

Association analysis of 31 common polymorphisms with type 2 diabetes and its related traits in Indian sib pairs

V. Gupta · D. G. Vinay · S. Rafiq · M. V. Kranthikumar · C. S. Janipalli · C. Giambartolomei · D. M. Evans · K. R. Mani · M. N. Sandeep · A. E. Taylor · S. Kinra · R. M. Sullivan · L. Bowen · N. J. Timpson · G. D. Smith · F. Dudbridge · D. Prabhakaran · Y. Ben-Shlomo · K. S. Reddy · S. Ebrahim · G. R. Chandak · for the Indian Migration Study Group

Received: 11 August 2011 / Accepted: 30 September 2011 / Published online: 4 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract

Aims/hypothesis Evaluation of the association of 31 common single nucleotide polymorphisms (SNPs) with fasting glucose, fasting insulin, HOMA-beta cell function (HOMA- β), HOMA-insulin resistance (HOMA-IR) and type 2 diabetes in the Indian population.

Methods We genotyped 3,089 sib pairs recruited in the Indian Migration Study from four cities in India (Lucknow, Nagpur, Hyderabad and Bangalore) for 31 SNPs in 24

genes previously associated with type 2 diabetes in European populations. We conducted within-sib-pair analysis for type 2 diabetes and its related quantitative traits.

Results The risk-allele frequencies of all the SNPs were comparable with those reported in western populations. We demonstrated significant associations of *CXCR4* (rs932206), *CDKAL1* (rs7756992) and *TCF7L2* (rs7903146, rs12255372) with fasting glucose, with β values of 0.007 ($p=0.05$), 0.01 ($p=0.01$), 0.007 ($p=0.05$),

Members of the Indian Migration Study Group are shown in the [Appendix](#).

Electronic supplementary material The online version of this article (doi:10.1007/s00125-011-2355-6) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

V. Gupta (✉) · K. S. Reddy · S. Ebrahim
South Asia Network for Chronic Disease,
Public Health Foundation of India,
C-1/52, Safdarjung Development Area,
New Delhi 110016, India
e-mail: udaiig@gmail.com

D. G. Vinay · M. V. Kranthikumar · C. S. Janipalli · K. R. Mani ·
M. N. Sandeep · G. R. Chandak (✉)
Centre for Cellular and Molecular Biology (CCMB),
Council of Scientific and Industrial Research (CSIR),
Habshiguda, Uppal Road,
Hyderabad 500007, India
e-mail: chandakgrc@ccmb.res.in

S. Rafiq · C. Giambartolomei · S. Kinra · R. M. Sullivan ·
L. Bowen · F. Dudbridge · S. Ebrahim
Faculty of Epidemiology and Population Health,
London School of Hygiene and Tropical Medicine,
London, UK

D. M. Evans · A. E. Taylor · N. J. Timpson · G. D. Smith ·
Y. Ben-Shlomo
School of Social and Community Medicine, University of Bristol,
Bristol, UK

D. M. Evans · A. E. Taylor · N. J. Timpson · G. D. Smith
MRC Centre for Causal Analyses in Translational Epidemiology,
University of Bristol,
Bristol, UK

F. Dudbridge · S. Ebrahim
Bloomsbury Centre for Genetic Epidemiology and Statistics,
London, UK

D. Prabhakaran
Centre for Chronic Disease Control,
New Delhi, India

V. Gupta · K. S. Reddy
Public Health Foundation of India,
New Delhi, India

0.01 ($p=0.003$) and 0.08 ($p=0.01$), respectively. Variants in *NOTCH2* (rs10923931), *TCF-2* (also known as *HNF1B*) (rs757210), *ADAM30* (rs2641348) and *CDKN2A/B* (rs10811661) significantly predicted fasting insulin, with β values of -0.06 ($p=0.04$), 0.05 ($p=0.05$), -0.08 ($p=0.01$) and -0.08 ($p=0.02$), respectively. For HOMA-IR, we detected associations with *TCF-2*, *ADAM30* and *CDKN2A/B*, with β values of 0.05 ($p=0.04$), -0.07 ($p=0.03$) and -0.08 ($p=0.02$), respectively. We also found significant associations of *ADAM30* ($\beta=-0.05$; $p=0.01$) and *CDKN2A/B* ($\beta=-0.05$; $p=0.03$) with HOMA- β . *THADA* variant (rs7578597) was associated with type 2 diabetes (OR 1.5; 95% CI 1.04, 2.22; $p=0.03$).

Conclusions/interpretation We validated the association of seven established loci with intermediate traits related to type 2 diabetes in an Indian population using a design resistant to population stratification.

Keywords Association · India · Intermediate traits · Polymorphisms · Type 2 diabetes

Abbreviations

AIIMS	All India Institute of Medical Sciences
GWAS	Genome-wide association study
LD	Linkage disequilibrium
MAF	Minor allele frequency
SNP	Single nucleotide polymorphism

Introduction

Type 2 diabetes is an outcome of chronic hyperglycaemia and is clinically characterised by impaired insulin secretion and insulin resistance [1, 2]. With 51 million individuals with type 2 diabetes [3], India could easily become an epicentre for research into the detection of common biological mechanisms through the genomics of diabetes and its related quantitative traits and their pathological phenotypic extensions. Several genome-wide association studies (GWASs) have confirmed associations of at least 38 common genetic variations with type 2 diabetes and related intermediate traits and many more are anticipated with growing sample sizes and better array performance [4–6]. In the light of the above discoveries, it is interesting to investigate whether the top GWAS signals in European populations are also associated with diabetes in populations of different ancestry, such as South Asians [7]. Therefore, validation of single nucleotide polymorphisms (SNPs) from different loci identified through GWASs is now an important step towards genomic medicine.

In India most genetic studies of type 2 diabetes published to date have used hospital-based case–control

designs for either establishing candidate genes in the metabolic pathways of interest [8, 9] or validating GWAS findings [10–13]. Further, the studies on quantitative traits of type 2 diabetes in India are few [14, 15] and have investigated a limited number of SNPs. For instance, Chauhan et al. [15] have recently validated two loci (*TCF7L2* and *PPARG*) associated with HOMA-beta cell function (HOMA- β) from eight loci tested in 2,500 control participants from the Indo-European linguistic group of India.

A high level of population stratification and ecological diversity has complicated genetic epidemiological studies in India. It has been estimated that the population differentiation between two neighbouring populations in India may be as large as that between northern and southern Europeans (average fixation index [F_{ST}] value of 0.0109 compared with $F_{ST}=0.003$ for northern and southern Europeans) [16]. Moreover, the endogamous structure of the Indian population has maintained a high level of linkage disequilibrium (LD) because of ancestral founder effects [16]. It has been suggested that genetic predisposition to diabetes may differ in Indians compared with Europeans [15]; however, in view of the endogamous structure and the high possibility of population stratification, these results need to be interpreted with caution.

As very few studies in India have tested the association of type 2 diabetes SNPs with related quantitative traits, we conducted the present study to evaluate the relationships between 31 GWAS-confirmed type 2 diabetes SNPs from 24 genes with type 2 diabetes and four available intermediate traits: fasting glucose; fasting insulin; HOMA-insulin resistance (HOMA-IR); and HOMA- β . We used a population-based cohort of 3,089 sib pairs (2,528 for quantitative traits and 561 for type 2 diabetes) drawn from four cities across India, a design resistant to population stratification as comparisons are made within sib pairs.

Methods

Study population The study included individuals recruited to the population-based Indian Migration Study (IMS) conducted during 2005–2007 in factories located in four cities (Lucknow, Nagpur, Hyderabad and Bangalore) of India. Briefly, each migrant worker and spouse were asked to invite one non-migrant full sibling of the same sex and closest to them in age still residing in their rural place of origin to take part in the study. This strategy resulted in rural-dwelling sibs being drawn from 20 of the 28 states in India, reflecting the migration patterns of the factory workforce and their spouses (for details see the Electronic supplementary material [ESM text]). Phenotypic information was available for 7,067 participants, of whom 6,780

individuals were true sib pairs. All the information was collected after receiving informed consent from the study participants. Ethical approval was obtained from the ethics committee of the All India Institute of Medical Sciences (AIIMS), New Delhi, India (reference number A-60/4/8/2004).

Biochemical phenotypes After the separation of plasma and serum, samples were transported monthly to AIIMS for the biochemical assays. Fasting plasma glucose was measured on the day of blood collection by local laboratories at each site using the GOD-PAP method and RANDOX kits (Randox Laboratories, Crumlin, UK) [17]. Fasting insulin was assayed in serum samples by the ELISA method, as a solid-phase two-site enzyme immunoassay, using kits from MERCODIA (Merckodia AB, Sylveniusgatan, Uppsala, Sweden) [18]. The quality of local assays was monitored by AIIMS, and checked regularly with external standards and internal duplicate assays. For quality assurance purposes, the Cardiac Biochemistry Laboratory, AIIMS, is part of the UK National External Quality Assessment (www.ukneqas.org.uk/). HOMA-IR and HOMA- β scores were calculated using standard equations [19]. Participants with fasting glucose ≥ 7 mmol/l or already diagnosed with diabetes (with or without medication) were classified as type 2 diabetes.

Genotyping and quality control Genomic DNA was isolated from all samples using the salt precipitation method and DNA samples were plated in 96-deep-well storage plates at a uniform concentration of 10 ng/ λ at the Centre for Cellular and Molecular Biology, Hyderabad, India. Each plate included eight repeat samples (~10%) as a quality control measure. We used the Sequenom Mass ARRAY technology to genotype 31 type-2-diabetes-related SNPs from 24 loci, which were discovered in various GWASs [20–23] as part of a common multiplex pool. The study was proposed in 2007 and started in 2008; therefore the SNPs included in the genotyping panel are part of GWAS results published up to that time. We also included some SNPs with p values $> 1 \times 10^{-7}$ to validate their associations in the Indian population. The genotyping success rate was $> 95\%$ and results of duplicate samples had $> 97\%$ concordance. The final analysis for quantitative traits was done on 2,528 sib pairs (5,056 individuals) because 2,011 of the 7,067 individuals with phenotypic information were excluded from the analysis (i.e. 301 pairs [602 individuals] had genotyping data for fewer than 50% of the 31 genotyped SNPs; 136 pairs [272 individuals] were not true siblings; 15 individuals were singletons; and 561 pairs [1,122 individuals] were concordant/discordant for type 2 diabetes). Type 2 diabetes as a binary trait was investigated using data from 561 pairs (1,122 individuals).

Sample size and power calculation Given a sample size of 2,528 sib pairs and minor allele frequency (MAF)=9–48%, our study had $\geq 80\%$ power to detect a quantitative trait locus explaining 0.7% of genetic variation at $\alpha=0.05$ and 1.5% of genetic variation at $\alpha=0.001$, for each of the tested traits. This was plausible in the context of genetic effect estimates observed for quantitative traits in European populations, which range from $r^2 < 0.001$ to $r^2 = 0.013$ (β range 0.008 to -0.28 , ESM Table 1). For type 2 diabetes as a binary outcome, based on the 561 sib pairs (487 discordant and 74 concordant pairs), we had 20% power to detect an OR of 1.17 (the average European effect size) given MAF of 23% (average MAF in this study) (ESM Tables 2 and 3). The above calculations were performed using the Genetic Power Calculator (<http://ibgwww.colorado.edu/~pshaun/gpc/>; option of ‘QTL association for sibships’) [24] for quantitative traits, and QUANTO 1.1 (<http://hydra.usc.edu/gxe/>) for type 2 diabetes as a binary trait (using the case–sibling option).

Statistical analyses Genotype frequencies were calculated in unrelated individuals residing in an urban location (a single randomly chosen member from each family) ($n=2,148$) and tested for departure from Hardy–Weinberg equilibrium using the exact test implemented in PLINK (version 0.99p; <http://pngu.mgh.harvard.edu/~purcell/PLINK/>) [25].

Association analysis was carried out using the orthogonal family-based model of Fulker et al. [26]. This is a mixed regression model in which the genetic effect is decomposed into fixed between- and within-family effects, with inference performed on the within-family effect. Shared environmental and residual genetic effects within families are modelled as a random effect. We fitted models in STATA (version 10) as the QTDT software [27], widely used for the Fulker model, is limited to analysis of quantitative traits. For drawing inferences, we have used the results after adjustment for BMI because adjustment with this covariate has reduced the residual variance in the regression model and provided more significant findings in comparison with the unadjusted model. In the ESM text we describe additional analyses of population structure and corresponding adjustments to association tests. Although the corrected overall critical p value using a false-discovery rate for each trait is equal to 0.002 [28], we have drawn inferences on the basis of uncorrected p values because there was strong prior evidence of associations of all the examined SNPs with type 2 diabetes [20–23].

Assuming additive models, we restricted the analysis to full sib pairs ($n=2,528$ pairs) after removing the pairs concordant/discordant for type 2 diabetes ($n=561$ pairs); for each SNP the major allele was the reference, and an effect estimate was calculated per copy of the minor allele.

Associations between SNPs and the natural log transformed z scores of quantitative phenotypes were tested after including age, sex and location (urban or rural) as covariates. We also tested separately for interaction between- and within-families genetic effect and age, sex and location (rural/urban) as effect modifiers in the models for quantitative traits.

For association analysis with the discrete trait of type 2 diabetes, we used a mixed logistic regression model using the same Fulker's decomposition of genotypes on 561 sib pairs (see ESM Table 4).

Results

Estimates of MAFs for 29 of the tested SNPs were available from the HapMap-GIH (Gujarati Indians in Houston) population (<http://hapmap.ncbi.nlm.nih.gov>) and were found to be similar to those in the study sample (ESM Table 2). The summary characteristics of the study participants are given in Table 1 (see also ESM Table 4). As age, sex and location (urban/rural) were associated with type 2 diabetes [29], these covariates were included in further association models.

We found that 7 of the 31 SNPs were associated with at least one of the four continuous traits related to type 2 diabetes after adjusting for BMI (Table 2; see also ESM Table 5 for unadjusted results).

Association analysis with fasting insulin and HOMA-IR Using within family-based approach, we found a variant of *ADAM30* (ADAM metalloproteinase domain 30, rs2641348, G allele) was associated with lower levels of fasting insulin ($\beta=-0.08$, 95% CI -0.02 , -0.02 , $p=0.009$) and HOMA-IR ($\beta=-0.07$, 95% CI -0.13 , -0.006 , $p=0.03$). Similarly, we also found an association of the

CDKN2A/B locus (rs10811661, C allele) with lower levels of fasting insulin ($\beta=-0.09$, 95% CI -0.16 , -0.02 , $p=0.02$) and HOMA-IR ($\beta=-0.09$, 95% CI -0.16 , -0.02 , $p=0.02$). Further, the T allele at rs10923931 variant in Notch homologue 2 (*NOTCH2*) showed a decreased level of fasting insulin ($\beta=-0.06$, 95% CI -0.12 , -0.001 , $p=0.04$), whereas, *TCF-2* (also known as *HNF1B*) (rs757210, A allele) was associated with higher levels of fasting insulin ($\beta=0.05$, 95% CI -0.001 , 0.11 , $p=0.05$) and HOMA-IR ($\beta=0.05$, 95% CI 0.003 , 0.11 , $p=0.04$). We observed a strong LD ($r^2>0.80$; range 0.81–0.92) between *ADAM30* and *NOTCH2* in the study population (ESM Table 6).

Association analysis with fasting glucose and HOMA- β We observed an association of the variant of the *CDKALI* locus (rs7756992, G allele) with an increased level of fasting glucose ($\beta=0.009$, 95% CI 0.002 , 0.02 , $p=0.01$). In addition, variants at the T allele at *TCF7L2* (transcription factor 7 like 2), rs7903146 ($\beta=0.007$, 95% CI -0.0001 , 0.01 , $p=0.05$) and rs12255372 ($\beta=0.01$, 95% CI 0.004 , 0.02 , $p=0.003$) predicted higher fasting glucose levels. In contrast, the *CXCR4* locus (rs932206, T allele) was associated with a lower level of fasting glucose ($\beta=0.009$, 95% CI -0.02 , -0.0005 , $p=0.04$), while we found association of *ADAM30* (rs2641348, G allele) ($\beta=-0.05$, 95% CI -0.10 , -0.01 , $p=0.01$) and the *CDKN2A/B* locus (rs10811661, C allele) ($\beta=-0.05$, 95% CI -0.11 , -0.004 , $p=0.03$) with reduced beta cell function.

Association analysis with type 2 diabetes In the present study, of the 31 SNPs tested for association with type 2 diabetes (ESM Table 2), only a variant in the *THADA* locus (rs7578597) was found to be associated with OR 1.5 (95% CI 1.04, 2.22, $p=0.03$) and this association remained significant even after adjustment for BMI. The remaining SNPs in this study had estimated ORs of 1.1 or less (ESM

Table 1 Characteristics of participants in the Indian Migration Study

Characteristic	Participants with diabetes	Participants without diabetes	p value ^a (t test)	Total participants
Sample size	635*	5,056	–	6,178
Men (%)	61.6	57.5	–	57.9
Age (years)	47.7 \pm 7.2	40.0 \pm 10.3	<0.0001	40.8 \pm 10.3
Age at diagnosis (years)	43.2 \pm 7.3	–	–	–
BMI (kg/m ²)	26.2 \pm 4.0	23.5 \pm 4.4	<0.0001	23.8 \pm 4.5
Fasting glucose (mmol/l)	7.8 \pm 3.1	5.0 \pm 0.6	<0.0001	5.3 \pm 1.4
Fasting insulin (pmol/l)	71.7 \pm 87.8	53.0 \pm 106.4	<0.0001	54.7 \pm 101.1
HOMA- β	57.6 \pm 49.9	88.8 \pm 60.2	<0.0001	85.6 \pm 59.9
HOMA-IR	1.49 \pm 2.05	0.95 \pm 0.95	<0.0001	1.0 \pm 1.12

Values are presented as mean \pm SD

^a Among 635 individuals, 487 belonged to pairs discordant for diabetes, and 148 individuals from pairs concordant for diabetes (i.e. 74 pairs) (total number of pairs discordant and concordant for diabetes was 561 [487+74])

Table 2 Within-sib-pair association estimates for quantitative traits related to type 2 diabetes (adjusted for BMI)

SNP (minor allele)	Gene	β coefficient ^a (95% CI), <i>p</i> value ^b		
		Fasting glucose (<i>n</i> =5,056)	Fasting insulin (<i>n</i> =5,054)	HOMA-IR (<i>n</i> =5,053)
rs1799854 (T)	<i>ABCC8</i>	-0.006 (-0.008, 0.006), 0.86	0.002 (-0.05, 0.06), 0.94	0.002 (-0.05, 0.05), 0.93
rs2641348 (G)	<i>ADAM30</i>	0.008 (-0.007, 0.009), 0.85	-0.08* (-0.15, -0.02), 0.01	-0.07* (-0.12, -0.005), 0.03
rs10490072 (C)	<i>BCL11A</i>	0.001 (-0.01, 0.012), 0.86	-0.04 (-0.13, 0.05), 0.37	-0.04 (-0.12, 0.04), 0.36
rs12779790 (G)	<i>CD123/CAMK1D</i>	0.005 (-0.004, 0.01), 0.28	0.008 (-0.06, 0.08), 0.82	0.007 (-0.06, 0.08), 0.84
rs10811661 (C)	<i>CDKN2A/B</i>	-0.003 (-0.01, 0.007), 0.58	-0.08* (-0.15, -0.01), 0.02	-0.08* (-0.15, -0.01), 0.02
rs10946398 (C)	<i>CDK4LI</i>	0.005 (-0.003, 0.01), 0.23	0.02 (-0.04, 0.08), 0.55	0.02 (-0.03, 0.08), 0.41
rs7754840 (C)	<i>CDK4LI</i>	0.004 (-0.003, 0.01), 0.24	0.01 (-0.04, 0.07), 0.61	0.02 (-0.03, 0.08), 0.43
rs7756992 (G)	<i>CDK4LI</i>	0.01* (0.002, 0.02), 0.01	-0.006 (-0.06, 0.05), 0.84	0.001 (-0.05, 0.06), 0.96
rs932206 (T)	<i>CXCR4</i>	-0.009* (-0.02, -0.0005), 0.04	-0.004 (-0.08, 0.07), 0.90	-0.02 (-0.09, 0.04), 0.49
rs1153188 (A)	<i>DGD</i>	-0.005 (-0.013, 0.003), 0.22	-0.04 (-0.10, 0.02), 0.24	-0.04 (-0.10, 0.02), 0.19
rs17044137 (A)	<i>FLJ39370^d</i>	0.003 (-0.007, 0.01), 0.53	-0.03 (-0.11, 0.05), 0.50	-0.02 (-0.11, 0.05), 0.53
rs9939609 (A)	<i>FTO</i>	-0.007 (-0.014, 0.0002), 0.06	0.01 (-0.04, 0.07), 0.65	0.01 (-0.04, 0.06), 0.65
rs5015480 (C)	<i>HHEX</i>	-0.004 (-0.01, 0.002), 0.20	0.009 (-0.04, 0.06), 0.71	0.01 (-0.04, 0.06), 0.63
rs1111875 (G)	<i>HHEX</i>	0.003 (-0.003, 0.01), 0.31	-0.02 (-0.07, 0.03), 0.53	-0.02 (-0.07, 0.03), 0.48
rs7923837 (G)	<i>HHEX</i>	0.004 (-0.003, 0.01), 0.26	-0.01 (-0.06, 0.04), 0.64	-0.01 (-0.06, 0.04), 0.63
rs2268573 (G)	<i>GCK</i>	0.004 (-0.002, 0.01), 0.23	0.03 (-0.02, 0.08), 0.20	0.03 (-0.02, 0.08), 0.28
rs2377892 (T)	<i>KCNQ1</i>	0.006 (-0.02, 0.03), 0.70	-0.04 (-0.26, 0.18), 0.73	-0.03 (-0.25, 0.18), 0.77
rs2876711 (C)	<i>KCTD12</i>	0.001 (-0.006, 0.008), 0.76	-0.02 (-0.07, 0.03), 0.47	-0.01 (-0.07, 0.04), 0.56
rs1256517 (C)	<i>LOC646279</i>	0.007 (-0.003, 0.02), 0.17	-0.03 (-0.11, 0.04), 0.37	-0.03 (-0.10, 0.04), 0.45
rs10823406 (A)	<i>NGN3</i>	-0.0003 (-0.008, 0.008), 0.99	-0.02 (-0.08, 0.04), 0.58	-0.02 (-0.08, 0.04), 0.48
rs10923931 (T)	<i>NOTCH2</i>	-0.0001 (-0.008, 0.008), 0.97	-0.06* (-0.12, -0.001), 0.04	-0.05 (-0.11, 0.008), 0.09
rs1801282 (G)	<i>PPARG</i>	-0.005 (-0.02, 0.005), 0.33	-0.01 (-0.09, 0.06), 0.72	-0.009 (-0.09, 0.07), 0.81
rs3856806 (T)	<i>PPARG</i>	-0.005 (-0.01, 0.004), 0.27	0.01 (-0.05, 0.08), 0.66	0.02 (-0.05, 0.09), 0.57
rs13266634 (T)	<i>SLC30A8</i>	-0.005 (-0.01, 0.003), 0.20	-0.006 (-0.07, 0.06), 0.84	-0.01 (-0.07, 0.05), 0.70
rs7578597 (T ^c)	<i>THADA</i>	-0.003 (-0.012, 0.006), 0.53	-0.02 (-0.09, 0.05), 0.57	-0.015 (-0.08, 0.05), 0.66
rs757210 (A)	<i>TCF2</i>	-0.001 (-0.008, 0.006), 0.78	0.05* (-0.001, 0.11), 0.05	0.05* (0.003, 0.11), 0.04
rs7903146 (T)	<i>TCF7L2</i>	0.007* (-0.0001, 0.01), 0.05	0.009 (-0.05, 0.06), 0.75	0.006 (-0.05, 0.06), 0.82
rs12255372 (T)	<i>TCF7L2</i>	0.01* (0.004, 0.02), 0.003	-0.009 (-0.06, 0.06), 0.98	-0.009 (-0.06, 0.06), 0.97
rs7961581 (C)	<i>TSPAN8/LGR5</i>	-0.0003 (-0.007, 0.007), 0.93	-0.02 (-0.08, 0.03), 0.39	-0.02 (-0.07, 0.03), 0.41
rs9472138 (T)	<i>VEGFA</i>	-0.002 (-0.01, 0.007), 0.68	-0.03 (-0.10, 0.04), 0.38	-0.02 (-0.09, 0.05), 0.57
rs10010131 (A)	<i>WFS1</i>	0.001 (-0.007, 0.009), 0.79	0.01 (-0.05, 0.07), 0.74	0.004 (-0.05, 0.06), 0.88

Significant association: **p*≤0.05

^a Within-sib-pairs association estimate model adjusted for three covariates (age, sex and location); all values are represented as β (95% CI), where β =coefficient from the linear regression on log-transformed *z* scores

^b Corrected overall critical *p* value using false-discovery rate for each trait=0.002

^c Major allele of locus *THADA* (for rest of the loci minor allele was taken as reference)

^d Also known as *C4ORF32*

Table 2) and we had insufficient power (because of the population-based design) to detect such a low effect size.

After applying corrections for 31 SNPs ($p < 0.002$ using the false discovery rate), we did not find any significant association with the studied phenotypes (fasting glucose, fasting insulin, HOMA-IR, HOMA- β). For comparison, we performed an additional analysis using Fulker's between-sib-pair model for total association with adjustment for a limited number of population clusters identified by STRUCTURE [30] (see ESM Fig. 1), but these results did not differ appreciably from the within-family association.

Discussion

We report associations of *ADAM30* and *CDKN2A/B* with fasting insulin levels, HOMA-IR and HOMA- β in India. Further, we report the associations of *TCF-2* and *NOTCH2* with fasting insulin level, and *TCF-2* with HOMA-IR. We also detected that *CXCR4*, *CDKAL1* and *TCF7L2* variants influence fasting glucose levels in the study population. In addition, we found the association of *THADA* with type 2 diabetes in the Indian population. The consistency of detecting association of these regions in the Indian population shows the importance of GWAS signals in quantifying the effects of type 2 diabetes.

In order to understand the mechanisms involved in beta cell function and glucose homeostasis, and to complement the genetic analysis of type 2 diabetes, large-scale meta-analysis of GWASs on continuous glycaemic phenotypes has been performed in the last 2 years [4, 5, 31]. These studies have found new genetic variants associated with quantitative metabolic phenotypes and have increased our understanding of the pathophysiology of type 2 diabetes. In addition, several studies have also shown positive associations of genetic variants related to type 2 diabetes with the continuous traits of this disease [32–34]. In India, few type 2 diabetes SNPs have been evaluated with quantitative phenotypes, so the relevance of these SNPs for glycaemic traits and their relation to the pathogenesis of type 2 diabetes is unclear for this population.

Of the 31 SNPs tested for association, we found that the minor alleles of *ADAM30* and *CDKN2A/B* variants predict reduced fasting insulin, HOMA-IR and HOMA- β , suggesting their role in insulin sensitivity and beta cell dysfunction. Several studies did not find any association of *ADAM30* with quantitative traits related to type 2 diabetes [33–35], whereas for *CDKN2A/B* few studies have shown evidence of impaired [36] and improved [37] beta cell function. Similar to European population studies, we also observed a strong LD between *ADAM30* and *NOTCH2* [23], and an association of *NOTCH2* with lower levels of fasting insulin [5]. Importantly, the variant,

rs2641348 in *ADAM30* is a non-synonymous SNP (L359P) and could be directly evaluated for functional genomic studies to determine the mechanism regulating insulin secretion and resistance [23]. Chauhan et al. [15] studied *CDKN2A/B* in an Indian population but did not observe an association with type-2-diabetes-related intermediate traits. Only one small study from India has shown an association of *CDC123/CAMK1D* with fasting insulin levels in a Sikh population, but was unable to find any evidence for *CDKN2A/B* and *NOTCH2*, probably because of the low power of the study [14]. *CDKN2A/B* has a role in pancreatic islets [38], and possibly influences insulin secretion by decreasing beta cell mass [36].

Gudmundsson et al. [39] observed the protective role of *TCF-2* variant against type 2 diabetes among Europeans and confirmed their findings in African and Chinese populations; however, these observations could not be replicated in an independent study in a Han Chinese population [40, 41]. Notably, we have reported the association of *TCF-2* with higher levels of fasting insulin and HOMA-IR, suggesting a role in insulin resistance.

CDKAL1 plays an important role in insulin secretion from pancreatic beta cells [42]. In a recent large meta-analysis of GWAS studies in non-diabetic individuals from European populations, *CDKAL1* was not associated with any of the quantitative traits [5]. Chauhan et al. failed to find any association of *CDKAL1* with glycaemic traits in Indians [15]. However, we found an association of the *CDKAL1* locus with increasing levels of fasting glucose, as observed by Voight et al. [4]. Further, Cauchi et al. [43] found an association of *CXCR4* with type 2 diabetes in a French population, and in the present study, we found an association of the *CXCR4* locus with increased levels of fasting glucose, suggesting its role in glucose homeostasis.

We also replicated in our population an association of *TCF7L2* variants with high fasting glucose as reported by Dupuis et al. [5] for Europeans. Recently, Chauhan et al. [15] found associations of eight SNPs with type 2 diabetes and also associations of *PPARG* and *TCF7L2* with HOMA- β . As reported elsewhere, *TCF7L2* is also the most promising gene harbouring common variations associated with type 2 diabetes in India because of the replication of results in all the studies [10–12]. However, evidence for an association with fasting glucose was positive only in the study by Chandak et al. [10], being negative in most Indian studies [11, 14, 15]. As our study has a large sample size and a design resistant to population stratification, it could therefore be seen as providing strong evidence in support of the role of *TCF7L2* in the homeostasis of fasting glucose levels in India.

Despite having reasonable power, we could replicate associations of only seven SNPs with various glycaemic phenotypes. Though difficult to explain, it is possible that

the remaining 24 SNPs might not play any role in influencing these continuous traits in Indians. An alternative, perhaps more likely, explanation is that the underlying causal variant is less well tagged by these SNPs in the Indian population owing to different allele frequencies and the extent of LD compared with Europeans.

As observed in the meta-analysis of GWASs by Zeggini et al. [23], we also demonstrated a strong effect (OR 1.5) of a *THADA* variant in predisposition to type 2 diabetes in our study population. This locus was previously studied among an Asian Sikh population, but no significant results were observed [14]. This variant of *THADA* (rs7578597) results in a missense mutation (T1187A) [23, 44] and is also known to lower measures of beta cell function [34], and hence it is a likely causal variant. The association of *THADA* with type 2 diabetes remained significant even after adjustment for BMI, suggesting an independent effect of the variant in predisposition to type 2 diabetes (ESM Table 2). In addition to low power for type 2 diabetes as a binary trait in our study, the lack of association of other SNPs may reflect their low effect sizes in this population as our sample estimates were 1.1 or lower. For instance, the loci *FTO*, *NGN3* (also known as *NEUROG3*), *TSPAN8* (p value > 0.05) could present false negative values because of the sample sizes available (ESM Table 2).

In addition, while comparing the effect sizes for type 2 diabetes and its related quantitative traits with Europeans (see ESM Tables 1 and 3), we did not find any evidence that the risk alleles from our Indian population were acting in the same direction as in the European population (see ESM Table 7). This suggests that further associations would have been present in our sample but failed to reach conventional levels of significance.

The major strengths of our study are the research design and the high number of SNPs associated with type 2 diabetes examined. In the Indian context, this is probably the largest study investigating the effect of 31 genetic variants on various intermediate traits predictive of type 2 diabetes. As the participants came from different parts of India, we can claim good representation of the Indian population and at the same time have been able to control for population stratification because of the model of sib-pair analysis. The present study is the first to report the associations of *ADAM30*, *NOTCH2*, *CDKN2A/B* and *TCF-2* with various quantitative glycaemic traits in an Indian population. The associations of *CDKAL* and *CXCR4* with fasting glucose in India are also observed for the first time in our study. We replicated the strong effects of *THADA* in predisposition of type 2 diabetes in this Indian population. Furthermore, the high level of population heterogeneity as described by Reich et al. [16], population stratification and small sample sizes have plagued case-control-based genetic association studies by providing spurious associations and inconsistent results of

replication studies. In this challenging situation, the present study has shown the benefit of a family-based design for conducting genetic epidemiological studies in India. The major limitation of our study was that, because of the cohort design, we had limited power for analysing binary traits and hence could not replicate published results for the majority of genes associated with type 2 diabetes. Although correction for multiple testing is not required for variants that already have strong priors of association with type 2 diabetes, the possibility exists that some of our weaker associations may represent false-positive findings.

In conclusion, using a large sample of sib pairs, we validated the association of seven type-2-diabetes-related loci with various continuous glycaemic traits in the Indian population. We also validated the strong association of *THADA* with type 2 diabetes in our study population. Furthermore, in the presence of a high level of population heterogeneity and substructure, we have also provided evidence of the utility of the sib-pair design for conducting genetic epidemiological studies for late-onset complex diseases in the Indian population.

Acknowledgements We are grateful to the Indian Migration Study Group, study participants and field staff for conducting the migration study. The Indian Migration Study was funded by the Wellcome Trust (grant number GR070797MF). The genetic work was funded by a project grant from the Wellcome Trust (083541/Z/07/Z). The funder had no role in study design, data collection, analysis or interpretation, writing the report, or the decision to submit the article for publication. The researchers are all independent from the funding source.

Contribution statement SE, GDS, FD, DP, YB-S, KSR, GRC, SK and IMSG contributed to the conception and design of the study. DGV, VG, MVKK, CSJ, KRM, MNS, SR, AET, SK, RMS, LB, NJT, CG, GDS, FD, DP, DME, YBS, SE and GRC contributed to the analysis and interpretation of data. VG, FD, SE, GRC and GDS were responsible for drafting the article. VG, FD, SE, GRC, YBS, DP, DGV, MVKK, CSJ, KRM, MNS, SR, DME, AET, SK, RMS, LB, NJT, CG, GDS and KSR contributed to critical revision of the paper. All authors have read the manuscript and gave their final approval for this version of the manuscript to be published.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Members of The Indian Migration Study Group: S. Ebrahim (Department of Epidemiology & Population Health, London

School of Hygiene & Tropical Medicine, London, UK and South Asia Network for Chronic Disease, Public Health Foundation of India, New Delhi, India); S. Kinra (Department of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, UK); K. S. Reddy (Public Health Foundation of India, New Delhi, India); D. Prabhakaran and T. Patel (Centre for Chronic Disease Control, New Delhi, India); L. Ramakrishnan (Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India); R. Gupta (Public Health Foundation of India, New Delhi, India); T. Lyngdoh (South Asia Network for Chronic Disease, Public Health Foundation of India, New Delhi, India); R. C. Ahuja and R. K. Saran (Department of Medicine, King George's Medical College and Institute of Clinical Epidemiology, Lucknow, India); P. Joshi and N. M. Thakre (Department of Medicine, Government Medical College, Nagpur, India); K. V. R. Sarma, S. M. Das, R. K. Jain, S. S. Potnis and S. Bhogadi (Department of Neurology, Krishna Institute of Medical Sciences, Hyderabad, India); A. V. Kurpad, M. Vaz, A. V. Barathi and M. Mohan (Department of Medicine, Dr. B.R. Ambedkar Medical College, Bangalore, India); C. S. Yajnik (KEM Hospital, Pune, India); G. D. Smith and Y. Ben-Shlomo (Department of Social Medicine, University of Bristol, Bristol).

References

- Genuth S, Alberti KGMM, Bennett P et al (2003) The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 26:3160–3167
- Prokopenko I, McCarthy MI, Lindgren CM (2008) Type 2 diabetes: new genes, new understanding. *Trends Genet* 24:613–621
- Unwin N, Whiting D, Gan D, Jacqmain O, Ghyoot G (2009) International Diabetes Federation Diabetes Atlas. IDF Diabetes Atlas, 4th edn. International Diabetes Federation, Brussels, Belgium
- Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579–589
- Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42:105–116
- Saxena R, Hivert MF, Langenberg C et al (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42:142–148
- Ioannidis JP, Thomas G, Daly MJ (2009) Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 10 (5):318–329
- Nair AK, Sugunan D, Kumar H, Anilkumar G (2010) Case-control analysis of SNPs in GLUT4, RBP4 and STRA6: association of SNPs in STRA6 with type 2 diabetes in a South Indian population. *PLoS One* 5:e11444
- Sugunan D, Nair AK, Kumar H, Gopalakrishnapillai A (2010) A case-control analysis of common variants in GIP with type 2 diabetes and related biochemical parameters in a South Indian population. *BMC Med Genet* 11:118
- Chandak GR, Janipalli CS, Bhaskar S (2007) Common variants in the TCF7L2 gene are strongly associated with T2D mellitus in the Indian population. *Diabetologia* 50:63–67
- Sanghera DK, Ortega L, Han S et al (2008) Impact of nine common T2D risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet* 9:59
- Gupta V, Khadgawat R, Ng HK, Kumar S, Aggarwal A, Rao VR, Sachdeva MP (2010) A validation study of type 2 diabetes-related variants of the TCF7L2, HHEX, KCNJ11, and ADIPOQ genes in one endogamous ethnic group of north India. *Ann Hum Genet* 74:361–368
- Chidambaram M, Radha V, Mohan V (2010) Replication of recently described type 2 diabetes gene variants in a South Indian population. *Metabolism* 59:1760–1766
- Sanghera DK, Been L, Ortega L et al (2009) 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* 54:162–168
- Chauhan G, Spurgeon CJ, Tabassum R et al (2010) Impact of common variants of PPARG, KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5,164 Indians. *Diabetes* 59:2068–2074
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461:489–494
- Trinder P (1968) Determination of glucose in blood using glucose oxidase with an alternative oxygen acceptor. *Ann Clin Biochem* 6:24–27
- Sjöstrand M, Gudbjörnsdottir S, Strindberg L, Lönnroth P (2005) Delayed transcapillary delivery of insulin to muscle interstitial fluid after oral glucose load in obese subjects. *Diabetes* 54:152–157
- Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC (1985) Homeostasis model assessment: insulin resistance and B cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412–419
- Scott LJ, Mohlke KL, Bonnycastle LL et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345
- Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for T2D and triglyceride levels. *Science* 316:1331–1336
- Sandhu MS, Weedon MN, Fawcett KA et al (2007) Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951–953
- Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645
- Purcell S, Chery SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150
- Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Fulker DW, Chery SS, Sham PC, Hewitt JK (1996) Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 64:259–267
- Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300
- Ebrahim S, Kinra S, Bowen L et al (2010) The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Med* 7:e1000268

30. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
31. Prokopenko I, Langenberg C, Florez JC et al (2009) Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 41:77–81
32. Tan JT, Nurbaya S, Gardner D, Ye S, Tai ES, Ng DP (2009) Genetic variation in KCNQ1 associates with fasting glucose and beta-cell function: a study of 3,734 subjects comprising three ethnicities living in Singapore. *Diabetes* 58:1445–1449
33. Stancáková A, Kuulasmaa T, Paananen J et al (2009) Association of 18 confirmed susceptibility loci for type 2 diabetes with indices of insulin release, proinsulin conversion, and insulin sensitivity in 5,327 nondiabetic Finnish men. *Diabetes* 58:2129–2136
34. Simonis-Bik AM, Nijpels G, van Haeften TW et al (2010) Gene variants in the novel type 2 diabetes loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B affect different aspects of pancreatic beta-cell function. *Diabetes* 59:293–301
35. Grarup N, Andersen G, Krarup NT et al (2008) Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* 57:2534–2540
36. Grarup N, Rose CS, Andersson EA et al (2007) Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* 56:3105–3111
37. Moore AF, Jablonski KA, McAteer JB et al (2008) Extension of type 2 diabetes genome-wide association scan results in the diabetes prevention program. *Diabetes* 57:2503–2510
38. Zeggini E, Weedon MN, Lindgren CM et al (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341
39. Gudmundsson J, Sulem P, Steinthorsdottir V et al (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39:977–983
40. Wen J, Ronn T, Olsso A et al (2010) 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
41. Han X, Luo Y, Ren Q et al (2010) Implication of genetic variants near SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, FTO, TCF2, KCNQ1, and WFS1 in type 2 diabetes in a Chinese population. *BMC Med Genet* 11:81
42. Pascoe L, Tura A, Patel SK et al (2007) Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* 56:3101–3104
43. Cauchi S, Meyre D, Durand E et al (2008) Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS One* 3:e2031
44. Staiger H, Machicao F, Kantartzis K et al (2008) Novel meta-analysis-derived type 2 diabetes risk loci do not determine prediabetic phenotypes. *PLoS One* 3:e3019