We performed a genome-wide association study (GWAS) and a multistage meta-analysis of type 2 diabetes (T2D) in Punjabi Sikhs from India. Our discovery GWAS in 1,616 individuals (842 case subjects) was followed by in silico replication of the top 513 independent single nucleotide polymorphisms (SNPs) (P < 10−7) in Punjabi Sikhs (n = 2,819; 801 case subjects). We further replicated 66 SNPs (P < 10−4) through genotyping in a Punjabi Sikh sample (n = 2,894; 1,711 case subjects). On combined meta-analysis in Sikh populations (n = 7,329; 3,354 case subjects), we identified a novel locus in association with T2D at 13q12 represented by a directly genotyped intron SNP (rs8552911, P = 1.82 × 10−7) in the SOCG gene. Next, we undertook in silico replication (stage 2b) of the top 513 signals (P < 10−5) in 25,157 non-Sikh South Asians (10,971 case subjects) and de novo genotyping of up to 31 top signals (P < 10−5) in 10,817 South Asians (5,157 case subjects) (stage 3b). In combined South Asian meta-analysis, we observed six suggestive associations (P < 10−3 to P < 10−2), including SNPs at HMG1L1/CTCFL, PLXNA4, SCAP, and chr5p11. Further evaluation of 31 top SNPs in 33,707 East Asians (16,746 case subjects) (stage 3c) and 47,117 Europeans (8,130 case subjects) (stage 3d), and joint meta-analysis of 128,177 individuals (44,358 case subjects) from 27 multiethnic studies, did not reveal any additional loci nor was there any evidence of replication for the new variant. Our findings provide new evidence on the presence of a population-specific signal in relation to T2D, which may provide additional insights into T2D pathogenesis. *Diabetes* 62:1746–1755, 2013

South Asians (people originating from the Indian subcontinent) comprise more than a quarter of the global population and contribute the highest number of patients with type 2 diabetes (T2D) (1). According to latest estimates, ~61 million people in India alone are currently afflicted with T2D, and their number is projected to increase to ~101 million by 2030 (2). Consequently, ~60% of the world’s coronary artery disease

---

**ORIGINAL ARTICLE**

**Genome-Wide Association Study Identifies a Novel Locus Contributing to Type 2 Diabetes Susceptibility in Sikhs of Punjabi Origin From India**

Richa Saxena,1 Danial Saleheen,2,3,4 Latonya F. Been,5 Martha L. Garavito,5 Timothy Braun,5 Andrew Bjonnnes,1 Robin Young,6 Weang Kee Ho,3 Asif Rasheed,2 Philippe Frossard,2 Xueling Sim,6,7 Neelam Hassanali,8 Venkatesan Radha,9 Manickam Chidambaram,9 Samuel Liu,9 Simon D. Rees,10,11 Daniel Peng-Keat Ng,12 Tien-Yin Wong,13,14,15 Toshimasa Yamauchi,16 Kazuo Hara,16,17 Yasushi Tanaka,18 Hiroshi Hirose,19 Mark I. McCarthy,8,20,21 Andrew P. Morris,20,21 DIAGRAM,8 MuTHER,8 AGEN,8 Abdul Basit,22 Anthony H. Barnett,11 Prasad Katulanda,8,23 David Matthews,8 Viswanathan Mohan,9,24 Gurpreet S. Wander,7 Jai Rup Singh,26 Narinder K. Mehra,7 Sarju Ralhan,25 M. Ilyas Kamboh,28 John J. Mulvihill,5 Hiroshi Maegawa,29 Kazuyuki Tobe,30 Shiro Maeda,31 Yoon S. Cho,32 E. Shyong Tai,12,33,34 M. Ann Kelly,10,11 John C. Chambers,35,36,37 Jaspal S. Kooner,35,36,38 Takashi Kadowaki,16 Panos Deloukas,39 Daniel J. Rader,40 John Danesh,3 and Dharambir K. Sanghera5

From the 1Center for Human Genetic Research and Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; the 2Center for Non-Communicable Diseases, Karachi, Pakistan; the 3Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.; the 4Department of Bio-statistics and Department of Epidemiology and Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; the 5Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; the 6Center for Statistical Genetics and Department of Statistics, University of Michigan, Ann Arbor, Michigan; the 7Centre for Molecular Epidemiology, National University of Singapore, Singapore; the 8Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, U.K.; the 9Madras Diabetes Research Foundation, Chennai, India; the 10College of Medical and Dental Sciences, University of Birmingham, Birmingham, U.K.; the 11Diabetes Centre, Heart of England National Health Service Foundation Trust, Birmingham, U.K.; the 12Saw Swee Hock School of Public Health, National University of Singapore, Singapore; the 13Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; the 14Department of Ophthalmology, National University of Singapore, Singapore; the 15Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia; the 16Department of Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Japan; the 17Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, The University of Tokyo, Tokyo, Japan; the 18Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine, Kanagawa, Japan; the 19Health Center, Keio University School of Medicine, Tokyo, Japan; the 20Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the 21Oxford Initiative for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, U.K.; the 22Baqai Institute of Diabetology and Endocrinology, Karachi, Pakistan; the 23Diabetes Research Unit, Department of Clinical Medicine, University of Colombo, Colombo, Sri Lanka; the 24Dr. Mohan’s Diabetes Specialities Centre, Chennai, India; the 25Hero Dayanand Medical College and Heart Institute, Ludhiana, Punjab, India; the 26Central University of Punjab, Bathinda, Punjab, India; the 27All India Institute of Medical Sciences and Research, New Delhi, India; the 28Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania; the 29Department of Medicine, Shiga University of Medical Science, Shiga, Japan; the 30First Department of Internal Medicine, University of Toyama, Toyama, Japan; the 31Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Kanagawa, Japan; the 32Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 200-702, Republic of Korea; the 33Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; the 34Duke-NUS Graduate Medical School Singapore, Singapore; the 35Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 200-702, Republic of Korea; the 36Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; the 37Duke-NUS Graduate Medical School Singapore, Singapore; the 38Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 200-702, Republic of Korea; the 39Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; the 40Department of Internal Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; the 41Department of Diabetes, Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Kanagawa, Japan; the 42Postgraduate Medical School, National University of Singapore, Singapore; the 43Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.; the 44Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Corresponding author: Dharambir K. Sanghera, dharambir-sanghera@ouhs.ac.uk

Received 13 August 2012 and accepted 22 November 2012. DOI: 10.2337/db12-1077

This article contains Supplementary Data online at http://diabetes.diabetesjournals.orglookup/suppl/doi:10.2337/db12-1077/DCL.

*Complete lists of the members of the DIAGRAM, MuTHER, and AGEN study groups can be found in the Supplementary Data online.*

© 2013 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons.org/licenses/by-nc-nd/3.0/ for details.

See accompanying commentary, p. 1369.
incidence of T2D and CAD is about three to five times higher in immigrant South Asians compared with Europeans (5). For these reasons, we conducted ethnic-specific genetic studies in a Sikh population to dissect genetic pathways that may contribute to T2D etiology in different ethnic groups.

The vast majority of genome-wide association studies (GWAS) on T2D so far have been performed on Europeans. Studies on non-European populations, especially those with unique demographic and cultural histories, are important for identifying population-specific linkage disequilibrium (LD) patterns and environmental factors that may modulate disease risk or protection. Interestingly, many, but not all, of the common loci originally identified in Europeans have been replicated in non-European groups (10–18). Recent GWAS in non-European populations have yielded intriguing new variants (19–21), including six novel signals in South Asians represented by single nucleotide polymorphisms (SNPs) near MTNR1B, ST6GAL1, VPS26A, HMG20A, AP3S2, and HNF4A in our recent meta-analysis of GWAS (22). Given the existence of marked genetic variability among South Asian communities, in addition to diversity in culture, language, caste system, physical appearance, and diet, they do not constitute a single homogeneous community (23). Therefore, screening populations with a different genetic and racial background or environmental exposures may improve insights about the disease and genetic risk factors (24).

People from India have a complex racial history complicated by the presence of a caste system that has prohibited interbreeding to a great extent and consequently separated people into numerous endogamous groups (25). The Sikhs, a relatively young, inbred population of 26 million (2% of the Indian population), are from the northwestern province of India and follow a distinct and unique religion born ~500 years ago in Punjab. They have an interesting background for “nontraditional” disease enrichment in the absence of conventional risk factors such as smoking, obesity, and a diet rich in meats (26). Sikhs do not smoke or chew tobacco because of religious and cultural compulsions, and ~50% of them are lifelong vegetarians. Despite the absence of these lifestyle-related risk factors, T2D and CAD have reached epidemic proportions in Sikhs. Our initial genetic studies in a Sikh cohort as part of the Asian Indian Diabetic Heart Study (AIDHS) or the Sikh Diabetes Study (SDBS) revealed an association of FTO and MTNR1B, ADIPOQ, and PPARG polymorphisms with T2D and CAD risk factors in the absence of obesity (11,27,28). In this investigation, we conducted a GWAS in a relatively homogenous Punjabi Sikh population of 1,850 individuals and performed multistage replication in up to 27 case-control studies of Punjabi, other South Asians, East Asians, and Caucasian ancestries (total n = 128,127; 44,358 T2D case and 83,769 control subjects) (Supplementary Tables 1 and 2). Study design of the discovery, replication, and meta-analysis phases was optimized to detect new population-specific and multiethnic T2D loci (Fig. 1). One important difference in the current study from our previous South Asian GWAS (22) is that in the previous study, the SNPs that were common between South Asians and Europeans were selected for replication based on the European Diabetes Genetics Replication and Meta-analysis (DIAGRAM) sample. However, in this study, the SNP selection was prioritized based on the top signals (P < 10−5) from our discovery Sikh cohort.

RESEARCH DESIGN AND METHODS

Participants. Participants were part of the Punjabi Sikh GWAS. Study samples and characteristics. Our primary Sikh GWAS (discovery) cohort used in this investigation is comprised of 1,616 individuals from the Punjabi Sikh population that was a part of the AIDHS (also named the SDBS). The AIDHS/SDBS has unique characteristics that are ideal for genetic studies. Sikhs are strictly a nonsmoking population, and ~50% of participants are teetotallers and life-long vegetarians. All individuals for the GWAS discovery cohort were recruited from one geographical location. Diagnosis of T2D was confirmed by medical records, self-report, and unannounced screening and measuring fasting glucose levels according to the guidelines of the American Diabetes Association (29), as described previously (11). Data on lipids, insulin, glucose, anthropometric measurements, education, socioeconomic status, job grade, diet, and physical activity were available on >95% of the AIDHS/SDBS individuals selected for this study. Dietary questions involving alcohol consumption were scored using a scale from 0 to 5; details are described elsewhere (26). T2D is often asymptomatic and remains undiagnosed for many years, especially in people from the developing world due to poor healthcare provisions. Therefore, it is reasonable to assume that the actual age of onset of T2D in Sikhs may range from 30 to 42 years of age compared with the observed age at diagnosis (46 years). This age is in sharp contrast to the mean age at onset of 60 years or above in developed countries (28,30). A medical record indicating either I) a fasting plasma glucose level ≥7.0 mmol/L (≥126 mg/dL) after a minimum 12-h fast or 2) a 2-h postglucose level of ≥11.1 mmol/L (≥200 mg/dL) estimated during a 2-h oral glucose tolerance test on more than one occasion, combined with symptoms of diabetes, confirmed the diagnosis. Impaired fasting glucose is defined as a fasting blood glucose level ≥5.6 mmol/L (≥100 mg/dL) but ≤7.0 mmol/L (≤126 mg/dL). Impaired glucose tolerance is defined as a 2-h OGTT > 7.5 mmol/L (≥140 mg/dL) but <11.1 mmol/L (≤200 mg/dL). The 2-h OGTT was performed according to the criteria of the World Health Organization (75-g oral load of glucose). BMI was calculated as weight (kg)/height (m)2, and waist-to-hip ratio was calculated as the ratio of abdomen or waist circumference to hip circumference. Subjects with type 1 diabetes, or those with a family member with type 1 diabetes, or rare forms of T2D subtypes (maturity-onset diabetes of the young) or secondary diabetes (e.g., hemochromatosis or hypertension) were excluded from the study. The selection of control subjects was based on a fasting glucose <100.8 mg/dL or a 2-h glucose <141.0 mg/dL. Subjects with impaired fasting glucose or impaired glucose tolerance were excluded when data were analyzed for association of the variants with T2D. All blood samples were obtained at the baseline visits. All participants signed a written informed consent for the investigations. The study was reviewed and approved by the University of Oklahoma Health Sciences Center Institutional Review Board, as well as the Human Subject Protection Committee at the participating hospitals and institutes in India.

South Asian cohorts. For stage 2a replication, the Sikh component of the London Life Sciences Population (LOLIPOP) study (22) comprised 2,919 individuals (801 T2D case and 2,018 control subjects). For stage 2b, the non-Sikh South Asian components of the LOLIPOP and the Pakistan Risk of Myocardial Infarction (Singapore Indian Eye Study [SINDI]) study (31), the Chennai Urban Rural Epidemiology Study (CURES) (32), the Diabetes Genetics in Pakistan (DGP) study, the UK Asian Diabetes Study (UKADS) (33), and the Sri Lankan Diabetes Study (SLDS) (34). Details of the contributing cohorts are provided in the Supplementary Data.

East Asian cohorts. Replication testing for stage 3c was carried out on a total of 33,707 East Asians, comprising 14,800 Japanese from RIKEN (n = 7,480).
Weinberg equilibrium (HWE) as were SNPs with call rate computation using the Impute 2 program (38).

Individuals with pi-hat estimated 1.0, so no adjustments were made (Supplementary Fig. 2).

Genomic control in simulations. After association analyses, the genomic control in our Sikh population sample and not the HapMap populations. If not included Sikhs, the 5 or 10 principal components used for this correction were estimated using our Sikh population sample and not the HapMap populations.

Statistical analyses

Association testing. Associations of SNPs with T2D among Europeans were estimated using the HapMap3 reference panel of multiethnic HapMap3 reference panel of estimated distribution of missing genotypes based on a set of known haplotypes and an estimated inbreeding coefficient (\(\theta\)), which determines the probability distribution of missing genotypes.

The Multiple Tissue Human Expression Resource (MuTHER; www.muther.ac.uk) includes lymphoblastic cell lines and skin and adipose tissue derived simultaneously from a subset of well-phenotyped Houston (GHI). Imputation yielded a total of 1,232,008 passing SNPs with MAF >1% in the Sikh GWAS. Imputed SNPs were analyzed using SNPTEST (38,40), adjusted for the covariates age, sex, BMI, and five principal components, which implements frequentist tests that calculate \(P\) values and parameter estimates and their standard errors that account for the uncertainty due to the probability distributions of the imputed genotypes, and included only those SNPs with an information score \(>0.5\) in the discovery sample as well as in all GWAS used for replication, a measure of the relative statistical information about the additive genetic effect being estimated. The genomic control value for imputed SNPs was 1.02. The inbreeding coefficient and measures of autogyosity were determined using the program PLINK. We identified runs of homozygosity using the metrics defined in Nalls et al. (41), evaluating 1-Mb autosomal regions with at least 50 adjacent SNPs, with a sliding window of 50 SNPs including no more than 2 SNPs with missing genotypes and one possible heterozygous genotype.

Stage 2 replication. We selected all independent association signals (\(r^2 <0.25\)) with \(P < 10^{-3}\) for lookup in GWAS of 1) the Sikh component of the LOLLIPOP GWAS (22) and 2) the non-Sikh South Asian components of the LOLLIPOP and PROMIS GWAS (22). A fixed-effect, inverse-variance meta-analysis (as implemented in METAL) (42) was used to combine the results for individual studies.

Stage 3 replication. Significant association results with \(P < 10^{-3}\) based on meta-analysis of stages 1, 2a, and 2b were selected for de novo or in silico replication in Sikh, South Asian, other Asian, and European populations. In addition, we selected SNPs from a Sikh-only meta-analysis of stages 1a and 2a for genotyping in an in-house Punjabi Sikh T2D case-control population. In our previous South Asian GWAS by Kooven et al. (22), 300 of the 3,200 samples of the AIDHS/SDS (used in replication) were genotyped using Illumina 660 Quad arrays, and the remaining samples (from 1,187 case and 1,632 control subjects) were genotyped using Sequenom MassARRAYs. However, in this study, in addition to GWAS set \((n = 1,016)\), SNPs were genotyped de novo on our remaining replication set \((n = 2,894)\). Signals with \(P < 10^{-3}\) after meta-analysis of stages 1, 2a, and 3a were also genotyped in the South Asian, other Asian, and European populations to test if they were specific to the Sikh ethnic group or spanned ethnicities. All meta-analyses were performed using a fixed-effects, inverse-variance meta-analysis implemented in METAL.

MuTHER Consortium. The Multiple Tissue Human Expression Resource (MuTHER; www.muther.ac.uk) includes lymphoblastic cell lines and skin and adipose tissue derived simultaneously from a subset of well-phenotyped
healthy female twins from the Twins UK adult registry. Whole-genome expression profiling of the samples, each with either two or three technical replicates, was performed using the Illumina HumanHT-12 v3 BeadChips according to the protocol supplied by the manufacturer. Log-transformed expression signals were normalized separately per tissue as follows. Quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was performed with a combination of Illumina arrays HumanHap300, HumanHap610Q, 1M-Duo, and 1.2M-Duo 1M. Untyped HapMap2 SNPs were imputed using the IMPUTE software package (v2). The number of adipose samples with genotypes and expression values was 77% Association between all SNPs (MAF >5%; IMPUTE info >0.8) within a gene or within 1 Mb of the gene transcription start or end site and normalized expression values was performed with the GenABEL/ProbABEL packages using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL nmscore test with imputed genotypes. Age and experimental batch were included as cofactors.

RESULTS
Punjabi Sikh discovery GWAS. Clinical characteristics of the stage 1 Punjabi Sikh T2D GWAS cohort and stage 2a and 2b replication cohorts are described in Supplementary Table 3. Principal components analysis revealed part of the population structure (Supplementary Fig. 1). After quality control, 524,216 directly genotyped SNPs in 1,616 subjects (842 case and 774 control subjects) from 1,850 total subjects were available for association testing after removing samples showing cryptic relatedness through identity-by-descent sharing. To increase genome coverage, genotypes were available for association testing after removing samples showing cryptic relatedness through identity-by-descent sharing. To increase genome coverage, genotypes were imputed for untyped SNPs using the HapMap3 multiethnic reference panel (see research design and methods), yielding a total of 1,232,008 SNPs for association analyses. The reason for choosing a more cosmopolitan panel and not restricting to the GIH was based on our own data showing equal diversity of the Sikhs from GIH and CEU, and based on previously described advantages of using a worldwide reference panel (39). We performed a GWAS for T2D adjusted for covariates age, sex, BMI, and five principal components (Supplementary Fig. 1); no evidence of inflation was observed (Supplementary Fig. 2A and B) (see research design and methods).

Replication and meta-analyses in Punjabi Sikh participants. We undertook a two-stage replication in T2D case-control samples of Punjabi Sikh ancestry (stages 2a and 3a in Fig. 1). Lead SNPs representing 513 novel, independent (\(r^2 < 0.25\)) association signals with \(P < 10^{-4}\) in the discovery GWAS (including only two previously known GWAS SNPs from TCF7L2 and IGF2BP2 and excluding 62 SNPs with \(P < 10^{-4}\) from other known T2D loci) were tested for in silico replication in the Punjabi Sikh subcomponent of the LOLIPOP GWAS comprising 801 T2D case and 2,018 control subjects (Supplementary Table 1). Top SNPs representing 66 putatively novel signals with \(P < 10^{-4}\) after stage 1 and 2a meta-analysis using a fixed effects, inverse-variance approach were directly genotyped in the stage 3a sample of 2,894 Punjabi Sikh individuals (1,711 T2D case and 1,183 control subjects) (Fig. 1 and Supplementary Table 2).

In a combined meta-analysis of the three Punjabi studies (\(n = 7,329\)), we identified one new locus reaching genome-wide significance (\(P < 5 \times 10^{-8}\)) along with robust replication of the established SNP rs7903146 in TCF7L2 (\(P = 3.32 \times 10^{-10}\)) in Sikhs (Figs. 2, 3, and 4). This novel association signal lies in a 164-kb region of strong LD at 13q12 (harboring genes gamma-sarcoglycan [SGCG] and sarcosin [SACS]) and is represented by a directly genotyped intronic SNP, rs9552911 in SGCG (odds ratio [OR] 0.67 [95% CI 0.58–0.77], \(P = 1.82 \times 10^{-8}\) for the minor “A” allele) (Table 1, Fig. 4, and Supplementary Table 5). Excluding BMI from the logistic regression model did not affect the association (Supplementary Table 6). Furthermore, including five additional principal components in the model did not attenuate the signal; indeed, the effect and significance were slightly improved (Supplementary Table 6).

The genetic variance (\(R^2\)) explained by this variant for the T2D phenotype in Punjabi Sikh discovery and replication sets was 1.57 and 1.34%, respectively. There were 15 additional independent loci with suggestive evidence (\(P < 10^{-5}\) to \(P < 10^{-7}\)) of association, including six unknown regions along with IGF2BP2, originally identified in Caucasians (43) (Supplementary Table 5). Meta-analysis results including non-Sikh Punjabis from PROMIS (Pakistan) revealed suggestive association (\(P < 10^{-4}\) to \(P < 10^{-7}\)) at SNPs from three new regions: chromosome 18q21 ZBTB7C (rs1893835), 20q13, near HMG1Li/CTCF/ RBM38/PCK1 (rs328506), and 5q33 (rs17053082) (Supplementary Table 7). Association results for 42 previously reported T2D loci in the Punjabi cohort are summarized in Supplementary Table 14. Most loci showed consistent effect in the same direction and 33 out of 42 were associated with T2D at \(P < 0.05\) in Sikhs.

Replication/evaluation and meta-analysis in other South Asians. In order to identify T2D association signals common to Punjabi and other South Asian populations, we tested the association of the 513 top independent signals (\(P < 10^{-3}\)) derived from the discovery cohort in GWAS from the LOLIPOP, PROMIS, and RACE studies as part of stage 2b replication (10,971 T2D case and 18,186 control subjects) (Fig. 1 and Supplementary Table 1). Thirty-one signals (\(P < 10^{-4}\) from an interim analysis with stage 2b) were further genotyped in 10,817 South Asians (5,157 T2D and 5,660 control subjects) (Fig. 1) as part of stage 3b replication. Clinical characteristics of the stage 3 replication cohorts are described in Supplementary Table 4. Combined South Asian meta-analysis revealed nominally significant association in six SNPs with MAF >5% (\(P < 10^{-4}\)), but only the two previously known SNPs in TCF7L2 and IGF2BP2 reached genome-wide significance (Table 1 and Supplementary Table 8). Suggestive novel signals included SNPs at chromosome 20q13 near PLXNA4 (rs1593304), 3p21 in SCPB (rs48508889), and 5p11 (rs1315082) (Supplementary Table 8). Further studies and replication in a larger sample will be required to validate these results and identify causal variants at these loci.

Multiethnic replication and meta-analysis. To identify T2D signals spanning ethnicities, we extended the replication of 31 SNPs with \(P < 10^{-4}\) in Punjabis and South Asians (stage 3b) to East Asians (AGEN+) and Europeans (DIAGRAM+) in stages 3c and 3d, respectively (Fig. 1). Upon meta-analysis of 31 loci in Asians (South Asians and AGEN+), genome-wide associations were only seen in TCF7L2 (rs7903146, \(P = 1.93 \times 10^{-3}\)) and IGF2BP2 (rs1470579, \(P = 1.54 \times 10^{-13}\)) (Supplementary Table 9). In joint multiethnic meta-analysis on 28,127 individuals from 27 studies, only two previously known loci, TCF7L2 (rs7903146, \(P = 8.53 \times 10^{-3}\)) and IGF2BP2 (rs1470579, \(P = 1.81 \times 10^{-15}\)), showed robust associations. Interestingly, none of the Punjabi hits could be independently confirmed in AGEN+ or DIAGRAM+ (notably, the lead rs9552911 variant from SGCG was monomorphic in DIAGRAM+ (Table 1 and Supplementary Table 10). Lookup of 50 kb upstream and downstream of SNPs within the SGCG locus in the publicly available data of
the Meta-Analyses of Glucose and Insulin-Related Traits Consortium (MAGIC) study on glycemic trait GWAS (44,45) revealed several nominal associations of SNPs with fasting blood glucose and 2-h glucose levels (Supplementary Fig. 3). Some of these SNPs also showed an association with fasting blood glucose and waist or waist-to-hip ratio in Sikhs (Supplementary Table 11), but none of these were in LD ($r^2$ >0.20) with our lead SNP.

**Gene expression studies.** We examined the expression of *SGCG* and neighboring genes (*FLJ46358*, *MIPEP*, *SACS*, and *sTNFRSF19*) within 1 Mb of the index SNP by cis expression quantitative trait locus (eQTL) analysis using adipose tissue, skin, and lymphoblastic cell line gene expression data from the MuTHER Consortium, comprising healthy female twins of European ancestry from Britain. Several SNPs in the *SGCG* region were associated with

---

**FIG. 2.** A: Manhattan plot showing primary genome-wide association analysis of the Punjabi Sikh discovery cohort using directly genotyped (524,216) SNPs. B: Manhattan plot shows imputed 1,232,008 SNPs on the x-axis and $-\log_{10}(P$ value of association on the y-axis. Locations of the three loci (including one novel locus at *SGCG*) reached genome-wide significance after combined analysis of the GWAS and replication data in Punjabi Sikhs.
significantly elevated ($P_{eQTL} < 10^{-4}$ to $10^{-9}$) expression of SGCG mRNA in adipose tissues (Supplementary Table 12 and Supplementary Fig. 4). One adipose eQTL from MuTHER (rs572303, $P_{eQTL} = 5.47 \times 10^{-4}$) located within SGCG showed a nominally significant association with increased waist circumference in Sikhs ($\beta = 0.67, P = 5.2 \times 10^{-4}$) (Supplementary Table 11). As shown in Supplementary Fig. 4, the LD patterns in the region (~1.46 Mb) surrounding the SGCG variant (rs9552911) varied in East Asians (JPT), Africans (YRI), Caucasians (CEU), Gujarati Indians (GIH), and Sikhs. Interestingly, in Caucasians and Yorubians, this variant was monomorphic. However, several alternative SNPs from this region in Europeans were nominally associated with fasting blood glucose (MAGIC study, $r^2$ ranging from 0.10 to 0.20 with the index SNP [rs9552911]) and mRNA expression of adipose cells in
the MuTHER study ($r^2$ ranging from 0.14 to 0.26 with the index SNP). These data suggest that population differences may underlie the weak LD. It is possible that a single causal variant may be responsible for these associations, but LD may differ between Sikhs, Europeans, and other populations.

**Comparative analysis of autozygosity.** We further looked to compare the distributions of inbreeding coefficients and autozygosity as described by Nalls et al. (41). As expected, the inbreeding coefficients in our sample were higher compared with two outbred populations of European Americans, Coriell, and Baltimore Longitudinal Study of Aging (BLSA) ($F = 0.041 \pm 0.018$ in Sikhs vs. $F = 0.007 \pm 0.019$ in Coriell and $F = -0.3 \pm 0.012$ in BLSA), as assessed by Nalls et al. (41). However, these results were similar to other Indian populations previously reported by Reich et al. (46). No significant difference in inbreeding was observed between case and control subjects ($P = 0.59$). Autozygosity analysis determined that there were 19 ± 7 homozygous segments > 1 Mb in length, with an average length of 2.0 ± 0.95 Mb. Hence, fewer but longer autozygous segments were found in our population than in outbred populations. No correlation of measures of autozygosity to age was observed ($P > 0.05$) across decades of age.

**DISCUSSION**

In this GWAS and multistage meta-analysis, a novel locus at 13q12 in the *SGCG* (rs9552911) gene was identified as associated with T2D susceptibility in Punjabi Sikhs from Northern India. *SGCG* is a member of the sarcoglycan complex of transmembrane glycoproteins mutated in autosomal recessive muscular dystrophy, in particular limb-girdle muscular dystrophy type 2C (LGMD2C). *SGCG* is expressed in skeletal muscle, and its high expression is also seen in vascular smooth muscle cells as well as in breast cancer cell lines (47,48). The red diamond signifies the OR and CI across all multiethnic samples.

**FIG. 4.** Forest plot showing the association of lead SNP in the *SGCG* (rs9552911) gene with T2D. For each study, the estimates of the ORs with 95% CI are shown. In addition, meta-analyses of Sikhs and South Asians and multiethnic studies are shown at the bottom. Meta-analysis in Sikhs shows a significant association of rs9552911 with T2D (OR 0.67 [0.58–0.77], $P = 1.82 \times 10^{-8}$). The red diamond signifies the OR and CI across all multiethnic samples.

*SGCG* targeted knockout mice displayed a variety of phenotypes, including dystrophic cardiomyopathy and defects in skeletal muscle, metabolism, homeostasis, growth, apoptosis, aging, and behavior (50–53). Mice lacking the sarcoglycan
### TABLE 1

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (Mb)</th>
<th>Chromosome</th>
<th>South Asian</th>
<th>Punjabi</th>
<th>Other South Asian</th>
<th>Duplicate</th>
<th>Replication</th>
<th>GWAS (a)</th>
<th>Nearest Gene</th>
<th>2013 sex</th>
<th>Nearest Gene</th>
<th>2013 sex</th>
<th>Nearest Gene</th>
<th>2013 sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2895607</td>
<td>19</td>
<td>11</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>2.00 (1.43-2.98)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
<td>1.00 (0.92-1.07)</td>
</tr>
</tbody>
</table>

GWAS, replication, and meta-analysis results of T2D loci identified in Punjabi Sikhs.
complex including SGC\textsuperscript{\textregistered} in adipose and skeletal muscle were shown to be glucose intolerant and exhibited whole-body insulin resistance due to impaired insulin-stimulated glucose uptake in skeletal muscle (54).

The allelic distribution of the less common “A” (protective) allele of rs9552911 ranged from 0.06 to 0.15 in South Asians and differed between other South Asians (0.11) and Punjabi Sikhs (0.08) (see details in Supplementary Table 13). Further replication in large independent datasets of South Asians and Punjabi Sikhs would be needed to confirm the pattern of observed association. In view of the complex racial history complicated by a well-defined caste system, Indian populations display a great deal of genetic and cultural diversity (55). Studies suggest that genetic affinity among endogamous communities in India is inversely correlated with geographic distance between them (23).

Therefore, it is possible that undetected causal variant(s) or multiple rare variants in LD with this marker arose on a haplotype tagged by rs9552911 in Punjabi Sikhs after divergence from other South and West Indian populations. This variation in the index SNP rs9552911 does not appear to be of recent origin, as suggested by comparative genomic analysis (Supplementary Fig. 5). Two important nuclear hormone receptors and transcription factors (peroxisome proliferator--activated receptor-\(\gamma\) [PPAR-\(\gamma\)] [1 and 2] and PPAR-\(\alpha\]) bind to the promoter and intron 1 of the SGC\textsuperscript{\textregistered} gene. Further, the maturity-onset diabetes of young 4 (MODY4) locus at chromosome 13q12, represented by insulin promoter factor 1 or PDX-1, lies next to the SGC\textsuperscript{\textregistered} locus. Therefore, further in-depth examination by targeted resequencing in the extended region and functional studies may reveal putative causative variants in this extended region and provide insight into the physiological relevance of the observed association.

In summary, our study identified a novel locus associated with T2D in a population of Punjabi Sikh ancestry from Northern India. These findings not only provide new information on previously unknown regions associated with T2D but demonstrate a putative population-specific association that could lead to additional biological insights into T2D pathogenesis.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases Grant R01-DK-082766 and a seed grant from the University of Oklahoma Health Sciences Center.

No potential conflicts of interest relevant to this article were reported.


The authors thank the research participants for their contribution and support for making this study possible. The authors also thank Bansari Mehta and Braden Juengel (University of Oklahoma Health Sciences Center) and Jackie Lane (Harvard Medical School) for their superb technical support in this study. The authors thank the collaborating groups of all the contributing sites including Dr. Nadereh Jafari and staff at the Northwestern University Genomics Core (Evanston, IL) for performing genome-wide genotyping and Dr. Andrew Brooks and staff at Rutgers University (New Brunswick, NJ) for replication genotyping. Study-specific acknowledgments are provided in the Supplementary Data.

REFERENCES


32. Chidambaram M, Radha V, Mohan V. Replication of recently described type 2 diabetes gene variants in a South Indian population. Metabolism 2010;59:1760–1766


