Polygenic scores of diabetes-related traits in subgroups of type 2 diabetes in India: a cohort study

Chittaranjan S. Yajnik,[a,**] Rucha Wagh,[a,b] Pooja Kunte,[a,c] Olof Asplund,[d] Emma Ahlqvist,[d] Dattatrey Bhat,[a] Sharvari R. Shukla,[a,e] and Rashmi B. Prasad[d,f,*]

[a]Diabetes Unit, Kamalnayan Bajaj Diabetology Research Centre, King Edward Memorial Hospital and Research Centre, Pune, 411011, India
[b]Symbiosis School of Biological Sciences, Symbiosis International (Deemed) University, Pune, 411021, India
[c]Diabetes and Islet Biology Group, School of Medicine, Western Sydney University, Campbelltown Campus, Sydney, 2560, NSW, Australia
[d]Department of Clinical Sciences, Diabetes and Endocrinology, CRC, Lund University, Malmö SE-205 02, Sweden
[e]Symbiosis Statistical Institute, Symbiosis International University, Pune, 411005, India
[f]Institute for Molecular Medicine Finland FIMM, Helsinki University, 00290, Helsinki, Finland

The Lancet Regional Health - Southeast Asia 2023;14: 100182
Published Online 1 May 2023
https://doi.org/10.1016/j.lansea.2023.100182

Summary

Background A machine-learning approach identified five subgroups of diabetes in Europeans which included severe autoimmune diabetes (SAID), severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD) and mild age-related diabetes (MARD) with partially distinct genetic aetiologies. We previously validated four of the non-autoimmune subgroups in people with young-onset type 2 diabetes (T2D) from the Indian WellGen study. Here, we aimed to apply European-derived centroids and genetic risk scores (GRSs) to the unselected (for age) WellGen to test their applicability and investigate the genetic aetiology of the Indian T2D subgroups.

Methods We applied European derived centroids and GRSs to T2D participants of Indian ancestry (WellGen, n = 2217, 821 genotyped) and compared them with normal glucose tolerant controls (Pune Maternal Nutrition Study, n = 461).

Findings SIDD was the predominant subgroup followed by MOD, whereas SIRD and MARD were less frequent. Weighted-GRS for T2D, obesity and lipid-related traits associated with T2D. We replicated some of the previous associations of GRS for T2D, insulin secretion, and BMI with SIDD and MOD. Unique to Indian subgroups was the association of GRS for (a) proinsulin with MOD and MARD, (b) liver-lipids with SIDD, SIRD and MOD, and (c) opposite effect of beta-cell GRS with SIDD and MARD, obesity GRS with MARD compared to Europeans. Genetic variants of fucosyltransferases were associated with T2D and MOD in Indians but not Europeans.

Interpretation The similarities emphasise the applicability of some of the European-derived GRSs to T2D and its subgroups in India while the differences highlight the need for large-scale studies to identify aetiologies in diverse ancestries. The data provide robust evidence for genetically distinct aetiologies for the T2D subgroups and at least partly mirror those seen in Europeans.

Funding Vetenskapsrådet, Diabetes Wellness, and Hjärt-Lungfonden (Sweden), DST (India), Wellcome Trust, Crafoord Foundation and Albert Påhlsson Foundation.

Abbreviations: SIDD, Severe insulin-deficient diabetes; MOD, Mild obesity-related diabetes; SIRD, Severe insulin-resistant diabetes; MARD, Mild age-related diabetes; PMNS, Pune maternal nutrition study; NGT, Normal glucose tolerance; GWAS, Genome-wide association study; GRS, Genetic risk score; BMI, Body mass index; VAT, Visceral adipose tissue; CIR, Corrected insulin response; ISI, Insulin sensitivity index; ISR, Insulin secretion rate; OGTT, Oral glucose tolerance test; IVGTT, Intravenous glucose tolerance test; SNP, Single nucleotide polymorphism; FUT, Fucosyltransferases

*Corresponding author. Department of Clinical Sciences, Diabetes and Endocrinology, CRC, Lund University, S-205 02 Malmö, Sweden.
**Corresponding author.
E-mail addresses: rashmi.prasad@med.lu.se (R.B. Prasad), csyajnik@gmail.com (C.S. Yajnik).

Keywords: Type 2 diabetes; Subgroups; Genetics of subgroups; India; Severe insulin-deficient diabetes; SIDD; Severe insulin-resistant diabetes; SIRD; Mild obesity-related diabetes; MOD; Mild age-related diabetes; MARD

Research in context

Evidence before this study

A recent study based on a European population (Swedish) investigated diabetes heterogeneity and identified five subgroups of diabetes, which included four subgroups of type 2 diabetes (T2D). The T2D subgroups were the severe insulin-deficient diabetes, severe insulin-resistant diabetes, mild-obesity-related diabetes, and mild age-related diabetes. The subgroups were also found to have distinct characteristics and also, partially distinct genetic aetiologies. Our own previous research based on Indians validated these clusters in young-onset T2D and found insulin deficiency to be the major driver of young T2D in India, contrary to earlier purported reports of insulin resistance.

Added value of this study

The four T2D subgroups described in our earlier study involving European population were validated in the unselected (no age exclusion criteria applied) Indian WellGen study. Like our previous study (based on T2D patients diagnosed before 45 years of age), the insulin deficient SIDD group was found to be predominant. Genetic scores derived from European studies for T2D, and related traits associated with T2D in India affirms the applicability of European derived scores in the Indian population. Some of the previously reported genetic associations with subgroups were replicated in the Indian WellGen study. Some associations were unique to the Indian WellGen study, suggesting the involvement of population-specific aetiologies and mechanisms.

Implications of all the available evidence

Our study provides useful clues to the genetic aetiologies and pathophysiological mechanisms in T2D in India. This information is invaluable for preclinical research and could be important clinically by providing biomarkers in the future to predict or diagnose the distinct subgroups and guide treatment.

Introduction

Five subgroups of diabetes were recognised in Swedish people with diabetes by applying a data-driven machine learning approach to commonly measured clinical parameters at diabetes diagnosis [including age, glycated haemoglobin (HbA1c), body mass index, presence of glutamic acid decarboxylase (GAD) autoantibodies and homeostatic model assessment (HOMA2) indices for insulin resistance and secretion].[1] The subgroups differ in relative contribution of beta cell dysfunction and insulin resistance, response to medication, and development of complications. These subgroups were robustly replicated in various populations of European and non-European ancestries.[2–4] The subgroups were partially genetically different with respect to diabetes-related traits, indicating possible etiological differences between the groups[5]

Indian people with type 2 diabetes (T2D) are diagnosed at a younger age and with lower BMI compared to their European counterparts.[6,7] This early onset of diabetes and lower BMI of Indian people with T2D may be partly because of genetic differences; some of which have been previously described.[8–12] However, a major contribution may be due to early life undernutrition in Indians.[13–15] A substantial variation was also reported in the prevalence and clinical characteristics of people with T2D in different states of India in the Indian Council for Medical Research-India Diabetes (ICMR-INDIAB) study.[16] Hence, there is a need to better characterise the genetic and developmental drivers of diabetes in Indian population.

We recently reported subgroups of people with young-onset T2D in India using the European patient derived centroids. Contrary to earlier reports of insulin resistance as the major driver of young-onset T2D, we found that insulin deficiency was the predominant pathophysiology.[2] This finding has implications in understanding the pathogenesis as well as the treatment of diabetes. Genetic studies are likely to validate the potential mechanisms underlying these findings and provide a rational basis for treatment. Therefore, the overarching aim was to investigate associations of genetic risk scores of diabetes-related traits (insulin secretion, insulin resistance, etc.) generated from European studies with T2D and its subgroups in an Indian cohort (WellGen) so as to dissect the underlying aetiology of T2D in India.

Methods

Participants

WellGen cohort (Wellcome Genetics study) and PMNS (Pune Maternal Nutrition Study)

People with T2D (n = 3111) being treated at the Diabetes Unit, KEM Hospital (Pune, India) and associated clinics were included in the WellGen cohort. T2D diagnosis was based on WHO guidelines using the following clinical criteria: age at diagnosis more than 20 years, no history of ketoacidosis, and response to treatment with oral glucose lowering agents.[2,17] People with a clinical diagnosis of type 1 diabetes (T1D) (diagnosis before 20 years of age and on continuous insulin treatment since

then, history of ketoacidosis), fibrocalculous pancreatic diabetes (FCPD) or monogenic diabetes were excluded. In absence of GAD autoantibody measurements, we tested for the presence of genetic T1D by applying a T1D genetic risk score previously validated in Indian population.[18]

Phenotypic measurements were made as described previously.[2] Briefly, clinical information including age, sex, age at T2D diagnosis, family history and socioeconomic status was obtained through a standardised questionnaire. Height and weight were measured using standardised methods.[19,20] Fasting plasma glucose and HbA1c were measured using standard methods.[19,20] Fasting C-peptide was measured by ELISA (Diagnostic Biochem Canade, ON, Canada). Fasting glucose and C-peptide measurements were used to calculate Homeostatic Model Assessment 2 estimates of β-cell function (HOMA2-B) and insulin resistance (HOMA2-IR).[21,22] The characteristics of the participants are shown in Table 1.

For controls, we included normal glucose tolerant (NGT) participants from the Pune Maternal Nutrition Study (PMNS), a birth cohort from the Diabetes Unit, KEM Hospital.[23] Briefly, ~800 families were serially followed-up for anthropometric and biochemical measurements every six years. The participants were classified as having diabetes or NGT based on American Diabetes Association (ADA) 2014 criteria after an oral glucose tolerance test (OGTT, 75 g anhydrous glucose). In this analysis, we have considered the parents who were NGT at 12-year follow-up as controls (Table 1). WellGen and PMNS participants were from the same geographic area and the same background population. Moreover, all the anthropometric and clinical measurements in WellGen and PMNS were performed in the same department using the same instruments and laboratory assays.

The Ethics Committee of the KEM Hospital Research Centre, Pune approved both the studies, WellGen (KEM/HRC/Dir.off/977) and PMNS (KEMHRC/VSP/Dir.Off/EC/065), and all participants signed a written informed consent.

All new diabetes in Scania—Malmö Diet and Cancer (ANDIS-MDC) cohort

The ANDIS-MDC cohort has been described previously.[5] Briefly the ANDIS (All new diabetes in Scania) project (http://andis.ludc.med.lu.se/) is aimed at recruiting all incident cases of diabetes within Scania (Skåne) County in southern Sweden. All healthcare providers in the region were invited. Individuals from ANDIS-MDC (n = 6986) with genome-wide association study (GWAS) data available were included in the current study.[1] For controls, participants from the MDC (Malmö Diet and Cancer) study including individuals from Malmö (the largest city in Scania, Sweden) and born between 1923 and 1950 were selected. Individuals without diabetes (n = 2744) from the MDC cardiovascular arm re-examination cohort (aged 61–85) were used as controls in the genetic analyses.[24]

The ANDIS and MDC study protocols were approved by the regional ethics review committee in Lund (nos. 584/2006, 2011/354, 2011/367, 2012/676, 2014/198, LU 51-90 and 532/2006). All participants gave informed written consent. Study participants received no compensation.

GWAS quality control and imputation

GWAS QC and imputation

Genome wide genotyping data was generated on WellGen and PMNS participants using Affymetrix SNP 6.0 Chips (Affymetrix, CA, USA) as previously described (generated at the Chandak lab, CCMB, Hyderabad, India).[2] Quality control was performed on the samples using the PLINK 1.9 software.[25] All individuals had a genotyping success rate of >95%. Sex checks derived from genetic data were performed to assess concordance with reported sex. Relatedness was assessed by IBS matrix assessment in PLINK and the included individuals were not related to each other by first, second

Characteristic	WellGen study (n = 2217)			NGT (PMNS, n = 461)			
	Male	Female	All	Male	Female	All	p value
No. of participants (%)	1224 (55.2)	993 (44.8)	2217	173 (37.5)	288 (62.5)	461	
Age, years[a]	42.04 (9.58)	41.88 (10.25)	41.97 (9.88)	39.80 (4.21)	33.31 (3.29)	35.56 (4.80)	1.35E-31
BMI, kg/m^2	25.46 (3.68)	27.27 (4.45)	26.27 (4.14)	21.16 (3.39)	19.74 (2.99)	20.25 (3.21)	8.59E-192
Fasting glucose, mg/dL	8.79 (3.15)	9.04 (3.32)	8.91 (3.23)	4.92 (0.45)	4.79 (0.36)	4.84 (0.40)	3.39E-141
HbA1c, mmol/mol	71.88 (22.59)	71 (22.62)	71.48 (22.6)	–	–	–	–
HbA1c, %	8.73 (2.07)	8.65 (2.07)	8.69 (2.07)	–	–	–	–
Fasting C-peptide, nmol/l	0.82 (0.48)	0.83 (0.48)	0.83 (0.48)	–	–	–	–
HOMA2-B	64.81 (44.3)	64.06 (46.17)	64.48 (45.14)	82.24 (36.67)	85.22 (32.94)	84.14 (34.33)	1.69E-45
HOMA2-IR	2.2 (1.38)	2.24 (1.35)	2.22 (1.37)	0.78 (0.51)	0.79 (0.44)	0.78 (0.47)	7.24E-57

Values are mean (SD). p-values are calculated using ANOVA (unequal variances) adjusted for age and sex. NGT: Normal glucose tolerance, PMNS: Pune Maternal Nutrition Study. [a]Age at diabetes diagnosis for WellGen, age at study for PMNS.

Table 1: **Clinical characteristic of the WellGen according to sex.**

or third degrees. Single nucleotide polymorphism (SNP) exclusion criteria included missingness threshold of >5%, minor allele frequency <1%, and Hardy–Weinberg equilibrium with p-value <0.05. Heterozygosity of the samples was checked and none of the samples were outside three SDs from the mean. No duplicate or related samples were found during data analysis, and monomorphic sites were removed. Imputation was performed on the Michigan Imputation server using 1000G Phase 3 v5 (GRCh37/hg1) as a reference panel along with Eagle v2.4 phasing and SAS (South Asian) as population type. SNPs of interest were extracted using the grep command and imputation scores were >0.4.

ANDIS study participants were genotyped with InfiniumCoreExome-24v1-1 BeadChip arrays (Illumina) at Lund University Diabetes Centre, Malmö, Sweden. MDC was genotyped at the Broad genotyping facility using Infinium OmniExpressExome-8 v.1.0 BeadChip arrays (Illumina). For each of the genetic datasets, standard quality control was performed as described previously.[5] Genotype imputation was performed at the Haplotype Reference Consortium v.1.0.3 Michigan Server.

Statistical methods

Subgrouping of people with T2D

K-means clustering was performed as described previously.[1] Briefly, participants with measurements above or below 5 SD from the mean for the clustering parameters were excluded from the analysis. Values outside the limits for HOMA2 (fasting glucose (FG) <3 mmol/L or C-peptide >3.5 nmol/ml) were capped to the proximal upper or lower limits. Participants with FG >25 mmol/L were excluded. To perform supervised clustering in relation to the European derived cluster coordinates, phenotype data (age at diagnosis, HbA1c, HOMA2B, HOMA2IR and BMI) in WellGen were scaled using the same scaling parameters (mean and standard deviation). Due to the unavailability of GADA data, we only included the non-autoimmune T2D subgroups (SIDD, SIRD, MOD and MARD).

Demographic data that were not normally distributed were log-transformed before the analysis. Phenotypic characteristics of the cases and controls as well as the four subgroups were compared using ANOVA (type III sum of squares for unequal variances) adjusted age and sex.

Genetic association and genetic risk scores (GRSs)

GRS traits. A previously described set of 403 SNPs for T2D including only genome-wide significant variants from Mahajan et al.[26] were selected (since this was the latest and largest T2D GWAS reported at this time point) and subsequently restricted to 384 SNPs as assessed in Mansour Aly et al.[5] (Supplementary Table S1, Supplementary Table S2). Of these, 381 SNPs were included whereas 3 SNPs were excluded due to non-availability of data (data not available for rs571342427, rs549498088, rs560716466 and their proxies). These SNPs weighted by their effect on T2D risk in the primary GWAS study were used to construct T2D-GRS (Supplementary Table S2).

GRSs was also constructed in WellGen data for fasting insulin, insulin secretion, insulin action, insulin action secretion (rs7454108 was replaced with proxy SNP rs3957146), adiposity, and impaired lipids weighted by their effect on T2D risk (Supplementary Table S2). These calculations were performed as previously described in the study by Mahajan and colleagues[26] based on clustering of T2D risk variants.

We next constructed weighted GRSs (wGRS) for insulin secretion [corrected insulin response (CIR), insulin secretion rate (ISR), and sensitivity (insulin sensitivity index (ISI))] weighted by their genetic effect on their respective measures in normal glucose tolerant (NGT) individuals. The SNPs for this analysis were extracted from their respective largest genetic studies and assessed in the T2D subgroup GWAS.[5,27–29] The wGRS for BMI, waist-hip ratio (WHR) and estimated visceral adipose tissue (VAT) (rs757318 was replaced with rs9304955) were constructed based on genome-wide association signals from their respective largest genetic studies.[30–32]

Finally, GRSs were also constructed for the five clusters obtained from the only genetic clustering for T2D till date, the T2D "soft" clustering approach, including proinsulin, beta cells, obesity, lipodystrophy, and liver lipids.[33] Proxy SNPs were selected based on linkage disequilibrium $r^2 > 0.8$. The GRS were presented as grouped based on trait characteristics (Supplementary Tables S1 and S2).

GRS assessment. All GRSs were computed using PLINK (version 1.9) weighted by their effect in the primary GWAS (Supplementary Table S2). GRS data were normalised using inverse normal transformation. Separate logistic regression models were generated to assess the association of the GRS with T2D and the subgroups [(1) T2D vs controls (2) SIDD vs controls (3) SIRD vs controls (4) MOD vs controls and (5) MARD vs controls] adjusting for sex. Outcome was presence or absence of T2D (or subgroup) for each individual. Odds ratios were used to interpret the strength of the association. Association of GRS for their corresponding phenotypes (GRS for Insulin secretion 1, Insulin secretion 2, Beta cells, Proinsulin, corrected insulin response (CIR), fasting insulin, Insulin secretion rate with HOMA2B, GRS for insulin sensitivity index (ISI), insulin action, insulin action secretion with HOMA2IR and GRS for BMI, WHR, VAT, lipodystrophy, liver lipids and adiposity with BMI) was performed using linear regression on the inverse normal transformed outcomes (adjusting for age and sex). Beta values were used to interpret the strength of the association. All statistical analyses were performed using R software (v4.0.3). An association was considered

statistically significant if it had a $p < 0.05$, given that this is a replication of previous findings.[5] For the analysis of association of all GRS with T2D, a second model was computed with age, sex, and BMI as covariates.

Sensitivity analysis with LD considerations. SNP list for each trait was assessed for LD within each list using PLINK. Two SNP pairs for T2D, one SNP pair for BMI, one for insulin action secretion GRS and seven for VAT were in LD (Supplementary Table S3). For each SNP pair in LD for a particular trait, one representative SNP was selected for the sensitivity analysis whereas the other was excluded. GRS analysis was performed as described before.

Sensitivity analysis with outliers. We performed a sensitivity analysis by including the participants who had clustering variables values outside mean ± 5 SD and were initially excluded from the primary analysis. We applied the European derived centroids to assign individuals to the T2D subgroups and subsequently performed the GRS association analysis.

Association of SNPs with T2D and subgroups. Association of individual SNPs with T2D risk was performed using logistic regression implemented in PLINK (version 1.9) with sex as a covariate. Power to detect true association for disease prevalence of 11%, risk allele frequency of 20% and relative risk of 1.5 at a modest significance level of 0.05 was 0.9 (http://csg.sph.umich.edu/abecasis/gas_power_calculator/index.html).

Individual SNPs used in the construction of GRSs for adiposity, BMI, obesity, WHR, VAT, liver lipids, impaired lipids and lipodystrophy were tested for association with BMI using linear regression adjusted for age and sex. SNPs used for the construction of GRSs for insulin secretion and resistance related traits were tested for association with the concordant traits (HOMA-B and HOMA-IR respectively) using linear regression in a model adjusted for age and sex. All outcome variables were inverse normal transformed prior to analysis. Association of selected SNPs with subgroups risk was performed using logistic regression implemented in PLINK (version 1.9) with sex as a covariate. BMI and age were not included as covariates since they are clustering parameters.

Type 1 diabetes genetic risk scores. In the absence of GAD autoantibody data, we applied a 'type 1' GRS previously validated in the Indian population[18] to 821 WellGen participants with available genotype data to estimate the proportion of those carrying autoimmune risk alleles as described in our previous study.[2] A positive control group comprised 261 individuals with type 1 diabetes, as described previously.[2] A negative control group comprised 461 participants with NGT (75 g OGTT; WHO 1999 criteria) from the PMNS.

Role of the funding source

The funders had no role in study design, data collection, data analysis, interpretation, or writing of the manuscript.

Results

Complete data on the subgrouping variables (age at diagnosis, BMI, HbA1c, HOMA2B and HOMA2IR) was available on 2217 people with diabetes (Supplementary Figure S1). These people with diabetes were similar to 894 people with diabetes who were not included with respect to age, duration of diabetes, BMI and HbA1c (Supplementary Table S4). The people with diabetes were on average 42 years old (55.2% males, 44.8% females) and duration since diagnosis of diabetes was ~10 years. The average BMI of the people with diabetes was 26.27 kg/m^2 (male 25.4 kg/m^2, female 27.3 kg/m^2), HbA1c was 8.69%, HOMA2B was 64.4 and HOMA2IR was 2.22 (Table 1). Of 2217 people with diabetes, 821 had GWAS data (Affymetrix SNP 6.0 Chips). These people with diabetes were marginally older and had lower fasting C-peptide concentrations compared to the rest (Supplementary Table S5). Characteristics of genotyped people with diabetes compared to NGT controls are presented in Supplementary Table S6. The T1D GRS distribution in people with diabetes confirmed the absence of genetic T1D in our participants (Supplementary Figure S2).

Weighted risk scores for T2D, insulin, glucose and related phenotypes constructed from European studies associate with T2D risk in India

We constructed genetic risk scores (n_{SNPs} = 381, T2D-GRS) based on previously reported genome-wide significant associations with T2D risk from European GWAS[26] in WellGen. T2D-GRS showed a strong association with T2D risk in Indian people with diabetes (OR per 1 SD increment [95% CI] = 1.55 [1.37–1.75]). wGRS for proinsulin levels (n_{SNPs} = 223), corrected insulin response (n_{SNPs} = 220), beta cells (n_{SNPs} = 27), insulin secretion (insulin secretion 1, n_{SNPs} = 8; insulin secretion 2, n_{SNPs} = −21), insulin action/secretion (n_{SNPs} = 37), and insulin action (n_{SNPs} = 16) were all associated with T2D risk in the WellGen study compared to NGT controls (Table 2).

In an additional model, we perform the associations of glucose-insulin trait-related GRS with T2D adjusting for age, sex, and BMI; and the associations persisted for the most cases. However, the associations of GRSs for beta cell, insulin action, and lipodystrophy with T2D were not statistically significant whereas ISI-GRS was significantly associated with T2D risk (Supplementary Table S7).

GRS	T2D (all subgroups N = 821) vs controls (N = 461)		SIDD (N = 369) vs controls (N = 461)		MOD (N = 268) vs controls (N = 461)		SIRD (N = 21) vs controls (N = 461)		MARD (N = 163) vs controls (N = 461)	
	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
T2D	1.55 (1.37, 1.75)	**4.222×10^{-12}**	1.66 (1.42, 1.93)	**9.937×10^{-11}**	1.5 (1.27, 1.76)	**8.014×10^{-07}**	1.15 (0.74, 1.78)	0.543	1.54 (1.27, 1.87)	**1.27×10^{-05}**
Body size and composition										
Obesity	0.87 (0.77, 0.99)	**0.029**	0.9 (0.77, 1.05)	0.17	0.87 (0.74, 1.02)	0.093	1.12 (0.7, 1.79)	0.63	0.79 (0.65, 0.97)	**0.021**
BMI	1.26 (1.12, 1.42)	**1.225×10^{-04}**	1.27 (1.1, 1.47)	**0.001**	1.31 (1.12, 1.53)	**6.180×10^{-04}**	1.37 (0.89, 2.11)	0.158	1.14 (0.95, 1.37)	0.165
VAT	1.09 (0.97, 1.22)	0.165	1.12 (0.97, 1.29)	0.11	1.16 (1, 1.35)	**0.052**	1.23 (0.79, 1.91)	0.353	0.92 (0.76, 1.11)	0.364
WHR	1.17 (1.04, 1.32)	**0.007**	1.22 (1.06, 1.41)	**0.005**	1.14 (0.98, 1.33)	0.086	1.26 (0.81, 1.95)	0.302	1.11 (0.92, 1.33)	0.282
Adiposity	1.14 (1.01, 1.28)	**0.033**	1.23 (1.07, 1.43)	**0.004**	1.09 (0.94, 1.27)	0.265	0.95 (0.61, 1.49)	0.822	1.05 (0.87, 1.27)	0.585
Insulin secretion										
Beta cells	0.89 (0.79, 1)	**0.047**	0.87 (0.75, 1)	**0.045**	0.94 (0.81, 1.09)	0.431	1.27 (0.82, 1.96)	0.284	0.81 (0.67, 0.97)	**0.025**
CIR	0.87 (0.77, 0.98)	**0.018**	0.86 (0.75, 0.99)	**0.036**	0.85 (0.73, 1)	**0.044**	0.98 (0.63, 1.51)	0.913512	0.9 (0.75, 1.08)	0.268
Insulin secretion 1	1.33 (1.18, 1.5)	**2.990×10^{-06}**	1.38 (1.19, 1.59)	**1.980×10^{-05}**	1.25 (1.08, 1.46)	**0.003**	1.48 (0.95, 2.3)	0.08101	1.33 (1.1, 1.61)	**0.003**
Insulin secretion 2	1.22 (1.08, 1.37)	**0.001**	1.15 (1, 1.32)	0.057	1.24 (1.07, 1.45)	**0.004**	0.94 (0.61, 1.43)	0.756627	1.32 (1.1, 1.59)	**0.003**
ISR	0.9 (0.8, 1.01)	0.07	0.86 (0.74, 0.99)	**0.031**	0.95 (0.82, 1.1)	0.498	0.95 (0.61, 1.48)	0.829754	0.89 (0.73, 1.07)	0.215
Insulin action										
Fasting insulin	0.98 (0.87, 1.1)	0.685	1.03 (0.89, 1.18)	0.714	0.96 (0.83, 1.12)	0.614	0.94 (0.61, 1.46)	0.793	0.94 (0.78, 1.13)	0.508
Insulin action secretion	1.17 (1.04, 1.32)	**0.007**	1.17 (1.01, 1.34)	**0.032**	1.09 (0.93, 1.27)	0.302	1.41 (0.89, 2.22)	0.146	1.45 (1.19, 1.78)	**2.590×10^{-04}**
Insulin action	1.16 (1.03, 1.3)	**0.014**	1.1 (0.95, 1.27)	0.192	1.17 (1, 1.36)	**0.049**	1.25 (0.8, 1.93)	0.325	1.26 (1.04, 1.53)	**0.016**
Proinsulin	1.15 (1.03, 1.29)	**0.017**	1.09 (0.94, 1.25)	0.255	1.2 (1.03, 1.4)	**0.017**	0.91 (0.59, 1.42)	0.686	1.23 (1.01, 1.48)	**0.035**
ISI	1.01 (0.9, 1.13)	0.919	1 (0.87, 1.15)	0.976	0.98 (0.84, 1.14)	0.819	0.86 (0.55, 1.33)	0.497	1.09 (0.9, 1.31)	0.388
Lipid metabolism										
Lipodystrophy	1.17 (1.04, 1.31)	**0.009**	1.16 (1, 1.34)	**0.042**	1.18 (1.01, 1.37)	**0.037**	1.12 (0.72, 1.73)	0.621	1.17 (0.97, 1.41)	0.098
Impaired lipids	1.22 (1.07, 1.38)	**0.002**	1.2 (1.03, 1.4)	**0.019**	1.26 (1.07, 1.49)	**0.006**	1.44 (0.88, 2.36)	0.142	1.17 (0.96, 1.44)	0.126
Liver lipids	1.28 (1.13, 1.44)	**6.770×10^{-05}**	1.24 (1.07, 1.43)	**0.004**	1.31 (1.12, 1.53)	**8.502×10^{-04}**	2.04 (1.26, 3.3)	**0.003**	1.16 (0.96, 1.41)	0.121

p-values calculated by logistic regression adjusted for sex for T2D and its subgroups compared to normal glucose tolerance (NGT). Level of significance, $p < 0.05$ (indicated in bold). OR: Odds ratio, T2D: Type 2 diabetes, SIDD: Severe insulin-deficiency diabetes, MOD: Mild obesity-related diabetes, SIRD: Severe insulin-resistant diabetes, MARD: Mild age-related diabetes, WHR: Waist-hip ratio, VAT: Visceral adipose tissue, CIR: Corrected insulin response, ISI: Insulin sensitivity index, ISR: Insulin secretion rate.

Table 2: **Associations of the GRS for diabetes-related traits with the T2D subgroups.**

In a separate analysis of wGRSs of glucose-insulin traits with the corresponding clustering parameters, HOMA2B and HOMA2IR, wGRS for ISI associated with HOMA2IR (Supplementary Table S8).

Weighted risk scores for body composition and lipid related phenotypes constructed from European studies associate with T2D risk in India

We next constructed GRSs based on previously reported genome wide significant associations in European populations for traits related to weight and blood lipids. GRSs for BMI (n_{SNPs} = 122), WHR (n_{SNPs} = 39), adiposity (n_{SNPs} = 6), lipodystrophy (n_{SNPs} = 18), impaired lipids (n_{SNPs} = 3), and liver lipids (n_{SNPs} = 3) were associated with T2D risk in WellGen compared to NGT (Table 2).

In a separate analysis of wGRSs of body composition and lipid traits with the clustering parameter BMI, wGRS for BMI, WHR, VAT, and adiposity associated with BMI (Supplementary Table S8).

Given the well-known substantial difference in the beta cell function and obesity parameters between Indians and Europeans,[6,7,34] we compared distribution of GRSs for related traits between these populations and investigated associations of GRSs with the phenotype. As a rule, there was no significant difference in the distribution of GRSs for fasting insulin, beta cells, BMI, and obesity between the two populations ($p > 0.05$). However, for a given GRS, Indians had lower fasting C-peptide, HOMA2B, BMI and obesity (BMI>30kg/m^2) (Fig. 1).

Subgroups of diabetes in India based on European patient-derived coordinates

To test the applicability of the diabetes subgrouping described in Ahlqvist et al.,[1] we applied European-derived centroids to 2217 Indian people with T2D from the WellGen study (Table 3). In the absence of GAD autoantibody data, we obtained the four expected subgroups. These included the severe insulin-deficient diabetes (SIDD) which was the most common (47.32%), followed by the mild obese diabetes (MOD) (30.94%), whereas the mild age-related diabetes (MARD) (18.58%) and severe insulin-resistant diabetes (SIRD) (3.16%) subgroups were less frequent (Table 3, Fig. 2). In the sex-stratified analysis, in male participants, the SIDD subgroup was most prevalent (53.92%), whereas in female participants, MOD was predominant (43.91%) (Fig. 2). The characteristics of all the subgroups reflected those seen in the European people with diabetes.

Association of weighted T2D genetic risk scores for insulin secretion and sensitivity measures with T2D subgroups

The European study showed that diabetes subgroups described previously have at least partially different genetic backgrounds. The SIRD subgroup seemed

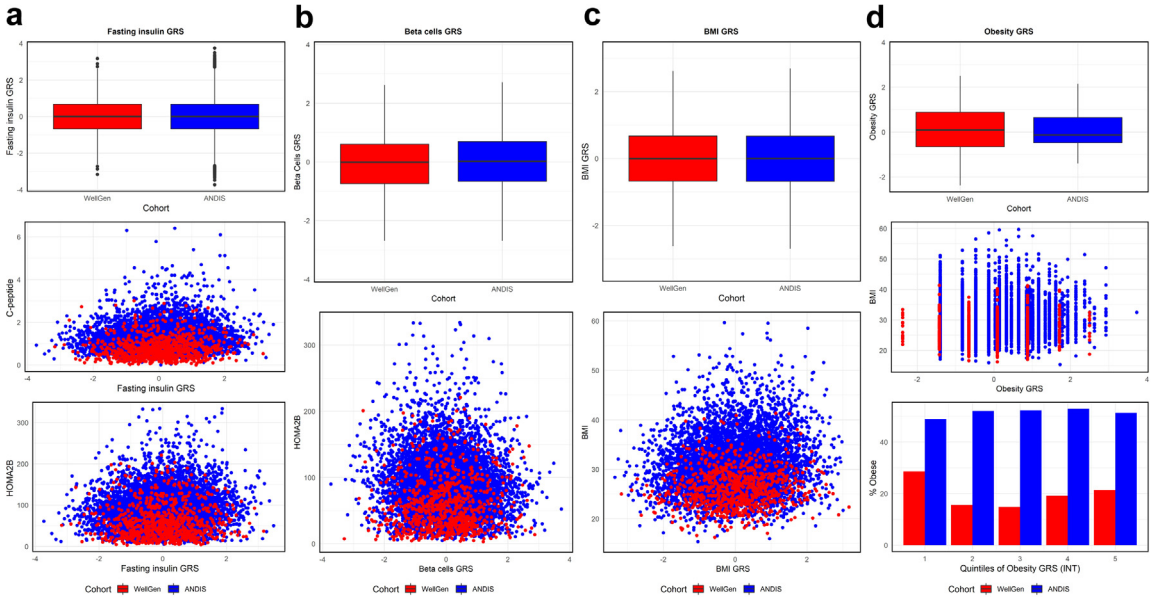

Fig. 1: Box and whisker plots show distributions of genetic risk scores (GRSs) for fasting insulin (a), beta-cells (b), BMI (c) and obesity (d) in WellGen (India) and ANDIS (Sweden) studies; and scatterplots of related phenotypes for the two populations. WellGen is shown in red and ANDIS in blue. For obesity GRS, proportion of patients with diabetes who are obese (BMI > 30 kg/m^2) are shown in quintiles of GRS. For the same value of GRS, WellGen patients have a lower C-peptide, homeostatic model assessment for beta cell function (HOMA2B), BMI and obesity prevalence compared to ANDIS patients.

Characteristic	WellGen (2217)				
	SIDD	MOD	SIRD	MARD	p value
No. of participants (%)	1049 (47.3)	686 (30.9)	70 (3.2)	412 (18.6)	
Age at diabetes diagnosis, years	40.03 (8.53)	38.22 (7.24)	52.6 (10.37)	51.33 (9.76)	1.01E-129
BMI, kg/m^2	25.2 (3.68)	28.77 (4.05)	29.22 (5.17)	24.34 (2.75)	4.96E-89
Fasting glucose, mg/dL	10.73 (3.35)	7.73 (2.27)	6.26 (1.26)	6.67 (1.35)	1.97E-198
HbA1c, mmol/mol	88.28 (19.31)	58.59 (12.71)	57.72 (14.14)	52.52 (10.49)	0.0001
HbA1c, %	10.23 (1.77)	7.51 (1.16)	7.43 (1.29)	6.96 (0.96)	0.0001
Fasting C-peptide, nmol/l	0.72 (0.42)	0.92 (0.48)	1.93 (0.52)	0.74 (0.37)	2.18E-56
HOMA2-B	39.09 (23.34)	80.52 (42.24)	184.29 (42.92)	82.05 (38.55)	3.06E-208
HOMA2-IR	2.17 (1.41)	2.33 (1.27)	4.57 (1.4)	1.78 (0.93)	4.51E-36

Values are mean (SD). p value corresponds to comparison between 4 subgroups and is calculated by ANOVA (type III SS for unequal samples sizes) adjusted for sex. SIDD: Severe insulin-deficient diabetes, SIRD: Severe insulin-resistant diabetes, MOD: Mild obesity-related diabetes, MARD: Mild age-related diabetes, T2D: Type 2 diabetes.

Table 3: **Clinical characteristic of the WellGen according to T2D subgroups.**

genetically different in that it had less involvement in beta-cell related pathways in its pathogenesis.

The GRS of known T2D loci (T2D-GRS) was associated with SIDD (OR per 1 SD increment [95% CI] = 1.66 [1.42–1.93]), MOD (1.5 [1.27–1.76]) and MARD (1.54 [1.27–1.87]), whereas it was not associated with SIRD (1.15 [0.74–1.78]) (Table 2, Fig. 3).

GRS for beta cells developed from Udler et al.[33] and GRS for first phase insulin response, measured as insulin secretion rate (ISR-GRS) based on serum C-peptide during an intravenous glucose tolerance test (IVGTT)[28] were associated with SIDD but not with SIRD, MOD or MARD. On the other hand, GRS for corrected insulin response (CIR-GRS) during an oral glucose tolerance test (OGTT)[27] and insulin secretion GRS from the study by Mahajan and colleagues,[35] were associated with SIDD, MOD and MARD but not with SIRD (Table 2, Fig. 3).

A GRS comprised of insulin action/secretion SNPs[35] was associated with SIDD and MARD, insulin action GRS[35] and proinsulin GRS constructed from the study by Udler and colleagues[33] were associated with MOD and MARD (Table 2, Fig. 3).

Association of wGRS for body composition and lipid related phenotypes with the T2D subgroups

GRS for obesity reported in the study by Udler and colleagues[33] associated with MARD, while a GRS for BMI associated with SIDD and MOD. A GRS for WHR and adiposity[35] associated with SIDD but not the other subgroups. GRS for lipodystrophy from the study by Udler and colleagues[33] was associated with SIDD and MOD, while liver lipid GRS with SIDD, MOD and notably also with SIRD. Impaired lipid GRS reported in the study by Mahajan and colleagues[35] was associated with SIDD and MOD (Table 2, Fig. 3).

Sensitivity analysis

Sensitivity analysis taking LD into account

We replicated the GRSs reported in the study by Mansour Aly and colleagues using the same SNPS to construct the GRSs. Given that LD structures could vary between populations, we performed a sensitivity analysis, by assessing LD between the SNPs within the GRS subsets in WellGen study. Very few SNPs in the BMI, VAT, insulin action secretion and T2D were in LD ($r^2 > 0.8$). We pruned the SNPs, selecting one representative SNP per pair and recalculated the GRS. The associations with T2D and subgroups remained the same as the results before pruning (Supplementary Table S9).

Sensitivity analysis including outliers

Thirty-three individuals had HOMA values outside mean ± 5 SD and were excluded. We performed a sensitivity analysis including the outlier which resulted in an additional 7 SIDD and 26 SIRD cases (Supplementary Figure S3, Supplementary Tables S10–S12). Of these, 18 had GWAS data available (12 were of SIRD subgroup, and 6 were of SIDD subgroup) (Supplementary Table S13). With the inclusion of these 18 individuals, the association of SIRD with liver lipids persisted, new associations with WHR ($p = 0.043$) and insulin secretion 1 ($p = 0.007$) GRSs were observed. However, the association of SIDD with beta cell GRS was not statistically significant ($p = 0.055$) (Supplementary Table S13).

Association of SNPs used to construct GRS with T2D risk

Several SNPs previously associated with T2D, insulin secretion and action associated with T2D risk as well as their respective traits in the Indian WellGen study ($p < 0.05$) most of them with directional consistency (Supplementary Tables S14 and S15).

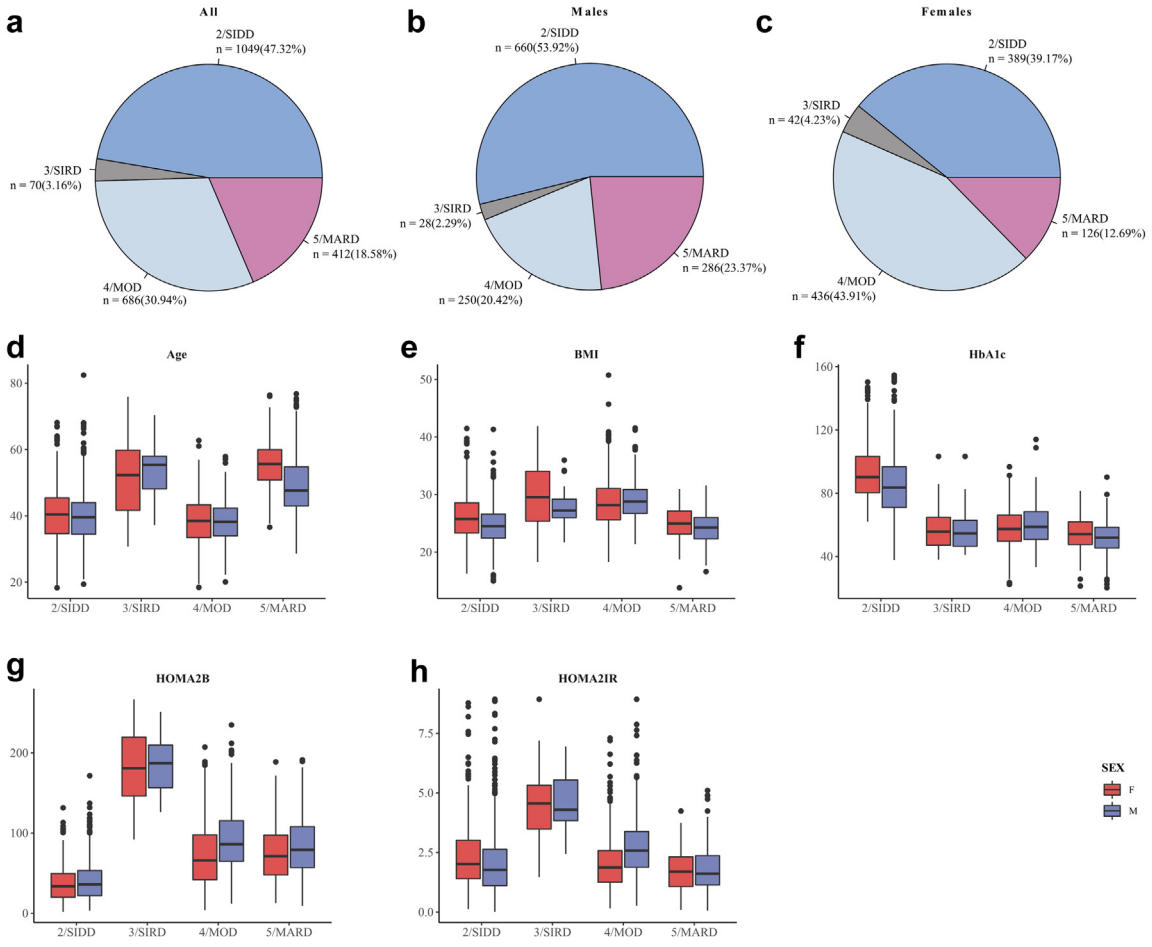

Fig. 2: Distribution of participants from the WellGen study in the predefined subgroups. (a–c) Distribution of WellGen participants, showing all participants (n = 2217) (a), men with diabetes (n = 1224) (b) and women with diabetes (n = 993) (c). Box plot of subgroup characteristics in the WellGen study. Distribution of age at diagnosis (d), BMI (e), HbA_{1c} (f), homeostatic model assessment for beta cell function (HOMA2-B) (g) and homeostatic model assessment for insulin resistance (HOMA2-IR) (h). SIDD–Severe insulin-deficient diabetes, SIRD–Severe insulin-resistant diabetes, MOD–Mild obesity-related diabetes, MARD–Mild age-related diabetes.

Genetic analysis of previously reported loci with subgroups

The TCF7L2 locus does not associate with SIRD

The strongest susceptibility signal for T2D from earlier GWAS was the rs7903146 variant in the *TCF7L2* locus.[36] This SNP associated with the SIDD, MOD and MARD subgroups but not SIRD in the WellGen cohort (Table 4).

LRMDA locus did not replicate the association with MOD

Association of rs10824307 with the MOD subgroup has been reported earlier.[5] This SNP was not associated with MOD or any of the diabetes subgroups in the WellGen study (Table 4).

India-specific T2D loci

Previous studies reported the association of rs9552911 SNP at the SGCG locus and rs998451 at the TMEM163 locus with T2D in India but not in other populations.[9,10] We did not find any association of these variants with any of the subgroups in WellGen (Table 4).

B12 locus FUT2/6 and MOD

Vitamin B12 deficiency has been shown to be endemic to the Indian population and has been linked to foetal growth, insulin resistance, and obesity.[37,38] A previous GWAS based on the Indian population identified variants in fucosyltransferase (FUT) genes (*FUT2/FUT6*) to be associated with vitamin B12 deficiency in Indians.[39] We therefore examined if the same variants are associated with T2D or with specific subgroups. We found that variants in *FUT2* (rs492602, rs601338, rs602662, and rs681343) were associated with T2D but not the variants in *FUT6* (rs3760776 and rs78060698). The same variants in

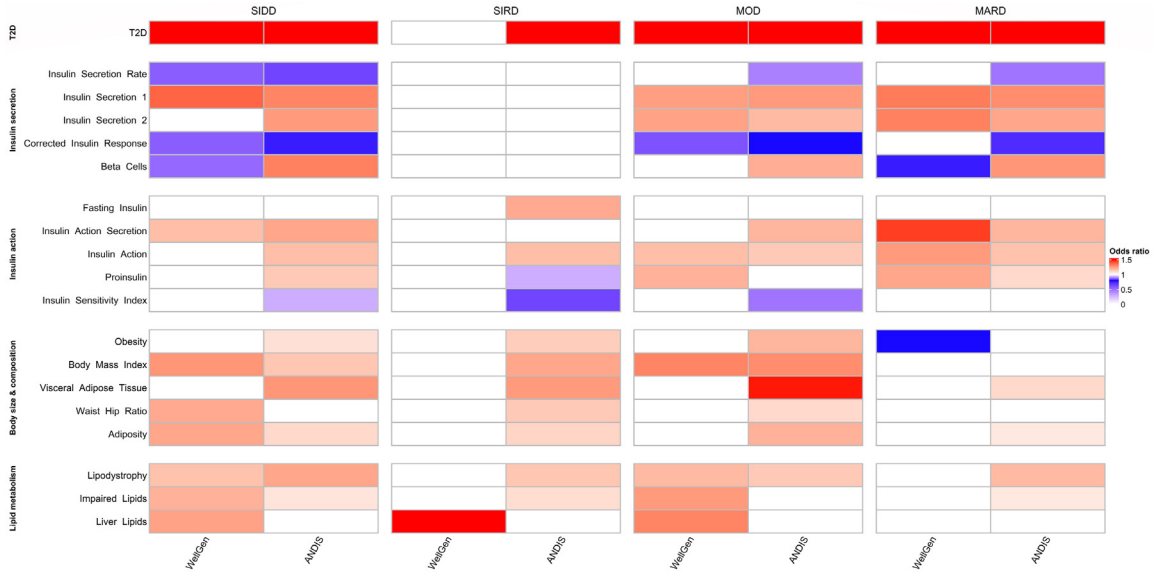

Fig. 3: Heatmap of associations of genetic risk scores (GRSs) for Type 2 diabetes-related traits with subgroups compared to NGT in two populations (WellGen and ANDIS). The odds ratios (OR) for statistically significant associations ($p < 0.05$) are presented. OR>1 represented in shades of red, OR<1 represented in shades of blue. SIDD–Severe insulin-deficient diabetes, SIRD–Severe insulin-resistant diabetes, MOD–Mild obesity-related diabetes, MARD–Mild age-related diabetes, T2D–Type 2 diabetes.

FUT2 and *FUT6* were also associated with the MOD but not with the other subgroups (Table 4). We next assessed for the same associations in the Swedish ANDIS study but did not see any association of FUT2/FUT6 variants with either T2D or subgroups (Supplementary Table S16).

Discussion

To our knowledge, this is the first study showing the association of European-derived genetic risk scores for T2D, insulin and glucose, body composition and lipid related traits with T2D as well as its subgroups in India. Similar to our previous findings,[2] the SIDD subgroup

Individual SNPs	T2D (N = 821) vs controls (N = 461)		SIDD (N = 369) vs controls (N = 461)		MOD (N = 268) vs controls (N = 461)		SIRD (N = 21) vs controls (N = 461)		MARD (N = 163) vs controls (N = 461)	
	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
Diabetes related SNPs										
TCF7L2 rs7903146	1.66 (1.31, 2.1)	**<0.001**	1.69 (1.27, 2.25)	**<0.001**	1.44 (1.06, 1.95)	**0.02**	2.11 (0.87, 5.14)	0.1	1.88 (1.29, 2.74)	**0.001**
ZNF503 rs10824307	1.01 (0.8, 1.28)	0.911	1.09 (0.83, 1.45)	0.529	0.92 (0.68, 1.25)	0.598	0.97 (0.4, 2.34)	0.946	1.05 (0.72, 1.52)	0.799
SGCG–rs9552911	1.2 (0.61, 2.39)	0.596	1.35 (0.58, 3.19)	0.488	1.36 (0.55, 3.35)	0.499	0 (0, Inf)	0.99	1.45 (0.47, 4.48)	0.523
TMEM163–rs998451	1.87 (0.4, 8.68)	0.423	3.19 (0.59, 17.16)	0.177	0.88 (0.08, 9.78)	0.919	0 (0, Inf)	0.991	0 (0, Inf)	0.984
Secretor status related SNPs										
FUT2 rs281377	1.18 (0.85, 1.63)	0.327	1.12 (0.75, 1.67)	0.584	1.28 (0.84, 1.96)	0.248	3.38 (0.85, 13.49)	0.083	0.97 (0.58, 1.63)	0.912
FUT2 rs492602	1.71 (1, 2.91)	**0.05**	1.17 (0.6, 2.28)	0.6504	2.46 (1.32, 4.61)	**0.005**	0.99 (0.12, 8.05)	0.991	1.47 (0.67, 3.22)	0.336
FUT2 rs601338	1.71 (1, 2.91)	**0.05**	1.17 (0.6, 2.28)	0.65	2.46 (1.32, 4.61)	**0.005**	0.99 (0.12, 8.05)	0.991	1.47 (0.67, 3.22)	0.336
FUT2 rs602662	1.73 (1.02, 2.95)	**0.044**	1.22 (0.63, 2.36)	0.557	2.45 (1.31, 4.59)	**0.005**	0.99 (0.12, 8.03)	0.99	1.48 (0.68, 3.24)	0.328
FUT2 rs681343	1.71 (1, 2.91)	**0.05**	1.17 (0.6, 2.28)	0.65	2.46 (1.32, 4.61)	**0.005**	0.99 (0.12, 8.05)	0.991	1.47 (0.67, 3.22)	0.336
FUT2 rs1800027	0.91 (0.21, 3.92)	0.904	0.32 (0.03, 3.14)	0.325	1.17 (0.19, 7.03)	0.868	6.45 (0.61, 68)	0.121	0.86 (0.08, 9.14)	0.904
FUT6 rs3760775	1.05 (0.62, 1.75)	0.8639	0.93 (0.49, 1.76)	0.826	0.8 (0.39, 1.65)	0.551	0 (0, Inf)	0.99	2.04 (1, 4.17)	0.051
FUT6 rs3760776	0.88 (0.51, 1.5)	0.632	0.84 (0.44, 1.64)	0.616	0.35 (0.13, 0.93)	**0.036**	1.76 (0.37, 8.44)	0.48	1.75 (0.84, 3.65)	0.136
FUT6 rs78060698	0.68 (0.37, 1.27)	0.226	0.65 (0.3, 1.42)	0.28	0.17 (0.04, 0.73)	**0.017**	0 (0, Inf)	0.987	1.81 (0.81, 4.01)	0.146
FUT6 rs708686	1.07 (0.74, 1.53)	0.721	0.94 (0.6, 1.47)	0.8	1.09 (0.68, 1.74)	0.732	1.18 (0.35, 4.02)	0.787	1.32 (0.76, 2.31)	0.322

p-values calculated by logistic regression adjusted for sex for age and sex for T2D and its subgroups compared to normal glucose tolerance (NGT). Level of significance, $p < 0.05$ (indicated in bold). OR: Odds ratio, T2D: Type 2 diabetes, SIDD: Severe insulin-deficiency diabetes, MOD: Mild obesity-related diabetes, SIRD: Severe insulin-resistant diabetes, MARD: Mild age-related diabetes.

Table 4: **Associations of individual loci with T2D and its subgroups.**

was predominant in this unselected (for age) cohort and the SIRD subgroup was the least prevalent. We leveraged genome-wide genotyping data and showed that T2D subgroups in India were similarly associated with European generated genetic risk scores for diabetes related traits as shown previously.[5] Our findings support that the above genetic risk scores are indeed associated with T2D risk in India, even in extended analyses allowing for differences in LD structures between populations as well as including outliers. More interestingly, the subgroups of T2D in India have partially distinct genetic pathophysiology reflected in many shared common genetic associations as well as some distinct signatures compared to that seen in Europe.

Genetic risk score for T2D associated with all T2D as well as SIDD, MOD and MARD subgroups, but not with SIRD in India, as seen in the Swedish study (Fig. 3). Unlike in the Europeans, the association of fasting insulin GRS with SIRD was not replicated in our study which may be attributed to the low proportion of SIRD individuals in our data. Interestingly, for each fasting insulin GRS, Indians had lower C-peptide concentrations and HOMA2B compared to Europeans (Fig. 1), highlighting the predominance of insulin deficiency in Indians.

We replicated the strong association of insulin secretion GRSs (insulin secretion 1, CIR, and insulin secretion rate) with the insulin deficient SIDD compared to the other subgroups similar to the Swedish ANDIS study. Interestingly, the direction of effect of the beta-cell GRS was opposite to that found in ANDIS, which could possibly reflect the diverse pathophysiology in the two populations (Fig. 3). While the beta-cell GRS distribution between the two population is comparable, the per GRS insulin secretion as reflected by beta-cell was lower in Indians compared to Europeans (Fig. 1).

The insulin action secretion GRS were found to be associated with SIDD and MARD but not MOD; while insulin action was found to be associated only with MOD and MARD but not SIDD, highlighting the subtle similarities and differences. Proinsulin GRS did not associate with any subgroup from Europe; however, it was associated with MOD and MARD in India, perhaps reflecting population differences in genetic aetiologies and early life experiences. It is interesting to note that animal models of intrauterine undernutrition demonstrated alterations in proinsulin secretion and metabolism[40] and South Asians have increased plasma levels of fasting proinsulin compared to white Caucasians.[41]

We also investigated European generated GRSs for body size and composition associations in T2D and subgroups expecting population-based differences given the distinct 'thin-fat' phenotype in Indians. Obesity GRS was not found to be associated with SIDD, SIRD and MOD but was inversely associated with MARD in Indians, while it had a consistent positive association with all the subgroups in Europeans (Fig. 3). A possible explanation is provided by Fig. 1 which shows that for each BMI and obesity GRS, Indians have a substantially lower BMI or obesity prevalence compared to the Europeans. The GRS for adiposity and WHR were associated with SIDD in India but not in Europe, perhaps reflecting the thin-fat-centrally obese phenotype of Indian participants with T2D.[42] Recent imaging studies of abdominal ectopic adiposity have highlighted excessive liver fat in South-Asians compared to European people with diabetes.[43] The presence of excessive liver fat in South-Asians is perhaps the reason for the exclusive association of liver lipid GRS with SIDD, SIRD and MOD, as well as the association of impaired lipid GRS with MOD in Indians but not in Europeans. The basic adiposity phenotype of Indians seems to extend even to the unique subgroup SIRD (including outliers) wherein the GRSs for VAT and WHR were also found to be associated, unlike in Europe, despite the low number of individuals in this group. We partially replicated the aetiological difference between the SIRD and MOD subgroups, with MOD having the highest GRS effects for BMI, WHR and VAT.

Like ANDIS, most of these insulin secretion GRS, fasting insulin as well as the TCF7L2 variant did not show an association with the insulin resistant SIRD. However, with the inclusion of outliers, the insulin secretion 1 GRS was found to be associated with SIRD, in contrast to ANDIS. This finding needs further confirmation in larger cohorts, and its possible implications for the pathophysiology of this unique subgroup (combining the demographic features of MOD and MARD but with a vigorous beta cell response to prevailing insulin resistance). Both GRSs for liver lipids and impaired lipids were found to be associated with SIRD (in analyses including and excluding outliers) and reflect the potential role of liver in its pathophysiology. Overall, this unique subgroup SIRD poses an intriguing picture of people with diabetes whose beta cells persist to perform despite severe insulin resistance.

T2D is a heterogeneous mixture of people with diabetes who have varied phenotypic combinations arising out of genetic and environmental exposures (early and late life) akin to the Heinz 57 model described by Graham Hitman or the Palette model described by Mark MacCarthy.[44] Our results highlight the genetic and phenotypic heterogeneity in the Indian and European populations. Several SNPs previously associated with T2D in studies of European and multi-ethnic ancestry also associated with T2D in the Indian WellGen study. However, these results should be interpreted cautiously, given the study power and modest effect sizes. We therefore tested the association of the combined effect of the T2D GRS which were replicated for association with T2D in India. However, the known T2D risk loci account for ~20% of T2D heritability, leaving a substantial proportion of unexplained or missing heritability, some of which can be attributed to population specific effects.

Environmental influences across the life course perhaps make a substantial contribution, part of which is explained by epigenetics as a consequence of gene–environment interactions.[45–47] Reputedly, the most important window for this is the periconceptional period which is influenced by maternal nutrition, metabolism, stress, and life course experiences of the mother.[45,48–50] We have reported a substantially lower birthweight in neonates from Indian subcontinent compared to neonates from UK at comparable birthweight GRS.[51] Lower birthweight is a risk factor for future diabetes and other cardiometabolic outcomes. Thus, the population-based differences could extend beyond genetics in that Indians and Europeans have very different life-course experiences.

India is ethnically, linguistically, and genetically diverse with at least four ancestries, European (north India), Dravidian (south India), Tibeto-Burman (north-east India), and Austro-Asiatic (fragmented in east and central India; spoken exclusively by the tribals).[52] Population studies have revealed a historic admixture pattern mirroring a North-South gradient [Ancestral North Indians (ANI) vs. Ancestral South Indians (ASI)].[53–55] A gene-flow from Europe to North India has been indicated in the manifestation of external phenotypes such as skin colour,[56] which could potentially be extended to metabolism. It could be speculated that there would be a higher degree of similarity between findings in Europe and North India whereas this would tend to reduce as one moves further South, given the increasing genetic diversity between the different geographies. Our results are based on a small cohort in western India (Pune, Maharashtra). However, the results which are replicated from European studies could be broadly applicable to the western-Indian population, and potentially extend also to other subsections. However, further studies in other subsections of the Indian population taking geographies and population structures into account will be crucial to further understand the aetiology of T2D in this diverse population.

We further explored the association of variants in TCF7L2 and LRMDA with the subgroups as a replication of our previous study, given that TCF7L2 is the most well-established and well known T2D locus, while LRMDA variant had a robust association with MOD.[5,57] We also included the TMEM163 locus variant, since this was an association specific to T2D in India.[9,10] Our choice of FUT2/6 variants is based on our research in the PMNS wherein we showed an interesting association between maternal vitamin B12 deficiency and insulin resistance in the offspring.[37,38] B12 deficiency is also associated with obesity.[58,59] Obesity and insulin resistance are two major risk factors for T2D. Indians as a group have a substantial prevalence of vitamin B12 deficiency. FUT2 and FUT6 variants showed a strong association with vitamin B12 deficiency in our previous study.[39] We found a significant association between FUT SNPs and T2D and its subgroups driven by MOD in Indian WellGen cohort, but not in the Swedish ANDIS cohort. Furthermore, there are no reports of B12 deficiency related to Swedish population.[60] This is in line with the proposed role for macronutrient excess in the Western populations and a suggested role for micronutrient imbalance in the low-income and middle-income countries related to aetiology of T2D.

Our findings show a resonance with the major Sustainable Development Goals—SDGs described by the UN (3. Good health and wellbeing, 10. Reduced inequalities, 1. No poverty, 2. Zero hunger, 17. Partnership for the goals and others) by highlighting considerable differences in the phenotype despite substantial similarities in the genetic associations of T2D subgroups. It is tempting to ascribe these to historical legacy and currently ongoing socioeconomic and nutritional factors across the two diverse populations.

One of the weaknesses is the low statistical power of the study to find genome-wide associations and the longer duration of diabetes compared to the European study which could influence clinical measurements. Previous studies have shown that substantially greater predictive power can usually be achieved by using GRS rather than a small number of genome-wide significant SNPs.[61–63] Despite low power, most of the GRSs were replicated in the Indian people with diabetes and provided us a genetic tool to support the hitherto described differences in the body composition and beta cell function between Indians and Europeans. We also note that the use of sensitivity analyses including the outlier values with the same statistical methods is not optimal, however, the fact that only one result became non-significant supports the idea that the outliers had no great effect on the outcomes thus increasing the confidence that outlier exclusion was not a strong driver of the results. Future studies may benefit from the use of non-parametrical statistical methods such as resampling to allow for inclusion of outliers in the primary analyses.

In conclusion, our data shows genetic similarities between European and Indian T2D at a broader level, emphasising the applicability of the European derived GRSs to the Indian population. Moreover, we provide robust evidence for genetically distinct aetiologies for the T2D subgroups in India which mirror at least in part those seen in Europeans, and this information is valuable for pre-clinical research as well as clinically important. Large-scale studies of the population-based differences would be interesting to unravel the unique characteristics of T2D in India.

Contributors

RBP and CSY contributed to the conception of the work. CSY, SRS, and DB contributed to data collection in Pune, India, and RBP, EA, and OA in Sweden. RPB, SRS, RW, OA, EA, and PK performed the data analysis. RW, PK, and RBP directly accessed and verified the underlying data. RBP and CSY drafted the article. All authors contributed to the interpretation of data and critical revision of the article. All authors gave final

approval of the version to be published. RBP and CSY are the guarantors of this work and as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Data sharing statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request with necessary permissions from Institutional Ethics Committee and Government of India HMSC permission.

Declaration of interests

The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work. Open access funding provided by Lund University. This study was supported by grants from the Indo-Swedish joint network grant from the Swedish Research Council (Vetenskapsrådet) and the Department of Science and Technology, India (DNR/Reg. No: 2015-06722 to RBP and C/3019/IFD/2018-2019 to CSY), Wellcome Trust (to CSY), Swedish Research Council (Vetenskapsrådet, 2021-02623), Diabetes Wellness Sweden (25-420 PG), The Swedish Heart and Lung Foundation (Hjärt-Lungfonden, 20180522), the Crafoord Foundation and the Albert Påhlsson Research Foundation to RBP. CSY was a visiting Professor to Danish Diabetes Academy (supported by Novo Nordisk Foundation) and Southern University of Denmark during 2016–2019.

Acknowledgments

We thank all people with diabetes for their support and willingness to participate. We thank A. Hattersley (University of Exeter Medical School, UK), G. Chandak and Suraj S. Nongmaithem [Genomic Research on Complex Diseases (GRC), CCMB, India], and S. Kale, S. SirDeshpande, K. Meenakumari, P. Yajnik, R. Ladkat, D. Raut, M. Deshmukh and C. Khole (Diabetes Unit, Kamalnayan Bajaj Diabetology Research Centre, King Edward Memorial Hospital and Research Centre, Pune, India) for their invaluable contribution to the WellGen study. We are grateful to participants and the staff of the Pune Maternal Nutrition Study for their contribution to the control data used in this study. We would like to profusely thanks Leif Groop for his invaluable input and advise. We gratefully acknowledge J. Postma, J. Kravic, and U. Blom-Nilsson (Department of Clinical Sciences, Diabetes and Endocrinology, CRC, Lund University, Malmö, Sweden) for excellent technical and administrative support and thank the ANDIS steering committee for their support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lansea.2023.100182.

References

1 Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6(5):361–369.
2 Prasad RB, Asplund O, Shukla SR, et al. Subgroups of patients with young-onset type 2 diabetes in India reveal insulin deficiency as a major driver. *Diabetologia*. 2022;65(1):65–78.
3 Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol*. 2019;7(1):9–11.
4 Zaharia OP, Strassburger K, Strom A, et al. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol*. 2019;7(9):684–694.
5 Mansour Aly D, Dwivedi OP, Prasad RB, et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat Genet*. 2021;53(11):1534–1542.
6 Narayan KMV, Kondal D, Daya N, et al. Incidence and pathophysiology of diabetes in South Asian adults living in India and Pakistan compared with US blacks and whites. *BMJ Open Diabetes Res Care*. 2021;9(1):e001927.
7 Yajnik CS, Yudkin JS. The Y-Y paradox. *Lancet*. 2004;363(9403):163.
8 Yajnik CS, Janipalli CS, Bhaskar S, et al. FTO gene variants are strongly associated with type 2 diabetes in South Asian Indians. *Diabetologia*. 2009;52(2):247–252.
9 Tabassum R, Chauhan G, Dwivedi OP, et al. Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes*. 2013;62(3):977–986.
10 Saxena R, Saleheen D, Been LF, et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes*. 2013;62(5):1746–1755.
11 Mahajan A, Spracklen CN, Zhang W, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*. 2022;54(5):560–572.
12 Siddiqui MK, Anjana RM, Dawed AY, et al. Young-onset diabetes in Asian Indians is associated with lower measured and genetically determined beta cell function. *Diabetologia*. 2022;65(6):973–983.
13 Bhargava SK, Sachdev HS, Fall CH, et al. Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *N Engl J Med*. 2004;350(9):865–875.
14 Yajnik CS. The lifecycle effects of nutrition and body size on adult adiposity, diabetes and cardiovascular disease. *Obes Rev*. 2002;3(3):217–224.
15 Yajnik CS, Deshpande SS, Jackson AA, et al. Vitamin B12 and folate concentrations during pregnancy and insulin resistance in the offspring: the Pune Maternal Nutrition Study. *Diabetologia*. 2008;51(1):29–38.
16 Anjana RM, Deepa M, Pradeepa R, et al. Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR-INDIAB population-based cross-sectional study. *Lancet Diabetes Endocrinol*. 2017;5(8):585–596.
17 Diabetes mellitus. Report of a WHO Study Group. *World Health Organ Tech Rep Ser*. 1985;727:1–113.
18 Harrison JW, Tallapragada DSP, Baptist A, et al. Type 1 diabetes genetic risk score is discriminative of diabetes in non-Europeans: evidence from a study in India. *Sci Rep*. 2020;10(1):9450.
19 Chandak GR, Janipalli CS, Bhaskar S, et al. Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia*. 2007;50(1):63–67.
20 Yajnik CS, Joglekar CV, Lubree HG, et al. Adiposity, inflammation and hyperglycaemia in rural and urban Indian men: coronary risk of insulin sensitivity in Indian subjects (CRISIS) study. *Diabetologia*. 2008;51(1):39–46.
21 Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28(7):412–419.
22 Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care*. 1998;21(12):2191–2192.
23 Rao S, Yajnik CS, Kanade A, et al. Intake of micronutrient-rich foods in rural Indian mothers is associated with the size of their babies at birth: Pune Maternal Nutrition Study. *J Nutr*. 2001;131(4):1217–1224.
24 Rosvall M, Persson M, Ostling G, et al. Risk factors for the progression of carotid intima-media thickness over a 16-year follow-up period: the Malmo Diet and Cancer Study. *Atherosclerosis*. 2015;239(2):615–621.
25 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.
26 Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505–1513.
27 Prokopenko I, Poon W, Magi R, et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet*. 2014;10(4):e1004235.
28 Wood AR, Jonsson A, Jackson AU, et al. A genome-wide association study of IVGTT-based measures of first-phase insulin secretion refines the underlying physiology of type 2 diabetes variants. *Diabetes*. 2017;66(8):2296–2309.
29 Walford GA, Gustafsson S, Rybin D, et al. Genome-wide association study of the modified stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes*. 2016;65(10):3200–3211.

30 Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197–206.
31 Shungin D, Winkler TW, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187–196.
32 Karlsson T, Rask-Andersen M, Pan G, et al. Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. *Nat Med*. 2019;25(9):1390–1395.
33 Udler MS, Kim J, von Grotthuss M, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med*. 2018;15(9):e1002654.
34 Yajnik CS. Early life origins of insulin resistance and type 2 diabetes in India and other Asian countries. *J Nutr*. 2004;134(1):205–210.
35 Mahajan A, Wessel J, Willems SM, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet*. 2018;50(4):559–571.
36 Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320–323.
37 Yajnik CS, Deshpande SS, Lubree HG, et al. Vitamin B12 deficiency and hyperhomocysteinemia in rural and urban Indians. *J Assoc Physicians India*. 2006;54:775–782.
38 Green R, Allen LH, Bjorke-Monsen AL, et al. Vitamin B12 deficiency. *Nat Rev Dis Primers*. 2017;3:17040.
39 Nongmaithem SS, Joglekar CV, Krishnaveni GV, et al. GWAS identifies population-specific new regulatory variants in FUT6 associated with plasma B12 concentrations in Indians. *Hum Mol Genet*. 2017;26(13):2551–2564.
40 Hales CN, Desai M, Ozanne SE, Crowther NJ. Fishing in the stream of diabetes: from measuring insulin to the control of fetal organogenesis. *Biochem Soc Trans*. 1996;24(2):341–350.
41 Nagi DK, Ali VM, Walji S, Jain SK, Yudkin JS. Hyperinsulinemia in nondiabetic Asian subjects using specific assays for insulin, intact proinsulin, and des-31, 32-proinsulin. *Diabetes Care*. 1996;19(1):39–42.
42 Shelgikar KM, Hockaday TD, Yajnik CS. Central rather than generalized obesity is related to hyperglycaemia in Asian Indian subjects. *Diabet Med*. 1991;8(8):712–717.
43 Iliodromiti S, McLaren J, Ghouri N, et al. Liver, visceral and subcutaneous fat in men and women of South Asian and white European descent: a systematic review and meta-analysis of new and published data. *Diabetologia*. 2022;66:44.
44 McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia*. 2017;60(5):793–799.
45 Fleming TP, Watkins AJ, Velazquez MA, et al. Origins of lifetime health around the time of conception: causes and consequences. *Lancet*. 2018;391(10132):1842–1852.
46 Gluckman PD. Epigenetics and metabolism in 2011: epigenetics, the life-course and metabolic disease. *Nat Rev Endocrinol*. 2011;8(2):74–76.
47 Gluckman PD, Hanson MA. Maternal constraint of fetal growth and its consequences. *Semin Fetal Neonatal Med*. 2004;9(5):419–425.
48 Stephenson J, Heslehurst N, Hall J, et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet*. 2018;391(10132):1830–1841.
49 Barker M, Dombrowski SU, Colbourn T, et al. Intervention strategies to improve nutrition and health behaviours before conception. *Lancet*. 2018;391(10132):1853–1864.
50 Hatem G, Hjort L, Asplund O, et al. Mapping the cord blood transcriptome of pregnancies affected by early maternal anemia to identify signatures of fetal programming. *J Clin Endocrinol Metab*. 2022;107(5):1303–1316.
51 Nongmaithem SS, Beaumont RN, Dedaniya A, et al. Babies of South Asian and European ancestry show similar associations with genetic risk score for birth weight despite the smaller size of South Asian newborns. *Diabetes*. 2022;71(4):821–836.
52 Basu A, Sarkar-Roy N, Majumder PP. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc Natl Acad Sci U S A*. 2016;113(6):1594–1599.
53 Venkatesan V, Lopez-Alvarenga JC, Arya R, et al. Burden of type 2 diabetes and associated cardiometabolic traits and their heritability estimates in endogamous ethnic groups of India: findings from the INDIGENIUS Consortium. *Front Endocrinol*. 2022;13:847692.
54 Indian Genome Variation C. Genetic landscape of the people of India: a canvas for disease gene exploration. *J Genet*. 2008;87(1):3–20.
55 Mittal B, Mittal RD, Kumar S. Designing genetic association studies for complex traits in India. *Indian J Med Res*. 2017;145(6):715–717.
56 Ali M, Liu X, Pillai EN, et al. Characterizing the genetic differences between two distinct migrant groups from Indo-European and Dravidian speaking populations in India. *BMC Genet*. 2014;15:86.
57 Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes*. 2015;6(1):87–123.
58 Allin KH, Friedrich N, Pietzner M, et al. Genetic determinants of serum vitamin B12 and their relation to body mass index. *Eur J Epidemiol*. 2017;32(2):125–134.
59 Razbekova M, Issanov A, Chan MY, et al. Genetic factors associated with obesity risks in a Kazakhstani population. *BMJ Nutr Prev Health*. 2021;4(1):90–101.
60 Lokk J, Nilsson M, Norberg B, Hultdin J, Sandstrom H, Westman G. Shifts in B12 opinions in primary health care of Sweden. *Scand J Public Health*. 2001;29(2):122–128.
61 International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748–752.
62 Agerbo E, Sullivan PF, Vilhjalmsson BJ, et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. *JAMA Psychiatry*. 2015;72(7):635–641.
63 Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet*. 2019;104(1):21–34.