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

Richa Saxena,[1] Danish Saleheen,[2,3,4] Latonya F. Been,[5] Martha L. Garavito,[5] Timothy Braun,[5] Andrew Bjonnes,[1] Robin Young,[3] Weang Kee Ho,[3] Asif Rasheed,[2] Philippe Frossard,[2] Xueling Sim,[6,7] Neelam Hassanali,[8] Venkatesan Radha,[9] Manickam Chidambaram,[9] Samuel Liju,[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] DIAGRAM,* MuTHER,* AGEN,* Abdul Basit,[22] Anthony H. Barnett,[11] Prasad Katulanda,[8,23] David Matthews,[8] Viswanathan Mohan,[9,24] Gurpreet S. Wander,[25] Jai Rup Singh,[26] Narinder K. Mehra,[27] 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. Sanghera[5]

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) ($P < 10^{-3}$) 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 intronic SNP (rs9552911, $P = 1.82 \times 10^{-8}$) in the SGCG gene. Next, we undertook in silico replication (stage 2b) of the top 513 signals ($P < 10^{-3}$) in 29,157 non-Sikh South Asians (10,971 case subjects) and de novo genotyping of up to 31 top signals ($P < 10^{-4}$) in 10,817 South Asians (5,157 case subjects) (stage 3b). In combined South Asian meta-analysis, we observed six suggestive associations ($P < 10^{-5}$ to $< 10^{-7}$), 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 128,127 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

From the [1]Center for Human Genetic Research and Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; the [2]Center for Non-Communicable Diseases, Karachi, Pakistan; the [3]Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.; the [4]Department of Biostatistics and Department of Epidemiology and Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; the [5]Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; the [6]Center for Statistical Genetics and Department of Statistics, University of Michigan, Ann Arbor, Michigan; the [7]Centre for Molecular Epidemiology, National University of Singapore, Singapore; the [8]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, U.K.; the [9]Madras Diabetes Research Foundation, Chennai, India; the [10]College of Medical and Dental Sciences, University of Birmingham, Birmingham, U.K.; the [11]Diabetes Centre, Heart of England National Health Service Foundation Trust, Birmingham, U.K.; the [12]Saw Swee Hock School of Public Health, National University of Singapore, Singapore; the [13]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; the [14]Department of Ophthalmology, National University of Singapore, Singapore; the [15]Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia; the [16]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; the [17]Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, The University of Tokyo, Tokyo, Japan; the [18]Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine, Kanagawa, Japan; the [19]Health Center, Keio University School of Medicine, Tokyo, Japan; the [20]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; the [21]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, U.K.; the [22]Baqai Institute of Diabetology and Endocrinology, Karachi, Pakistan; the [23]Diabetes Research Unit, Department of Clinical Medicine, University of Colombo, Colombo, Sri Lanka; [24]Dr. Mohan's Diabetes Specialities Centre, Chennai,

India; the [25]Hero Dayanand Medical College and Heart Institute, Ludhiana, Punjab, India; the [26]Central University of Punjab, Bathinda, Punjab, India; the [27]All India Institute of Medical Sciences and Research, New Delhi, India; the [28]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania; the [29]Department of Medicine, Shiga University of Medical Science, Shiga, Japan; the [30]First Department of Internal Medicine, University of Toyama, Toyama, Japan; the [31]Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Kanagawa, Japan; the [32]Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do 200-702, Republic of Korea; the [33]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; the [34]Duke-NUS Graduate Medical School Singapore, Singapore; the [35]Ealing Hospital National Health Service Trust, Middlesex, U.K.; the [36]Imperial College Healthcare National Health Service Trust, London, U.K.; [37]Epidemiology and Biostatistics, Imperial College London, London, U.K.; the [38]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, U.K.; the [39]Wellcome Trust Sanger Institute, Hinxton, U.K.; and the [40]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

(CAD), a principal cause of mortality in individuals with T2D, is expected to occur in India (3). There is considerable ethnic difference in the prevalence and progression of T2D and CAD. In addition to environmental factors, genetic factors influence disease susceptibility (4). The incidence of T2D and CAD is about three to five times higher in immigrant South Asians compared with Euro-Caucasians, and the age of onset of T2D is roughly a decade earlier in South Asians than in Europeans (5–7). The higher prevalence of T2D among South Asians settled in developed countries compared with the host population reflects the genetic and ethnic predisposition to cardiometabolic disease under an adverse environment and the joint effects of genes and environment in the predisposition to T2D (8). 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 (9). 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 *GRB14*, *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 (SDS) revealed an association of *FTO* and *MTNR1B*, *ADIPOQ*, and *PPARG* polymorphisms with T2D and 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 Asian, East Asian, 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 ($P < 10^{-3}$) from our discovery Sikh cohort.

## RESEARCH DESIGN AND METHODS

**Participants.** Participants were part of the Punjabi Sikh GWAS.
**Study sample 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 SDS). The AIDHS/SDS has unique characteristics that are ideal for genetic studies. Sikhs are strictly a nonsmoking population, and ~50% of participants are teetotalers and life-long vegetarians. All individuals for the GWAS discovery cohort were recruited from one geographical location. Diagnosis of T2D was confirmed by scrutinizing medical records for symptoms and use of medication 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/SDS 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 39 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 (5,26,30). A medical record indicating either *1*) 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 >7.8 mmol/L (>140 mg/dL) but <11.1 mmol/L (<200 mg/dL). The 2-h OGTTs were 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 (from, e.g., hemochromatosis or pancreatitis) 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 Study (PROMIS; and the Risk Assessment of Cerebrovascular Events [RACE] study) GWAS (22) comprised 29,157 individuals (10,971 case and 18,186 control subjects) (22). Stage 3a Punjabi-specific replication was carried out on 2,894 individuals (1,711 case and 1,183 control subjects) of Punjabi ancestry from India as part of AIDHS/SDS, and replication testing among South Asians for stage 3b was carried out among 10,817 participants (5,157 case and 5,660 control subjects), which were part of the following studies: Asian Indians from the Singapore Indian Eye (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 14,890 Japanese from RIKEN ($n$ = 7,480
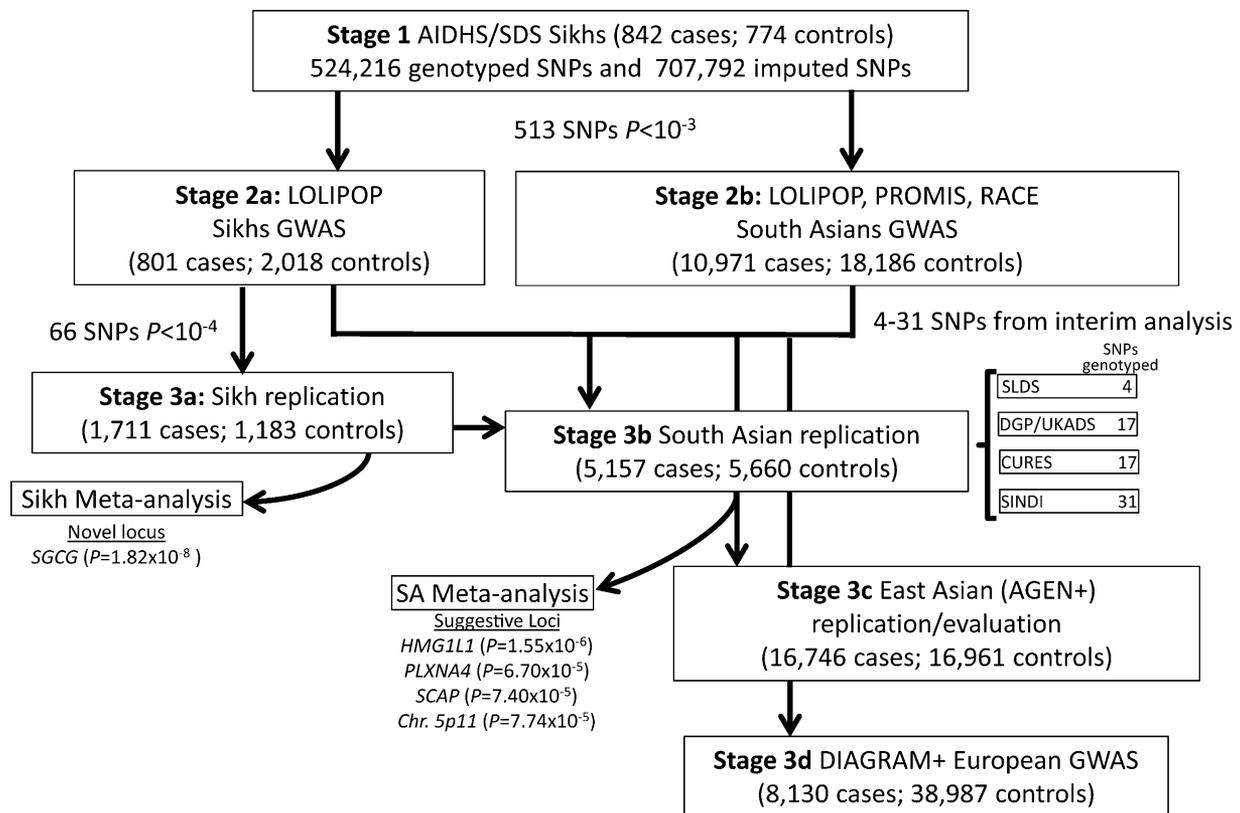
**FIG. 1. Summary of study design and outcome of key findings.**

genotyped) and BioBank Japan ($n = 7,410$ GWAS) (19,35) and 18,817 individuals of East Asian ancestry as part of the Asian Genetic Epidemiology Network (AGEN) with genotype data available from eight GWAS (21).

**DIAGRAM (Euro-Caucasians).** Associations of SNPs with T2D among Europeans were tested in silico using results from the genome-wide association phase of the DIAGRAM study comprising 47,117 subjects (36).

**Genotyping and quality control.** Genomic DNA was extracted from buffy coats using QiaAmp blood kits (Qiagen, Chatsworth, CA) or by the salting-out procedure (37). Stage 1 genome-wide genotyping was performed using a Human 660W-Quad BeadChip panel (Illumina, Inc., San Diego, CA). We performed pairwise identity-by-state clustering in PLINK across all individuals to assess population stratification; no population outliers were detected. Related individuals with pi-hat >0.3 and samples with <93% call rate were excluded, as were SNPs with call rate <95%. Also excluded were SNPs with Hardy-Weinberg equilibrium (HWE) $P < 10^{-6}$ or minor allele frequency (MAF) <1%. After quality control, 524,216 directly genotyped SNPs in 1,616 subjects (842 case and 774 control subjects) were available for association testing.

Genotyping for de novo SNPs in the replication samples was performed by Sequenom MassArray (BioMark HD MX/HX Genetic Analysis System; Fluidigm) or KASPAR (LGC Genomics KBioscience, London, U.K.). Samples and SNPs with <95% call rate were excluded, as were those that deviated from HWE at $P < 10^{-3}$. The associations of SNPs with T2D were tested in each cohort separately.

**Statistical analyses**

*Association testing.* Associations of SNPs with T2D were tested using logistic regression and an additive genetic model. Age, sex, BMI, and 5 or 10 principal components to adjust for residual population stratification were included as covariates. As the existing HapMap2 or HapMap3 and 1000 Genomes data do not include Sikhs, the 5 or 10 principal components used for this correction were estimated using our Sikh population sample and not the HapMap populations. After association analyses, the genomic control inflation factor ($\lambda$) was 1.0, so no adjustments were made (Supplementary Fig. 2A and B).

In addition to the analysis of directly genotyped SNPs, we performed imputation using the Impute 2 program (38–40), which determines the probability distribution of missing genotypes based on a set of known haplotypes and an estimated fine-scale recombination map. Imputation was based on the entire multiethnic HapMap3 reference panel of ~1.5 million autosomal SNPs with MAF >1% in 1,011 individuals from Africa, Asia, Europe, and the Americas (including 1,362,138 SNPs from the Indian population of 100 Gujaratis from

Houston [GIH]). 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 autozygosity 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 LOLIPOP GWAS (22) and *2*) the non-Sikh South Asian components of the LOLIPOP 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 Kooner 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 ($n = 1,616$), SNPs were genotyped de novo on our remaining replication set ($n = 2,894$). Signals with $P < 10^{-4}$ 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. $Log_2$-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 1.2MDuo 1M. Untyped HapMap2 SNPs were imputed using the IMPUTE software package (v2). The number of adipose samples with genotypes and expression values is 776. 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 mmscore score 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 little 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 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 $P < 10^{-3}$ in the discovery GWAS (including only two previously known GWAS SNPs from *TCF7L2* and *IGF2BP2* and excluding 62 SNPs with $P < 10^{-3}$ 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* ($P = 3.32 \times 10^{-19}$) 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 sacsin [*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 $< 10^{-7}$) 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 ($P < 10^{-5}$ to $< 10^{-7}$) at SNPs from three new regions: chromosome 18q21 *ZBTB7C* (rs1893835), 20q13, near *HMG1L1/CTCFL/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% ($P \leq 10^{-4}$), 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 *HMG1L1/CTCFL/RBM38/PCK1* (rs328506), 7q32 near *PLXNA4* (rs1593304), 3p21 in *SCAP* (rs4858889), and 5p11 (rs13155082) (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, $P = 1.93 \times 10^{-38}$) and *IGF2BP2* (rs1470579, $P = 1.54 \times 10^{-13}$) (Supplementary Table 9). In joint multiethnic meta-analysis on 128,127 individuals from 27 studies, only two previously known loci, *TCF7L2* (rs7903146, $P = 8.53 \times 10^{-75}$) and *IGF2BP2* (rs1470579, $P = 1.81 \times 10^{-19}$), 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
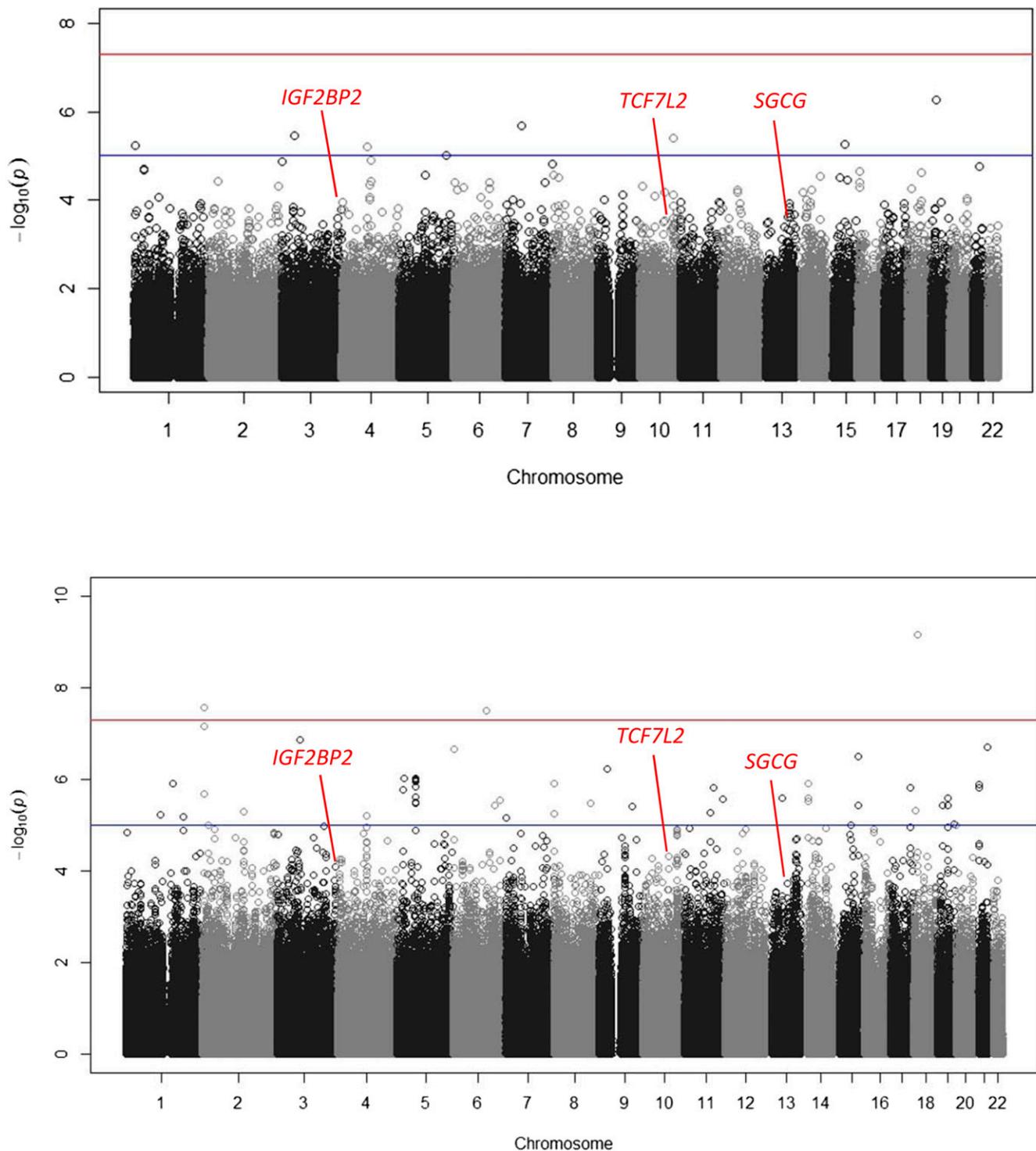
FIG. 2. *A*: Manhattan plot showing primary genome-wide association analysis of the Punjabi Sikh discovery cohort using directly genotyped (524,216) SNPs. *B*: Manhattan plot shows imputed 1,232,008 SNPs on the *x*-axis and $-\log_{10} P$ value of association on the *y*-axis. Locations of the three loci (including one novel locus at SGCG) reached genome-wide significance after combined analysis of the GWAS and replication data in Punjabi Sikhs.

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 s*TNFRSF19)* 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
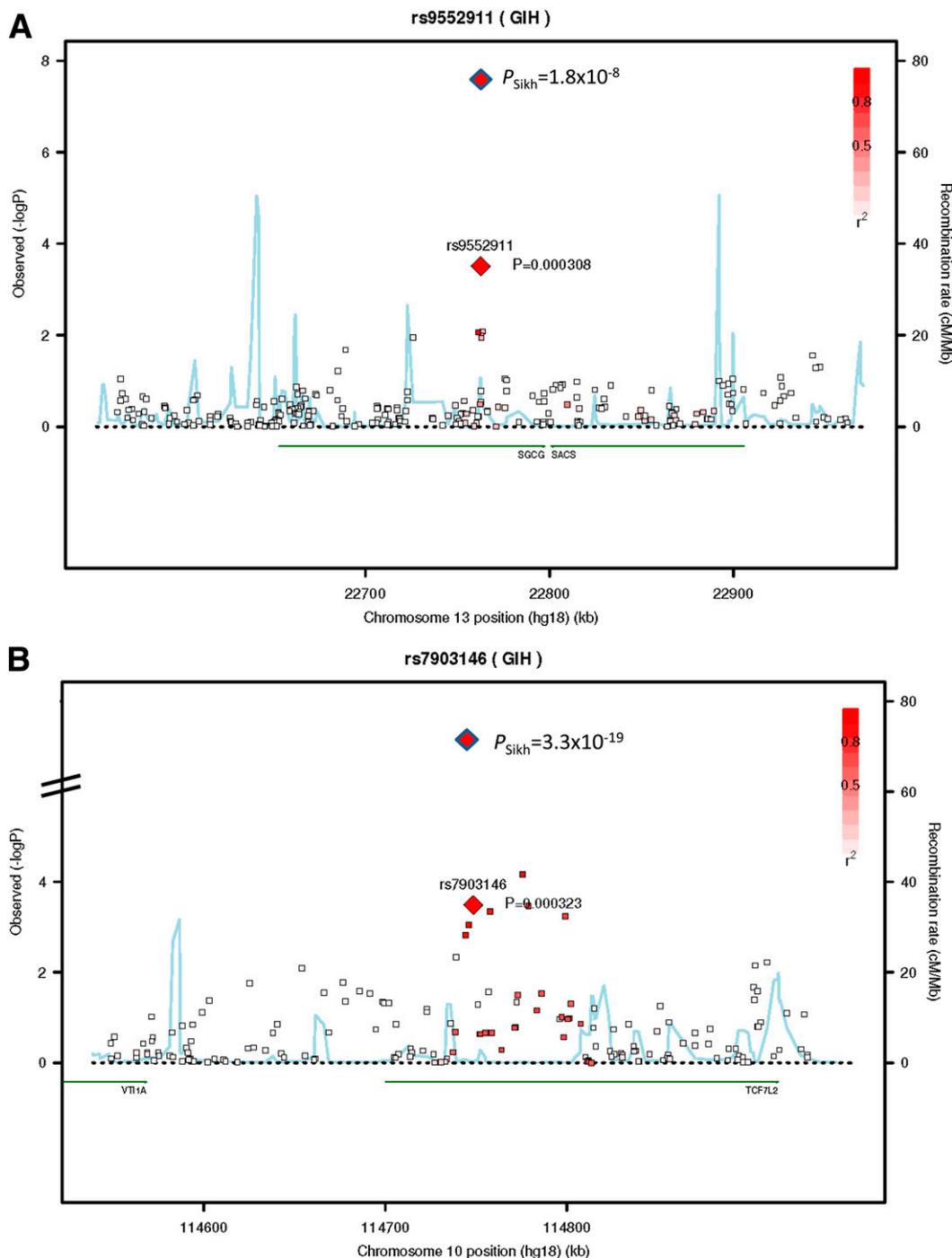
FIG. 3. *A*: Regional association plot for a new T2D locus detected at 13q12 in the *SGCG* gene from the genome-wide meta-analysis in Sikhs. *B*: A strong confirmation of SNPs in the *TCF7L2* gene in Sikh meta-analysis. In these plots, the SNPs showing the most strongly associated signal are depicted as a red diamond with blue border for the combined stage 1, 2a, and 3a results for meta-analysis, and the red diamond with black border shows evidence of association for the stage 1 results. Each square in color shows a SNP with the color scale relating the $r^2$ value for that SNP and the top SNP taken from the HapMap 3 GIH panel. We present LD using the GIH panel, the closest HapMap population to the Sikhs; however, we note that there could still be differential LD between the reference panel and the Sikh population. At the bottom of the plot, the locations of known genes in the region are shown.

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 ($\beta = 0.67$, $P = 5.2 \times 10^{-2}$) (Supplementary Table 11). As shown in Supplementary Fig. 4, the LD patterns in the region ($\sim$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
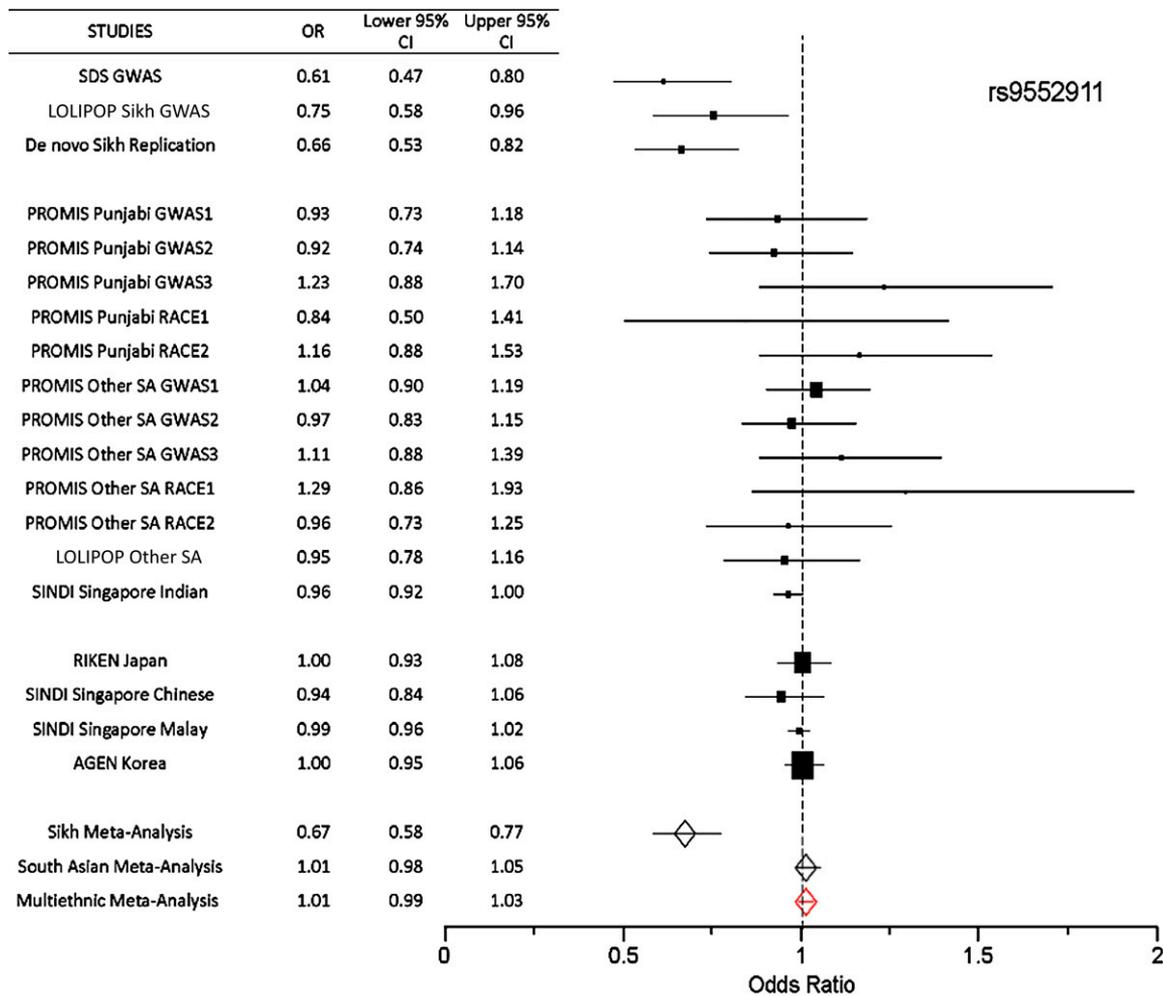
| STUDIES | OR | Lower 95% CI | Upper 95% CI |
|---|---|---|---|
| SDS GWAS | 0.61 | 0.47 | 0.80 |
| LOLIPOP Sikh GWAS | 0.75 | 0.58 | 0.96 |
| De novo Sikh Replication | 0.66 | 0.53 | 0.82 |
| PROMIS Punjabi GWAS1 | 0.93 | 0.73 | 1.18 |
| PROMIS Punjabi GWAS2 | 0.92 | 0.74 | 1.14 |
| PROMIS Punjabi GWAS3 | 1.23 | 0.88 | 1.70 |
| PROMIS Punjabi RACE1 | 0.84 | 0.50 | 1.41 |
| PROMIS Punjabi RACE2 | 1.16 | 0.88 | 1.53 |
| PROMIS Other SA GWAS1 | 1.04 | 0.90 | 1.19 |
| PROMIS Other SA GWAS2 | 0.97 | 0.83 | 1.15 |
| PROMIS Other SA GWAS3 | 1.11 | 0.88 | 1.39 |
| PROMIS Other SA RACE1 | 1.29 | 0.86 | 1.93 |
| PROMIS Other SA RACE2 | 0.96 | 0.73 | 1.25 |
| LOLIPOP Other SA | 0.95 | 0.78 | 1.16 |
| SINDI Singapore Indian | 0.96 | 0.92 | 1.00 |
| RIKEN Japan | 1.00 | 0.93 | 1.08 |
| SINDI Singapore Chinese | 0.94 | 0.84 | 1.06 |
| SINDI Singapore Malay | 0.99 | 0.96 | 1.02 |
| AGEN Korea | 1.00 | 0.95 | 1.06 |
| Sikh Meta-Analysis | 0.67 | 0.58 | 0.77 |
| South Asian Meta-Analysis | 1.01 | 0.98 | 1.05 |
| Multiethnic Meta-Analysis | 1.01 | 0.99 | 1.03 |

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.

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 $19 \pm 7$ homozygous segments >1 Mb in length, with an average length of $2.0 \pm 0.95$ 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* gene (rs9552911, $P = 1.82 \times 10^{-8}$) 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). Founder mutations in *SGCG* that cause LGMD2C predate migration of the Romani gypsies of Europe out of India around 1100 AD (49). Due to complete endogamy, this genetically isolated community had an increased incidence of autosomal recessive LGMD2C. *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**
GWAS, replication, and meta-analysis results of T2D loci identified in Punjabi Sikhs

| SNP | Chromosome position | Nearest gene | Effect/other allele | EAF Sikh | EAF CEU | Punjabi Sikh Discovery GWAS (stage 1) OR (95% CI), P value | Punjabi Sikh Replication (stages 2a, 3a) OR (95% CI), P value | Other South Asian Replication (stages 2b, 3b) OR (95% CI), P value | East Asian Evaluation (stage 3c) OR (95% CI), P value | European Evaluation (stage 3d) OR (95% CI), P value | Punjabi meta-analysis OR (95% CI), P value | South Asian meta-analysis OR (95% CI), P value | Multiethnic meta-analysis OR (95% CI), P value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Novel T2D locus** | | | | | | | | | | | | | |
| rs9552911 | 13 | 23864657 | SGCG | A/G | 0.08 | 0 | 0.61 (0.47–0.80), $P = 3.08 \times 10^{-4}$ | 0.69 (0.59–0.82), $P = 1.14 \times 10^{-5}$ | 1.04 (1.00–1.07), $P = 0.037$ | 1.00 (0.98–1.03), $P = 0.77$ | — | 0.67 (0.58–0.77), $P = 1.82 \times 10^{-8}$ | 1.01 (0.98–1.05), $P = 0.47$ | 1.01 (0.99–1.03), $P = 0.49$ |
| **Previously reported T2D loci** | | | | | | | | | | | | | |
| rs1470579 | 3 | 185511687 | IGF2BP2 | A/C | 0.59 | 0.5 | 0.76 (0.66–0.88), $P = 2.53 \times 10^{-4}$ | 0.87 (0.8–0.95), $P = 1.87 \times 10^{-3}$ | 0.95 (0.93–0.97), $P = 1.79 \times 10^{-6}$ | 0.94 (0.92–0.97), $P = 7.21 \times 10^{-6}$ | 0.88 (0.92–0.84), $P = 2.17 \times 10^{-6}$ | 0.84 (0.78–0.9), $P = 5.02 \times 10^{-6}$ | 0.94 (0.92–0.96), $P = 4.19 \times 10^{-9}$ | 0.94 (0.92–0.95), $P = 1.81 \times 10^{-19}$ |
| rs7903146 | 10 | 114758349 | TCF7L2 | T/C | 0.35 | 0.28 | 1.31 (1.13–1.52), $P = 3.23 \times 10^{-4}$ | 1.50 (1.36–1.65), $P = 7.83 \times 10^{-17}$ | 1.13 (1.10–1.16), $P = 6.12 \times 10^{-25}$ | 1.12 (1.06–1.19), $P = 1.06 \times 10^{-4}$ | 1.40 (1.34–1.46), $P = 2.21 \times 10^{-51}$ | 1.44 (1.33–1.56), $P = 3.32 \times 10^{-19}$ | 1.15 (1.13–1.18), $P = 2.71 \times 10^{-35}$ | 1.19 (1.16–1.21), $P = 8.53 \times 10^{-75}$ |

**Novel T2D locus rs9552911**

| Punjabi Sikh only Studies | n case subjects | n control subjects | Genotype call rate | Controls HWE P value | MAF case subjects | MAF control subjects | OR (95% CI) | P value | Meta-analysis OR (95% CI) and P value | Het P value |
|---|---|---|---|---|---|---|---|---|---|---|
| Punjabi Sikhs SDS/AIDHS GWAS discovery | 842 | 774 | 1 | 0.39 | 0.06 | 0.09 | 0.61 (0.47–0.80) | 0.00031 | | |
| Punjabi Sikhs LOLIPOP Sikh GWAS | 801 | 2,017 | 1 | 0.88 | 0.07 | 0.08 | 0.75 (0.58–0.96) | 0.021 | 0.67 (0.58–0.77) $P = 1.82 \times 10^{-8}$ | 0.56 |
| Punjabi Sikhs De novo Sikh Replication | 977 | 1,493 | 0.97 | 0.049 | 0.06 | 0.09 | 0.66 (0.53–0.82) | 0.00014 | | |

All *P* values are two sided. CEU, Euro-Caucasians; EAF, effect allele frequency; Het, heterozygous.

complex including SGCG 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-γ [PPAR-γ] [1 and 2] and PPAR-α) bind to the promoter and intron 1 of the SGCG 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 SGCG 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.

## REFERENCES

1. Basnyat B, Rajapaksa LC. Cardiovascular and infectious diseases in South Asia: the double whammy. BMJ 2004;328:781
2. International Diabetes Federation. *Diabetes Atlas*. Belgium, International Diabetes Federation, 2011
3. Ghaffar A, Reddy KS, Singhi M. Burden of non-communicable diseases in South Asia. BMJ 2004;328:807–810
4. Zimmet P. Type 2 (non-insulin-dependent) diabetes—an epidemiological overview. Diabetologia 1982;22:399–411
5. McKeigue PM, Pierpoint T, Ferrie JE, Marmot MG. Relationship of glucose intolerance and hyperinsulinaemia to body fat pattern in south Asians and Europeans. Diabetologia 1992;35:785–791
6. Oldroyd J, Banerjee M, Heald A, Cruickshank K. Diabetes and ethnic minorities. Postgrad Med J 2005;81:486–490
7. Nakagami T, Qiao Q, Carstensen B, et al.; DECODE-DECODA Study Group. Age, body mass index and type 2 diabetes-associations modified by ethnicity. Diabetologia 2003;46:1063–1070
8. Zimmet PZ. Kelly West Lecture 1991. Challenges in diabetes epidemiology—from West to the rest. Diabetes Care 1992;15:232–252
9. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356–369
10. Sanghera DK, Nath SK, Ortega L, et al. TCF7L2 polymorphisms are associated with type 2 diabetes in Khatri Sikhs from North India: genetic variation affects lipid levels. Ann Hum Genet 2008;72:499–509
11. 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
12. Hu C, Zhang R, Wang C, et al. A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population. Diabetologia 2009;52:451–456
13. 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
14. Ng MC, Park KS, Oh B, et al. Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. Diabetes 2008;57:2226–2233
15. Sanghera DK, Been L, Ortega L, et al. Testing the association of novel meta-analysis-derived diabetes risk genes with type II diabetes and related metabolic traits in Asian Indian Sikhs. J Hum Genet 2009;54:162–168
16. Omori S, Tanaka Y, Horikoshi M, et al. Replication study for the association of new meta-analysis-derived risk loci with susceptibility to type 2 diabetes in 6,244 Japanese individuals. Diabetologia 2009;52:1554–1560
17. Schleinitz D, Tönjes A, Böttcher Y, et al. Lack of significant effects of the type 2 diabetes susceptibility loci JAZF1, CDC123/CAMK1D, NOTCH2, ADAMTS9, THADA, and TSPAN8/LGR5 on diabetes and quantitative metabolic traits. Horm Metab Res 2010;42:14–22
18. Rees SD, Hydrie MZ, Shera AS, et al. Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations. Diabetologia 2011;54:1368–1374
19. Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet 2010;42:864–868
20. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet 2008;40:1098–1102

21. Cho YS, Chen CH, Hu C, et al.; DIAGRAM Consortium; MuTHER Consortium. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 2012;44:67–72

22. Kooner JS, Saleheen D, Sim X, et al.; DIAGRAM; MuTHER. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 2011;43:984–989

23. Kivisild T, Bamshad MJ, Kaldma K, et al. Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. Curr Biol 1999;9:1331–1334

24. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. Nat Rev Genet 2008;9:9–14

25. Singh K, Bhalla V, Kaul L. *People of India*. Vol. X. Oxford, Oxford University Press, 1994, p. 1–3

26. Sanghera DK, Bhatti JS, Bhatti GK, et al. The Khatri Sikh Diabetes Study (SDS): study design, methodology, sample collection, and initial results. Hum Biol 2006;78:43–63

27. Sanghera DK, Demirci FY, Been L, et al. PPARG and ADIPOQ gene polymorphisms increase type 2 diabetes mellitus risk in Asian Indian Sikhs: Pro12Ala still remains as the strongest predictor. Metabolism 2010;59:492–501

28. Been LF, Hatfield JL, Shankar A, et al. A low frequency variant within the GWAS locus of MTNR1B affects fasting glucose concentrations: Genetic risk is modulated by obesity. Nutr Metab Cardiovasc Dis 2012; 22:944–951

29. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2004;27(Suppl. 1):S5–S10

30. McKeigue PM, Marmot MG, Syndercombe Court YD, Cottier DE, Rahman S, Riemersma RA. Diabetes, hyperinsulinaemia, and coronary risk factors in Bangladeshis in east London. Br Heart J 1988;60:390–396

31. Lavanya R, Jeganathan VS, Zheng Y, et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. Ophthalmic Epidemiol 2009;16:325–336

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

33. Rees SD, Islam M, Hydrie MZ, et al. An FTO variant is associated with type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. Diabet Med 2011;28:673–680

34. Katulanda P, Constantine GR, Mahesh JG, et al. Prevalence and projections of diabetes and pre-diabetes in adults in Sri Lanka—Sri Lanka Diabetes, Cardiovascular Study (SLDCS). Diabet Med 2008;25:1062–1069

35. Gille H, Strahl T, Shaw PE. Activation of ternary complex factor Elk-1 by stress-activated protein kinases. Curr Biol 1995;5:1191–1200

36. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589

37. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215

38. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906–913

39. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda) 2011;1:457–470

40. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;11:499–511

41. Nalls MA, Simon-Sanchez J, Gibbs JR, et al. Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. PLoS Genet 2009;5:e1000415

42. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191

43. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885

44. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105–116

45. Saxena R, Hivert MF, Langenberg C, et al.; GIANT consortium; MAGIC investigators. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 2010;42:142–148

46. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature 2009;461:489–494

47. Avondo F, Roncaglia P, Crescenzio N, et al. Fibroblasts from patients with Diamond-Blackfan anaemia show abnormal expression of genes involved in protein synthesis, amino acid metabolism and cancer. BMC Genomics 2009;10:442

48. Barresi R, Moore SA, Stolle CA, Mendell JR, Campbell KP. Expression of gamma-sarcoglycan in smooth muscle and its interaction with the smooth muscle sarcoglycan-sarcospan complex. J Biol Chem 2000;275:38554–38560

49. Piccolo F, Jeanpierre M, Leturcq F, et al. A founder mutation in the gamma-sarcoglycan gene of gypsies possibly predating their migration out of India. Hum Mol Genet 1996;5:2019–2022

50. Goldstein JA, Kelly SM, LoPresti PP, et al. SMAD signaling drives heart and muscle dysfunction in a *Drosophila* model of muscular dystrophy. Hum Mol Genet 2011;20:894–904

51. Townsend D, Yasuda S, McNally E, Metzger JM. Distinct pathophysiological mechanisms of cardiomyopathy in hearts lacking dystrophin or the sarcoglycan complex. FASEB J 2011;25:3106–3114

52. Wheeler MT, Korcarz CE, Collins KA, et al. Secondary coronary artery vasospasm promotes cardiomyopathy progression. Am J Pathol 2004;164:1063–1071

53. Hack AA, Ly CT, Jiang F, et al. Gamma-sarcoglycan deficiency leads to muscle membrane defects and apoptosis independent of dystrophin. J Cell Biol 1998;142:1279–1287

54. Groh S, Zong H, Goddeeris MM, et al. Sarcoglycan complex: implications for metabolic defects in muscular dystrophies. J Biol Chem 2009;284:19178–19182

55. Bamshad M, Kivisild T, Watkins WS, et al. Genetic evidence on the origins of Indian caste populations. Genome Res 2001;11:994–1004