Common and distinct genetic architecture of age at diagnosis of diabetes in South Indian

Srinivasan, Sundar; Liju, Samuel; Sathish, Natarajan; Siddiqui, Moneeza K.; Anjana, Ranjit Mohan; Pearson, Ewan R.

Published in:
Diabetes Care

DOI:
10.2337/dc23-0243

Publication date:
2023

Common and distinct genetic architecture of age at diagnosis of diabetes in South Indian and European populations

Sundararajan Srinivasan[1], Samuel Liju[2], Natarajan Sathish[2], Moneeza K Siddiqui[1], Ranjit Mohan Anjana[2], Ewan R. Pearson[1], Alexander S.F. Doney[1], Viswanathan Mohan MD[2], Venkatesan Radha [2], Colin N.A Palmer[1*]

[1] Division of Population Health and Genomics, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, DD1 9SY, United Kingdom.

[2]Madras Diabetes Research Foundation and Dr Mohan's Diabetes Specialities Centre, Chennai, India

Short title: GWAS of Age at diagnosis of Type 2 diabetes

ADDRESS FOR CORRESPONDENCE:

Professor Colin NA. Palmer
Email: c.n.a.palmer@dundee.ac.uk. Phone: +441382383155
Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, United Kingdom.

Manuscript word count (not including tables, Figs, and references): 3772

Number of tables and figures: 4

This research was funded in whole, or in part, by the Wellcome Trust [Grant number 072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence.

Twitter Summary

Our meta-analysis identified two genetic loci associated with early onset of Type 2 Diabetes. Genome wide polygenic scores performed better in South Indians than Europeans.

OBJECTIVE

South Asians are diagnosed with type 2 diabetes (T2D) more than a decade earlier in life than seen in European populations. We hypothesised that studying the genomics of age of diagnosis in these populations may give insight into earlier age diagnosis of T2D among individuals of South Asian descent.

RESEARCH DESIGN AND METHODS

We conducted a meta-analysis of GWAS of age at diagnosis of T2D in 34,001 individuals from four independent cohorts of European and South Asian Indians.

RESULTS

We identified two signals near the *TCF7L2* and *CDKAL1* associated with age at the onset of T2D. The strongest genome-wide significant variants at chromosome 10q25·3 in *TCF7L2* (rs7903146; $p = 2{\cdot}4 \times 10^{-12}$, Beta = -0·436; SE = 0·02) and chromosome 6 p22·3 in *CDKAL1* (rs9368219; $p = 2{\cdot}29 \times 10^{-8}$; Beta = -0·053; SE=0·01) were directionally consistent across ethnic groups and present at similar frequencies, however both loci harboured additional independent signals that were only present in the South Indian Cohorts. A genome wide signal was also obtained at chromosome 10q26·12 in *WDR11* (rs3011366; $p = 3.255 \times 10^{-8}$; Beta = 1·44; SE=0·25) specifically in the South Indian cohorts. Heritability estimates for the age diagnosis were much stronger in South Indian compared to Europeans, and a polygenic risk score was constructed using a South Indians, which explained about 2% trait variance.

CONCLUSIONS

Our findings provide a better understanding of ethnic differences in the age at diagnosis and indicate the potential importance of ethnic differences in the genetic architecture underpinning T2D.

Article Highlights

- Transethnic meta-analysis revealed an association of *TCF7L2* and *CDKAL1* with early age at diagnosis of T2D in South Indians
- Our findings highlight the disparity in heritability estimates of type 2 diabetes age diagnosis in South Indians and Europeans and explain polygenic variation in South Indians.
- Genome wide polygenic scores based on South Indians GWAS performed better in South Indians (R^2= ~2 %) than Europeans (R^2= ~ 0.1%).

Type 2 diabetes (T2D) is a multifactorial disease characterised by impaired insulin action and pancreatic islet dysfunction. The global prevalence of T2D is a pivotal driver of cardiovascular and renal disease(1–3) affecting hundreds of millions of people globally and is responsible for long-term complications, decreased quality of life, and increased mortality(4–7). Improved understanding the intrinsic genomic and phenotypic heterogeneity driving T2D has major potential for improvement of its clinical management and reducing morbidity and mortality. South Asian Indians have an earlier age of onset of diabetes compared to Europeans and mounting evidence suggests this is associated with earlier mortality emphasising the need to delay or prevent the onset of T2D in this ethnic group(8,9). South Asians with newly diagnosed diabetes may have a higher risk for microvascular complications than Europeans(10). Recent studies highlight that higher cardiovascular mortality and disease risk is associated with early onset of T2D diagnosis compared to individuals with delayed onset of the disease(11). South Asians (individuals originating from India, Pakistan, and Bangladesh) are genetically more diverse than the white Europeans, and the prevalence of T2D is much higher in this ethnic group than other ethnic backgrounds(12–14).

Currently nearly 250 genetic loci (more than 400 unique genetic variants) have been identified that influence T2D(2,15,16). Several of these genetic loci have only been identified in European study populations. A trans-ethnic meta-analysis of European and East Asian populations reported several T2D risk variants with significant allelic frequency heterogeneity(12,13). Such frequency differences between ethnic populations affects the power to detect genomic signals within a specific ethnic subgroup. A recent study reported that migrant South Asians are more insulin resistant and have poorer β-cell function at a younger age than White Europeans. Previously identified genetic variants explained about (~10%) of the heritability of T2D(14) .

Despite advancement in genetic research tools, South Asian Indian specific studies are minimal compared to European ancestry. To our knowledge, no GWAS has been published that addresses the age at diagnosis of T2D in people of South Asian Indian ethnicity and compare this with European populations. We aimed to identify novel genetic determinants that influence the risk of younger age of diagnosis in two distinct ethnic backgrounds more specifically in South Asian Indians and Europeans. We aimed to develop, evaluate, and understand a T2D age at diagnosis PRS in in cohorts from South India (DMDSC) and Europe (GoSHARE). In this multicentre study, we focused on inter-ancestry differences in the genetics of age diagnosis of T2D that might influence ethnic-ancestry differences in health outcomes in general and T2D.

Research Design and Methods

Study participants

We included participants from four independent cohorts: Dr Mohan's Diabetes Specialties Centre (DMDSC)(17), Chennai, India, Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS)(18), Genetics of Scottish Health Research Register (GoSHARE)(19) and the United Kingdom Biobank (UKBB)(20). DMDSC is a chain of diabetes hospitals and clinics established in 1991 in Chennai, Southern India, which currently includes 50 clinics in various locations across 8 states in India(17). A total of 560,000 patients with type 2 diabetes to date, are provided with a unique identification number at their first visit, and clinical, anthropometric, and biochemical data are updated at each subsequent visit. Each patient underwent a comprehensive evaluation for screening and assessment of diabetes and presence of chronic complications at the time of their registered first visit, and these tests were repeated subsequently. All these data are collected and stored in the common diabetes electronic medical records (DEMR) system. The Research unit of DMDSC is named as the Madras Diabetes Research Foundation (MDRF) and is accredited by College of American Pathologists (CAP) as well as by the National Accreditation Board for Testing and Calibration Laboratories (NABL) for various biochemical tests. GoDARTS consists of 18,306 participants from the Tayside region of Scotland, of which 10,149 participants were recruited based on their diagnosis of type 2 diabetes(18). GoSHARE currently comprised of a biobank of around 74,000 individuals across NHS Fife, and NHS Tayside(19). Both cohort participants' have provided a sample of blood for genetic analysis and informed consent to link their genetic information to the anonymized electronic health records. UKBB is a large prospective general population cohort. A total of 502,628 individuals who were recruited between 2006-2010 at ages between 40 and 69 years from across the UK and provided electronically signed consent to use

their self-reported answers on socio-demographic, lifestyle, ethnicity, a range of physical measures and blood or urine or saliva samples(20).

All research has been conducted under the principles of the Declaration of Helsinki and approved by corresponding institutional review boards. All study participants provided written informed consent, and Institutional Ethics Committees approved the study. This study is reported according to the Strengthening the Reporting of Genetic Association Studies (STREGA) reporting guideline(21) (Supplementary material).

Phenotyping – Age at Diagnosis of Type 2 Diabetes

We included 8,295 type 2 diabetes patients from the DMDSC cohort of South Indians, whose first clinical visit was within one year of diagnosis. Age at diagnosis was recorded at first visit to the DMDSC. For individuals who are diagnosed during the visit at DMDSC, the age of diagnosis is recorded at that time. Diabetes is diagnosed by general practitioners using WHO criteria(22) for diagnosis and by the oral glucose tolerance test or fasting and/or random glucose test of HbA1c test. All study participants underwent a structured assessment, including detailed family history at the DMDSC. We excluded patients with type 1 diabetes or if positive for GAD65 and ZnT8 antibodies during treatment and follow-up. if positive for GAD65 and ZnT8 antibodies during diagnosis or follow-up. Although this study is cross-sectional, diabetes classifications were applied retrospectively to ensure that the population under study only included individuals with type 2 diabetes. We selected the study participants in the GoDARTS, and GoSHARE based on the following inclusion criteria, aged between 20 and 80 years, and the type 2 diabetes status is monitored continuously and can be updated by primary or secondary physician or community nurses. Algorithms are in place that track use of insulin and other oral hypoglycemic agents. Individuals who are originally classed as having T2D can be re-classified as T1D if they were

diagnosed under the age of 30 years while also having routine insulin use as per WHO guidelines and recorded in the Scottish Care Information-Diabetes system(23). The age at diagnosis is available as part of the Scottish Care Information (SCI) Diabetes Collaboration data recording system which is centrally updated. This is the most precise estimate of when the disease was diagnosed(24). We identified 14,552 T2D participants within the UKBB cohort using diabetes diagnosis by the doctor (data-field code 2443), started insulin after a year diagnosis (2986) and the self-reported ethnicity (21000) and excluded participants with outlying principal components.

Genotyping, Quality Control, and Imputation

Blood samples were collected from the DMDSC participants. A total of 5,801 patients with T2D were genotyped by the Illumina using the global screening arrays version 1.0 (GSA v1·0) (5,801 T2D patients) and 2,494 patients with T2D were genotyped by the Illumina using the GSA v2·0. All genotyped samples were converted to PLINK format files using Illumina Genome Studio v2·04. We excluded samples if their call rate was less than 95% and genetically inferred sex discordance with phenotype data, batch effects, heterozygosity > 3 standard deviation, sample duplicates (IBD score > 0.8). We excluded the SNPs with less than 97 % call rate and Hardy Weinberg Equilibrium (HWE) p-value less than 1×10^{-6} (autosomal variants only). QC assessment was performed independently for DMDSC cohorts before and after phasing and imputation against the haplotype reference consortium (HRC r1·1) panel.

Genotyping of GoDARTS and SHARE cohorts were derived from various platforms: Affymetrix 6·0 (Affymetrix, Santa Clara), Illumina Omni Express-12VI platform and GSA v·2·0, respectively. A total of 11,154 (6,999 GoDARTS and 4,155 GoSHARE) participants were considered after excluding those individuals failing to meet QC criteria. Individual genotype call rate (< 95%), heterozygosity>3 SD from the mean and the highly related sample's identity by descent. We then

carried out SNP-level QC by excluding markers < 97 % call rate, Hardy-Weinberg $p < 1 \times 10^{-6}$. PLINK version 1.7 and 1.9 were used for quality control assessment and data pre-processing for imputation. Ancestry outliers were identified by principal component analysis in each cohort. The genotype data from all three cohorts were imputed against the haplotype reference consortium (HRC r1.1) reference panel. Monomorphic markers or imputation quality score < 0.4 markers were excluded in the post-imputation data.

Ethnic-Specific Meta-analysis of GWAS

Genome-wide association analyses were performed independently for each cohort using an additive model while adjusting for sex (Supplementary Fig 5 -8). It is worth to highlight those previous studies highlighted South Asian in general have weakest age-adjusted association between BMI in those with T2D and those without diabetes(25). In our own datasets, we also find that BMI is the weakest predictor of age of diagnosis of T2D in South Indian cohort but is so in white Europeans (Supplementary Fig 10 & table S4). We estimated allelic effects using the using a linear mixed model as implemented in BOLT-LMM version 2·3·2 which accounts for relatedness and any population stratification and SNPTESTv2·5 in each cohort accordingly. We performed a meta-analysis based on ancestry: South Asian Indians specific analyses include the DMDSC cohort, a unique South Indian population, and migrated South Asians in the UKBB and the European specific analyses include the GoDARTS, GoSHARE, and White Europeans in the UKBB. We performed the meta-analyses using a fixed-effect method in METAL software(26), which assumes the effect allele is the same for each study within an ancestry. We then conducted trans-ancestry meta-analyses using the HRC imputed data up to 26·2 million SNPs directly genotyped or successfully imputed at high quality across all the study cohorts. Heterogeneity across these studies was assessed by the I^2 (low to high) and Cochran's Q statistics as reported by METAL. The Forest

plot was generated using the metafor package. We annotated the genetic variants using the University of California Santa Cruz (UCSC) Genome resource based on the Genome Reference Consortium Human genome build 37.

Conditional analysis

We performed conditional analyses to identify additional secondary signals across the lead SNPs within the South Indian population.

SNP-based heritability

We used the summary statistics data from the South Indians and Europeans specific meta-analyses to estimate the SNP-based heritability in a liability scale using Linkage Disequilibrium Score Regression (LDSC) software(27).

Genome-wide Polygenic risk scores for age at diagnosis of type 2 diabetes

For the Polygenic risk scores (PRS), we considered summary statistics of DMDSC samples. The PRSice tool generates the scores by the weighted sum of the risk allele carried by individuals based on effect estimate. We removed DNA polymorphisms with ambiguous strands (A/T or C/G) from the score derivation. SNPs were clumped to a more significant SNP in LD ($r2 \geq 0.10$) within a 500 kb window. The PRS calculation considered several p-value thresholds (0·001, 0·05, and 0·1).

Role of the funding source

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

Results

A total of 34,001 participants with T2D were included for this study after-quality control filtering: 8,295 T2D patients of South Indian ancestry from Dr Mohan's Diabetes Specialties Centre (DMDSC) a unique and homogenous population as shown in Supplementary Fig 1 , 6,999 patients of European ancestry from Genetics of Diabetes Audit and Research Tayside and Scotland (GoDARTS)(18), 4,155 patients of European ancestry from Scottish Health Research Register (GoSHARE)(19) and 14,552 T2D participants from United Kingdom Biobank (UKBB)(20). A detailed illustration of the study design is presented in Supplementary Fig S1. We identified participants of European (N=13,744) and South Asian Indians (808) in the UKBB using principal components (PCs) analysis of genome-wide data and found that this was consistent with self-reported ancestry information (Supplementary Fig S2). The population characteristics of the cohorts are described in Supplementary Table 1. Notably, we observed the average age of diagnosis of T2D in the South Indians was 40 years, whereas in the white Europeans, the mean age of diagnosis was 59 & 58.2 years and 54.6 years in the UKBB.

SNP-based heritability

Using the LDSC tools, we estimated the SNP-based heritability for age of diagnosis of T2D in South Indians was 17% (SE 6%) but was only 5% (SE 2%) for Europeans.

Trans-ethnic Meta-analysis of GWAS for age of T2D diagnosis

Our meta-analysis revealed two previously known T2D loci at chromosome 10 q25·2 near Transcription factor 7-like 2 (rs79603146, *TCF7L2*, $P < 2{\cdot}4 \times 10^{-12}$, Beta -0·436; SE = 0·02; P-het =0.01) and at chromosome 6p22·3 cyclin-dependent kinase 5 (*CDK5*) regulatory subunit-associated protein 1-like 1 (*CDKAL1)* (rs9368219, $P < 2{\cdot}29 \times 10^{-8}$, Beta -0·053; SE 0·01; P-het =0·007) associated with T2D age-diagnosis (Fig1 & Supplementary Fig S9). The allelic

frequency of *CDKAL1* is more common in South Indian cohort (MAF = 0·26) compared to Caucasians (MAF =0·18) The lead SNPs at *TCF7L2* and *CDKAL1* (Fig 1 and & Supplementary Fig S9) locus demonstrated consistent allelic direction across all cohorts, with the risk alleles associated with lower age of diagnosis, however, a large difference was observed in the size of the estimate of the effects between the South Indian and European cohorts explaining that variation in allelic effect estimates is presumably due to their genetic ancestry. Interestingly the effect size of the variants was much lower in the cohorts of European descent. Ethnic-specific meta-analysis results are presented in supplementary tables (excel sheet 1). There was no evidence of population stratification in the meta-analysis (genomic inflation factor, $\lambda = 1{\cdot}007$).

Stratification of Type 2 Diabetes by Age of diagnosis

As the two ethnic groups were very different in the mean age of diagnosis, we explored the extent to which the observed differences in allelic effect size may be determined by the heterogeneity in age of onset between the populations, therefore stratified by of age of onset. DMDSC T2D individuals under different age groups were provided in the supplementary table S5. Only small proportion of individuals were diagnosed over 55 years.

Based on the European mean age of diagnosis, the study participants in both ethnicities were stratified into the early-diagnosis T2D group (20-55 years) (Fig 1) and late diagnosis T2D (diagnosed after 55 years) (Fig 1). We found that the effect size of both the *TCF7L2* and *CDKAL1* variants was more pronounced in the early onset group regardless of ethnicity (Fig 1 & Supplementary Fig S3). These variants have very little effect on age of diagnosis (Table 2) in those with diabetes diagnosed after 55 years of age in either ethnicity (Fig 1). Although this variant is also nominally associated with age of diagnosis in the later onset group in Europeans, we observed high heterogeneity (I^2=79%) between young and older onset (Supplementary Fig S11).

The role of other T2D variants in Age of T2D diagnosis

We identified several previously reported T2D variants as suggestive signals ($P < 1 \times 10^{-5}$) in this age of diagnosis of T2D trans-ethnic meta-analyses (Supplementary Table excel sheet 1). In particular, the risk variant nearby *SEC24B* at chromosome location 4q25 (rs76170449, $P< 1{\cdot}79\times10^{-7}$) is also associated with cardiovascular traits, and 3p24·3 *ZNF385D* (rs17011243, $P< 1{\cdot}13\times10^{-5}$) associated with T2D in prior GWAS studies. In addition to the other suggestive signals, we detected potential common variants at chromosome location 16p13.3 (rs1977100, *TPSD1*, $P < 3{\cdot}40 \times 10^{-6}$) and 17q21·2 (rs684214, *MLX*, $P< 2.40 \times10^{-6}$) with no difference in their effect estimates between two distinct ethnic groups (Supplementary Table). We replicated previously reported South Asian T2D genome-wide signals(15,28–31) with suggestive evidence or a nominal association for age of diagnosis of T2D in the trans ethnic meta-analyses and ancestry-specific groups. Most of the formerly associated T2D loci from earlier GWAS showed consistent effect estimates in South Indian and European subjects. These include *LPL, SLC30A8, GCKR, THADA, HNF1A, TPCN2, GRB14, SIX3, WDR11, SPC25, CENTD2, MLX, APS32, WFS1, ST6GAL1, KNCQ1 and IGF2BP2*.

Meta-analysis of South Indian cohorts

In the meta-analysis of only the South Indian cohorts, we also found an additional novel genome-wide signal at chromosome 10 q26·12 near the *WDR11* region (rs3011366, $P < 3.255\times10^{-8}$, Beta 1·44, SE 0·26). However, this variant was not associated the age-of-diagnosis in the European cohorts (Table 1). WDR11 encodes the WD repeat domain family, which involves signal transduction and cell cycle progression. Previous GWAS studies in the European populations and UK Biobank T2D participants have reported that *WDR11* (rs3011366) was associated primarily with fasting glucose(32). In-silico lookups in the Common Metabolic Diseases Knowledge Portal indicate that this SNP near *WDR11* is also associated with youth-onset T2D with a nominal

significance level in trans-ancestry cohorts(33). A recent case-control meta-analysis highlighted the association of *WDR11* gene with T2D in East Asians(13) but the association was shown for a different allele in Europeans, East and South Asians. The Conditional analyses conducted in South Indian ancestry (Supplementary Table S3), indicated two independent secondary signals at *TCF7L2* (rs570193324, q25·2, P < 3·2E-05, Beta 9·8, MAF 0·002, R2 0·0006) and *CDKAL1* (rs143316471, P < 0·0054, Beta -5·3, MAF 0·003, R2 0·005). Allelic frequency for both independent signals were rare in European cohorts compared to South Indian Cohorts. The regional plot for independent association of *TCF7L2* and *CDKAL1* is shown in Supplementary Fig S4

Meta-analysis of European cohorts

In the analyses unique to White Europeans, we did not observe any genome-wide signal in the European specific meta-analyses. However, we observed suggestive association of a missense variant rs2232328 near *SPC25*, an established variant for fasting blood glucose and type 2 diabetes(34). Several other SNPs reached suggestive significance for age of diagnosis of T2D, and the direction of effect was consistent across all cohorts of European descent (Supplementary Table excel sheet 3). Notably, rs17843614 near *HLA-DQB1* reached suggestive significance. While *HLA-DQB1* is a well-established T1D locus, there has been strong and consistent evidence about the association of rs17843614 with Type 2 Diabetes(35).

PRS analysis reveals polygenic effects for Age at the onset of Type 2 Diabetes

PRS are emerging as more informative clinical screening and prediction tool with an increasing number of robust genomic variants identified through more extensive genetic association studies(36). To investigate whether different genetic variants shared between ethnicities were conferring the risk of the onset of T2D: The PRS was derived as the weighted sum of risk alleles based on beta values from the DMDSC 1 GWAS summary statistics. Results were presented based

on 8,232,187 SNPs after the quality control and clumped based on LD ($r^2 \geq 0.1$) within a 500 kb window. Then we validated this using the DMDSC 2 of South Indian data which contained no overlapping participants; next we assessed the performance of this South Indian GRS in the European cohort (GoSHARE) (Fig 2). The PRS replicated strongly between the South Indian cohorts. On the other hand, the South Indian derived PRS explained less than 0·1% of the variance in age of diagnosis of T2D in the GoSHARE cohort (European ancestry).

Discussion

In this study, we undertook a trans ethnic meta-analysis of age of diagnosis of T2D in 34,001 T2D individuals from two diverse ancestral backgrounds, European and South Asian Indian, revealing a differential role for established T2D susceptibility loci in determining age of onset of diabetes. Interestingly the well-established T2D signal at *TCF7L2* was much more strongly associated with age of onset in the South Indian population compared to the European cohort, despite the allele frequency not differing between these ancestral groups. We showed that this difference was due to the distribution of age of onset of diabetes within the two ancestral groups, with the *TCF7L2* effect being largely observed in those diagnosed before the age of 50 in both ancestral groups (Table 2). This is consistent with the concept that early onset disease would have a stronger genetic component; indeed, when we looked at the overall heritability estimates for age of diagnosis of T2D, the heritability was much stronger in the younger South Indian population with diabetes when compared to the more elderly European population with diabetes. We also found evidence for ethnic-specific signals that were associated with an early age at diagnosis of T2D in South Asian Indians that were very rare in the European cohort. Our findings emphasize and support our recently reported finding that South Indians have greater genetic beta-cell dysfunction compared to Europeans(37).

The role of beta cell function as a driver for the early age onset of T2D in South Indians is well supported from our ethnic specific *TCF7L2* and *CDKAL1* signals. The *TCF7L2* variation has led to an upsurge risk of early onset of T2D among African Americans(38). It is worth to highlight that the current study reinforces the earlier studies on South Asian T2D genetics(39). In addition, *WDR11* has previously been associated with fasting glucose(32), type 2 diabetes susceptibility(32) and youth-onset T2D(33).

Strengths and Limitations

One of the strengths of this study is that we address the genetic basis for the large differences in age of diagnosis of T2D genetics between two ethnic groups. To date, this is the first study that demonstrates the genome wide PRS of age at diagnosis of T2D in South Indians. We believe that our polygenic risk scores derived from Asian Indian based GWAS can be useful for the population specific studies in South Asians although it cannot be used to predict in Europeans where the age diagnosis of T2D is much higher and thus likely, the genetics of age at diagnosis of T2D can be different. While T1D genetic risk score is portable between the Indian and Scottish (Supplementary Fig S12). We also note that transferability of PRS across different ethnic groups demands careful evaluation.

One of the limitations in our study is the modest sample size of South Asian Indian samples for the GWAS study, which limits our ability to identify associations with low-frequency variants. Next, our study cohort was limited only to the South Indian population and South Asian living in UK; thus, the findings and the biological interpretation of the significant South Indian T2D polygenic effects reported here, needs further validation using an independent South Asian Indian cohort. GAD65 and ZnT8 antibody testing was carried out in the South Indian cohort to ensure exclusion of individuals with type 1 diabetes from analysis.

In conclusion, our study demonstrated the association of several previously established loci in European T2D GWAS for age of diagnosis of T2DM. However, we observed substantial heterogeneity in both the effect sizes and/or the allele frequencies between the ethnic groups. Furthermore, the higher heritability estimates of age of onset of type 2 diabetes in South Indians demonstrates the importance of further study of the genetic architecture of age of onset of type 2 diabetes in the ancestral group of South Asians.

Article Information

Acknowledgements

We are thankful to all the families who took part in this study. We are grateful to the GoDARTS, SHARE and DMDSC teams, including interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants, and nurses, for their cooperation in recruiting them. We would like to acknowledge Dundee Health Informatics Centre (HIC) for managing and providing anonymised data.

Guarantor's statement

Prof. Colin NA Palmer and Dr. Sundararajan Srinivasan are the guarantor 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.

Funding

This work was supported by the National Institute for Health Research using Official Development Assistance (ODA) funding [INSPIRED 16/136/102]. The Wellcome Trust United Kingdom Type 2 Diabetes Case-Control Collection (supporting GoDARTS) was funded by The Wellcome Trust (072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT program. The present study was conducted using the UK Biobank Resource under application No. 20405.

Duality of Interest

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

Author Contributions

SS and CNAP designed the study with input from ERP, VR and VM. VR, RMA, VM and CNAP oversaw data collection. VR, SS, RMA, CNAP and VM coordinated the study. SS, MKS, RMA, AD, ERP AND CNAP had access to and verified the raw data. SL and NS, and in conjunction with the VR and SS, carried out data curation and genotyping. The data analysis, and interpretation were carried

out by SS. SS and CNAP wrote the first draft of the paper, and all other authors contributed and revised it critically for important intellectual content. All authors approved the final version for publication, and CNAP had final responsibility for the decision to submit. Funding was acquired by CNAP.

Conflict of Interest / Disclosures
We declare no competing interest.

Data sharing
Summary data might be made available upon reasonable request via email to the lead and corresponding author.

References

1. Kooner JS, Saleheen D, Sim X, Sehmi J. Genome-wide association study in people of South Asian ancestry identifies six novel susceptibility loci for type 2 diabetes. 2013;43(10):984–9.
2. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536(7614):41–7.
3. Cho YS, Chen CH, Hu C, Long J, Hee Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2012;44(1):67–72.
4. HR, Rao Gundu M V. Type 2 Diabetes in South Asian: Epidemiology, Risk Factors and Prevention. 1/e. 10.5005/jp/books/10994; 2007. 400 p.
5. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V., Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90.
6. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579–89.
7. Saeedi P, Salpea P, Karuranga S, Petersohn I, Malanda B, Gregg EW, et al. Mortality attributable to diabetes in 20-79 years old adults, 2019 estimates: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pr. 2020;162:108086.
8. Mohan V, Deepa M, Deepa R, Shanthirani CS, Farooq S, Ganesan A, et al. Secular trends in the prevalence of diabetes and impaired glucose tolerance in urban South India - The Chennai Urban Rural Epidemiology Study (CURES-17). Diabetologia. 2006;49(6):1175–8.
9. Anjana RM, Rani CSS, Deepa M, Pradeepa R, Sudha V, Nair HD, et al. Incidence of diabetes and prediabetes and predictors of progression among Asian Indians: 10-year follow-up of the Chennai urban rural epidemiology study (CURES). Diabetes Care. 2015;38(8):1441–8.
10. Shah A, Kanaya AM. Diabetes and associated complications in the South Asian population. Curr Cardiol Rep. 2014;16(5).
11. Sattar N, Rawshani A, Franzén S, Rawshani A, Svensson AM, Rosengren A, et al. Age at Diagnosis of Type 2 Diabetes Mellitus and Associations With Cardiovascular and Mortality Risks: Findings From the Swedish National Diabetes Registry. Circulation. 2019;139(19):2228–37.
12. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature. 2020;582(7811):240–5.
13. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, 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):560–72.
14. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018;9(1):1:14.

15. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Dia APM (SAT2D). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46(3):234–44.
16. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018 Nov 1;50(11):1505–13.
17. Pradeepa R, Prabu AV, Jebarani S, Subhashini S, Mohan V. Use of a large diabetes electronic medical record system in India: Clinical and research applications. J Diabetes Sci Technol. 2011;5(3):543–52.
18. Hébert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, et al. Cohort profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). Int J Epidemiol. 2018;47(2):380–1.
19. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: The Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. BMJ Open. 2017;7(2):1–8.
20. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.
21. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA)—an extension of the STROBE statement. Genet Epidemiol. 2009;33(7):581–98.
22. Organization WH, Federation ID. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia : report of a WHO/IDF consultation. Geneva PP - Geneva: World Health Organization;
23. Mcgurnaghan SJ, Blackbourn LAK, Caparrotta TM, Mellor J, Barnett A, Collier A, et al. Cohort profile: The Scottish Diabetes Research Network national diabetes cohort - A population-based cohort of people with diabetes in Scotland. BMJ Open. 2022;12(10):1–10.
24. Siddiqui MK, Hall C, Cunningham SG, McCrimmon R, Morris A, Leese GP, et al. Using Data to Improve the Management of Diabetes: The Tayside Experience. Diabetes Care. 2022 Dec 1;45(12):2828–37.
25. Wright AK, Welsh P, Gill JMR, Kontopantelis E, Emsley R, Buchan I, et al. Age-, sex- and ethnicity-related differences in body weight, blood pressure, HbA1c and lipid levels at the diagnosis of type 2 diabetes relative to people without diabetes. Diabetologia. 2020;63(8):1542–53.
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Appl NOTE. 2010;26(17):2190–1.
27. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–5.
28. Chambers JC, Abbott J, Zhang W, Turro E, Scott WR, Tan ST, et al. The South Asian genome. PLoS One. 2014;9(8).
29. Liju S, Chidambaram M, Mohan V, Radha V. Impact of type 2 diabetes variants

identified through genome-wide association studies in early-onset type 2 diabetes from South Indian population. Genomics and Informatics. 2020;18(3):1–12.
30. Chidambaram M, Liju S, Saboo B, Sathyavani K, Viswanathan V, Pankratz N, et al. Replication of genome-wide association signals in Asian Indians with early-onset type 2 diabetes. Acta Diabetol. 2016;53(6):915–23.
31. Sim X, Ong RTH, Suo C, Tay WT, Liu J, Ng DPK, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. PLoS Genet. 2011 Apr;7(4):1001363.
32. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105–16.
33. Srinivasan S, Chen L, Todd J, Divers J, Gidding S, Chernausek S, et al. The First Genome-Wide Association Study for Type 2 Diabetes in Youth: The Progress in Diabetes Genetics in Youth (ProDiGY) Consortium. Diabetes. 2021;70(4).
34. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012;44(6):659–69.
35. Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I, Grarup N, Sebastian D, Rodriguez-Fos E, et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat Commun. 2018;9(1):1–14.
36. Choi SW, Shin T, Mak H, Reilly PFO. A guide to performing Polygenic Risk Score analyses Introduction to Polygenic Risk Scores. Nat Protoc. 2021;15(9):2759–72.
37. Siddiqui MK, Anjana RM, Dawed AY, Martoeau C, Srinivasan S, Saravanan J, et al. Young-onset diabetes in Asian Indians is associated with lower measured and genetically determined beta cell function. Diabetologia. 2022;65(6):973–83.
38. Dabelea D, Dolan LM, D'Agostino R, Hernandez AM, McAteer JB, Hamman RF, et al. Association testing of TCF7L2 polymorphisms with type 2 diabetes in multi-ethnic youth. Diabetologia. 2011;54(3):535–9.
39. Saxena R, Saleheen D, Been LF, Garavito ML, Braun T, Bjonnes A, 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):1746–55.

Table 1. Summary statistics of the most significant SNPs from meta-analysis

SNP	CHR	POS	EA/NEA	Study Cohort	BETA	SE	EAF	P value	Het P.value	Gene name
rs7903146	10	114758349	T/C	DMDSC data freeze1	−1·26	0·20	0·35	1.0×10^{-10}		*TCF7L2*
				DMDSC data freeze 2	−0·40	0·36	0·34	$2\cdot7\times10^{-03}$	0·02	
				UKBB(SAS)	−0·33	0·37	0·35	0·3	$2\cdot7\times10^{-06}$	
				META (SAS)	−0·92	0.16	0·34	$1\cdot1\times10^{-08}$	0·01	
				GoDARTS	−0·02	0·17	0·34	0·11		
				GoSHARE	−0·90	0·26	0·32	0·0008		
				UKBB (Europeans)	−0·35	0·08	0·35	$1\cdot3\times10^{-05}$		
				META (Europeans)	−0·02	0·02	0·34	0·004		
				Trans-ethnic Meta	−0·05	0·08	0·35	$2\cdot4\times10^{-12}$		
rs9368219	6	20674691		DMDSC data freeze 1	−1·20	0·21	0·25	$4\cdot3\times10^{-08}$		*CDKAL1*
				DMDSC data freeze 2	−0·38	0·38	0·25	$6\cdot1\times10^{-02}$	0·002	
				UKBB(SAS)	−0·59	0·40	0·27	0.14	0·02	
				META (SAS)	−0·92	0·17	0·26	$6\cdot6\times10^{-08}$	0·007	
				GoDARTS	−0·03	0·02	0·18	0·004		
				GoSHARE	−0·80	0·30	0·19	0·007		
				UKBB (Europeans)	−0·17	0·09	0·17	0·07		
				META (Europeans)	−0·05	0·02	0·19	0·03		
				Trans-ethnic Meta	−0·05	0·23	0·21	2.3×10^{-08}		
rs3011366	10	122554701	G/A	DMDSC data freeze 1	1·35	0·32	0·10	$3\cdot1\times10^{-05}$		*WDR11*
				DMDSC data freeze 2	1·02	0·57	0·10	0·07	0·25	
				UKBB(SAS)	2·44	0·68	0·08	$3\cdot4\times10^{-04}$	0·85	
				META (SAS)	1·44	0·26	0·09	3.3×10^{-08}	0·0001	
				GoDARTS	0·21	0·11	0·01	0·36		
				GoSHARE	−0·31	1·09	0·01	0·77		
				UKBB (Europeans)	−0·21	0·41	0·01	0·60		
				META (Europeans)	−0·09	0·08	0·01	0·20		
				Trans-Ethnic Meta	0·21	0·07	0·01	0·01		

CHR, Chromosome; POS, Position; EA, Effect Allele; NEA, non-effect allele; SE, Standard Error; EAF, Effect allele frequency.

Table 2. Summary statistics of the Lead SNPs from meta-analysis using stratified age at diagnosis cohort

Cohort with Diabetes	SNP	EA/NEA	Study Cohort	Sample Size N	BETA	SE	EAF	P value	Het P.value	Gene Name
Young Onset Diabetes (20-55Years)	**rs7903146**	T/C	DMDSC data freeze 1	5191	−1·10	0.20	0·35	8.6×10^{-8}		*TCF7L2*
			DMDSC data freeze 2	2102	−0·62	0.29	0·34	0.03	*0·13*	
			UKBB(SAS)	542	−0·38	0.35	0·35	0·4	*0.78*	
			META (SAS)	7835	−0·84	0·15	0·34	<0.0001		
			GoDARTS	768	−0·43	0·26	0·34	0·11		
			GoSHARE	543	−0·37	0·35	0·36	0·0008		
			UKBB (Europeans)	4991	−0·26	0·01		0.015		
			META (Europeans)	6302	−0·29	0·08		0·001		
Older Onset Diabetes (Over 55years)	**rs7903146**	T/C	DMDSC data freeze 1	610	0.11	0·27	0·35	0.64	*0.93*	*TCF7L2*
			DMDSC data freeze 2	392	−0.08	0·47	0·33	0.54	*0·93*	
			UKBB(SAS)	266	0.02	0·27	0·36	0.70		
			META (SAS)	1268	0.05	0·17	0·34	0.80		
			GoDARTS	6231	−0·08	0·14	0·33	0·55		
			GoSHARE	3612	−0·03	0.18	0·36	0·83		
			UKBB (Europeans)	8753	−0.10	0·05		0.06		
			META (Europeans)	18596	−0·09	0·04		0·03		
Young Onset Diabetes (20-55Years)	**rs9368219**	T/C	DMDSC data freeze 1	5191	−1·06	0·19	0·25	4.6×10^{-8}	*0·08*	*CDKAL1*
			DMDSC data freeze 2	2102	−0·11	0·30	0·25	0.06	*0.48*	
			UKBB(SAS)	542	−0·53	0·38	0·27	0·04		
			META (SAS)	7835	−0·74	0·29	0·18	<0.001		
			GoDARTS	768	−0·05	0·30	0·19	0·08		
			GoSHARE	543	−0·50	0.39	0·17	0·07		
			UKBB (Europeans)	4991	−0·008	0·12		0.9		
			META (Europeans)	6302	−0·05	0·01		0·63		
Older Onset Diabetes (Over 55years)	**rs9368219**	T/C	DMDSC data freeze 1	610	0.11	0·27	0·25	0.68	*0·95*	*CDKAL1*
			DMDSC data freeze 2	392	0.02	0·47	0·25	0.54	*0.21*	
			UKBB(SAS)	266	0.16	0·25	0·27	0·5		
			META (SAS)	1268	0·12	0·29	0·18	0.39		
			GoDARTS	6231	−0·04	0·15	0·19	0·08		
			GoSHARE	3612	−0·38	0.21	0·17	0·07		
			UKBB (Europeans)	8753	0.002·	0·06		0.9		
			META (Europeans)	18596	−0·02	0·12		0·59		

CHR, Chromosome; POS, Position; EA, Effect Allele; NEA, non-effect allele; SE, Standard Error; EAF, Effect allele frequency.

Figure Legend

Figure1. Forest plot for the top significant SNP rs7903146 near *TCF7L2* (Effect allele - T) in individuals with T2D. DMDSC Dr. Mohan Diabetes speciality Clinic, UKBSAS UK Biobank South Asians. GoDARTS, Genetics of Diabetes Audit and Research in Tayside Scotland; GoSHARE, Genetics of Scottish Health Research Register; UKBB Euro, United Kingdom Biobank Europeans. A) Overall meta-analysis of GWAS of age of diagnosis of T2D; B) South Asian Indians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); C) South Asian Indians with later onset of T2D (Over 55 years); D) Europeans with earlier onset of T2D (Age at diagnosis between 20 – 55 years); E) Europeans with later onset of T2D (Over 55 years).

Figure 2. Performance of South Indian PRS in European population. PRS generated using South Indian summary data from DMDSC cohort 1 and tested in European samples GoSHARE (N 4,155) and South Indian independent cohort (DMDSC 2 (N 2,494)) for polygenic risk prediction of age of onset.