 5% in at least one of the four ethnic groups; and (iii) relatively evenly spaced across the genomic region[12].

GWAS represents an important development beyond family-based linkage studies and candidate gene approach. However, there are still areas for improvement in the GWAS field.

First, the genome-wide scan method utilized in this dissertation may only detect SNPs that are common (>5%) in our study population and therefore may miss those SNPs are rare and with low allele frequency. Rare variants may play a critical role in fully understanding the etiology of metabolic diseases. Second, we have yet to identify an independent cohort to replicate our findings in both the GWAS Pathway Analysis and the family-based GWAS using a new score test.

5.2 Future Research Directions

The primary goal in this dissertation was to examine the association between genetic variants with risk of metabolic diseases (T2D and/or CVD) and diabetes-related quantitative traits in both candidate gene and genome-wide scan settings. As mentioned earlier, we plan to validate our findings in genome-wide approach in another independent cohort. Also, stronger summary statistics can be developed in GWASPA to assess the strength of association at the pathway level. For example, a gene score can be constructed by weighting the SNP association effect on vascular disease (our disease outcome of interest) according to the specific effect sizes to better capture the pathways with genes of multiple independent associated SNPs.

Recently, a novel statistical approach has been adopted to take into account the correlations among SNPs within genes within a biological pathway[13,14]. A new multi-SNP analytical approach has also been applied to all candidate pathways of interest, for example pathways that linked with inflammatory and endothelial function[15,16] for T2D, to identify those that contains SNPs for which the cases and controls are discriminated and deduce those pathways' role in the development of metabolic diseases. It is based on the rationale that cases and controls will show more within-group similarity than across-group similarity for those SNPs

in the inflammatory and endothelial function related pathways if this set of SNPs is associated with the development of metabolic diseases[17]. These are possible analytical approaches that worth further investigation. In addition, more GWASPA method should be developed to incorporate epistastic effects among associated alleles in a pathway[18,19] and include a collection of genes sets and pathways in metabolic diseases that makes use of the high-throughput functional genomics data[20].

In this post-GWAS era, combining primary results in meta-analyses has become a widely used approach to re-examine the association of SNPs among some candidate genes with certain phenotypic traits such as those related to metabolic diseases[21-26]. In the setting of genome-wide scan for phenotypic trait genetic loci, it was previously reported that pooling the primary raw data from independent genome-wide scans is preferable to meta-analyses combining the primary results for power considerations in identifying genetic loci and lessening sources of variation. The pooling approach would be more superior if heterogeneity of participants can be reduced by conducting analyses in subgroups where within-group variation can be diminished[27]. Therefore, it may worth further investigation into the association between genetic variants with our metabolic-disease-related outcomes by combining primary data from other cohorts that offers genome-wide data in future.

Aside from genetic variation at the DNA level, epigenetic events and expression levels are also presumed to play a role in the development of metabolic diseases, especially obesity[28-31]. Epigenetic and expression research may play an important role in the development of obesity with the knowledge that family studies have offered vast evidence for moderate to high heritability of the traits. Further, complex gene-gene and gene-environment interactions may also shed some light on the development of these two traits[29]. Last but not least, it may also be

interesting to investigate the possibility of developing an analytical framework to evaluate both linkage and association at the same time. This approach would take the data from both pedigrees with relationship structures and case-control samples. This may be a powerful approach for detecting novel genetic factors related to trait loci beyond those that can be identified by case-control genome-wide scans alone[31].

SNPs are one type of genetic variants that has been emphasized in this dissertation. However, SNPs only account for a fraction of metabolic diseases heritability and leave much "missing" variation in the genetic architecture of the disease. Emerging data now indicate that structural variants (SVs), particularly copy number variants (CNVs), are an important form of genetic variation that may impact individual susceptibilities to complex diseases[9,32,33] and may explain some missing heritability. Next-generation sequencing platforms have also launched the revolution of genomics and directed their effects on studies of association between genetic variations, including CNVs, with complex diseases[34-36]. This technology is expected to raise the read length to thousands of base pairs or more in order to facilitate SNPs, CNVs and other types of SVs detection via a complete assembly of human genomes. It may also be capable of discovering copy-invariant SVs such as inversions. On the other hand, it is also superior in detecting smaller events and in verifying the exact location of variation breakpoints, as well as providing improved breakpoint resolution, copy-number accuracy, sensitivity and specificity through increasing coverage[36].

With more available data and tools to detect diseases-related genes that contribute to the development of metabolic diseases, we may not be far from increasing the wealth and quality of genetic information to improve the diagnosis, treatment, and prevention of metabolic diseases, especially for T2D, in present and the coming decade.

5.3 References

1. Busch CP, Hegele RA. Genetic determinants of type 2 diabetes mellitus. Clin Genet 2001;60:243-54.

2. Froguel P, Velho G. Genetic determinants of type 2 diabetes. Recent Prog Horm Res 2001;56:91-105.

3. O'Rahilly S, Barroso I, Wareham NJ. Genetic factors in type 2 diabetes: the end of the beginning? Science 2005;307:370-3.

4. Casas JP, Cooper J, Miller GJ, Hingorani AD, Humphries SE. Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies. Ann Hum Genet 2006;70:145-69.

5. Pollex RL, Hegele RA. Genetic determinants of the metabolic syndrome. Nat Clin Pract Cardiovasc Med 2006;3:482-9.

6. Yang W, Kelly T, He J. Genetic epidemiology of obesity. Epidemiol Rev 2007;29:49-61.

7. Soranzo N. Genetic determinants of variability in glycated hemoglobin (HbA(1c)) in humans: review of recent progress and prospects for use in diabetes care. Curr Diab Rep 2011;11:562-9.

8. Jin W, Patti ME. Genetic determinants and molecular pathways in the pathogenesis of Type 2 diabetes. Clin Sci (Lond) 2009;116:99-111.

9. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.

10. Kwok PY. Genomics. Genetic association by whole-genome analysis? Science 2001;294:1669-70.

11. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391-7.

12. Wang L, Liu S, Niu T, Xu X. SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. BMC Bioinformatics 2005;6:60.

13. Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene and pathway analysis. Eur J Hum Genet 2010;18:1045-53.

14. Yaspan BL, Bush WS, Torstenson ES, et al. Genetic analysis of biological pathway data through genomic randomization. Hum Genet 2011;129:563-71.

15. Liu S, Tinker L, Song Y, et al. A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. Arch Intern Med 2007;167:1676-85.

16. Song Y, Manson JE, Tinker L, et al. Circulating levels of endothelial adhesion molecules and risk of diabetes in an ethnically diverse cohort of women. Diabetes 2007;56:1898-904.

17. Braun R, Buetow K. Pathways of distinction analysis: a new technique for multi-SNP analysis of GWAS data. PLoS Genet;7:e1002101.

18. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. Am J Hum Genet 2010;86:6-22.

19. Jia P, Wang L, Meltzer HY, Zhao Z. Pathway-based analysis of GWAS datasets: effective but caution required. Int J Neuropsychopharmacol 2011;14:567-72.

20. Weng L, Macciardi F, Subramanian A, et al. SNP-based pathway enrichment analysis for genome-wide association studies. BMC Bioinformatics;12:99.

21. Franklin CS, Aulchenko YS, Huffman JE, et al. The TCF7L2 diabetes risk variant is associated with HbA(C) levels: a genome-wide association meta-analysis. Ann Hum Genet 2010;74:471-8.

22. Gjesing AP, Nielsen AA, Brandslund I, et al. Studies of a genetic variant in HK1 in relation to quantitative metabolic traits and to the prevalence of type 2 diabetes. BMC Med Genet 2011;12:99.

23. Kettunen J, Silander K, Saarela O, et al. European lactase persistence genotype shows evidence of association with increase in body mass index. Hum Mol Genet;19:1129-36.

24. Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A(C) levels via glycemic and nonglycemic pathways. Diabetes 2010;59:3229-39.

25. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010;42:937-48.

26. Wen W, Cho YS, Zheng W, et al. Meta-analysis identifies common variants associated with body mass index in east Asians. Nat Genet 2012;44:307-11.

27. Sammalisto S, Hiekkalinna T, Schwander K, et al. Genome-wide linkage screen for stature and body mass index in 3.032 families: evidence for sex- and population-specific genetic effects. Eur J Hum Genet 2009;17:258-66.

28. Haemer MA, Huang TT, Daniels SR. The effect of neurohormonal factors, epigenetic factors, and gut microbiota on risk of obesity. Prev Chronic Dis 2009;6:A96.

29. Hinney A, Vogel CI, Hebebrand J. From monogenic to polygenic obesity: recent advances. Eur Child Adolesc Psychiatry;19:297-310.

30. Newnham JP, Pennell CE, Lye SJ, Rampono J, Challis JR. Early life origins of obesity. Obstet Gynecol Clin North Am 2009;36:227-44, xii.

31. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. Nat Rev Genet;12:465-74.

32. Craddock N, Hurles ME, Cardin N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010;464:713-20.

33. McCarroll SA. Extending genome-wide association studies to copy-number variation. Hum Mol Genet 2008;17:R135-42.

34. Bateman A, Quackenbush J. Bioinformatics for next generation sequencing. Bioinformatics 2009;25:429.

35. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 2009;10:R32.

36. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods 2009;6:S13-20.