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 onset of Type 2 diabetes in South Indians and White Europeans

Manuscript word count (not including tables, figures, and references): 4092

ADDRESS FOR CORRESPONDENCE:

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

Abstract

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 onset in these populations may give insight into earlier age at onset of T2D among individuals of South Asian descent. We conducted a meta-analysis of GWAS of age at diagnosis of T2D in 34,001 individuals from four independent cohorts of European and Asian Indians. We identified two signals near the *TCF7L2* and *CDKAL1* associated with age at onset of T2D. The strongest genome-wide significant variants at chromosome 10q25.3 in *TCF7L2* (lead SNP rs7903146; $p = 2.4 \times 10^{-12}$; p-het =0.01; Beta = -0.436; SE = 0.02) and chromosome 6 p22.3 in *CDKAL1* (rs9368219; $p = 2.29 \times 10^{-8}$; p-het =0.007; 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 Asian cohorts. A genome wide signal was also obtained at chromosome 10q26.12 in *WDR11* (rs3011366; $p = 2.4 \times 10^{-8}$; p-het =0.25; Beta = 1.44; SE=0.25) specifically in the South Asian Indian cohorts. Our study estimated 17% of heritability (h^2) for age onset of T2D among South Indians; higher than the 5% heritability we observed in Europeans suggesting different genetic architectures between these populations. We further estimated the Asian Indian genome-wide polygenic risk score for T2D risk and identified that the variance explained by the Indian specific polygenic risk score (PRS) for age at onset of T2D is substantially higher (2%) in an independent cohort of Asian Indians compared to that seen in White Europeans (<0.1%). This reveals that although variants do exist that are shared between the two population there are also genetic disparities of age onset of T2D between the ethnicities.

Author Summary

Recent large multi-ancestry genome-wide meta-analyses have identified over 277 genetic loci associated with type 2 diabetes (T2D). Most association studies for T2D have focused on populations of European ancestry, and the specific genetic architecture of T2D in the South Asian population has not been extensively investigated. Increasing evidence indicates that the prevalence of T2D is higher in South Asians with a strikingly earlier age at diagnosis when compared to Europeans. Genome-wide association studies (GWAS) on age at age at diagnosis of T2D are limited and can help elucidate underlying mechanisms. We conducted the largest meta-analysis to date of genome-wide association studies of age at diagnosis as a surrogate for age of onset of T2D in South Indians and Europeans. Our trans-ancestry meta-analysis identified genetic loci near *TCF7L2* and *CDKAL1* with substantial ethnic heterogeneity in both allele frequencies and effect sizes in early diagnosis with T2D. Two novel variants near *TCF7L2* (rs570193324) and *CDKAL1* (rs143316471) were associated with age of diagnosis of T2D only in South Indians and were present at much lower frequency in the European populations. Heritability estimates for age of onset were much stronger in Asian Indians compared to Europeans and a polygenic risk score was constructed using a South Indians, which explained about 2% trait variance compared to European ancestry (<0.1%). Our novel findings provide a better understanding of ethnic differences in the age at onset of T2D and indicate the potential importance of ethnic differences in the genetic architecture underpinning T2D.

Introduction

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 [14–16].

Currently nearly 250 genetic loci (more than 400 unique genetic variants) have been identified that influence T2D[2,12,13]. 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[14]. 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 [17]. Previously identified genetic variants explained about (~10%) of the heritability of T2D [15] .

Despite advancement in genetic research tools, South 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 Asians and Europeans. We aimed to develop, evaluate, and understand a T2D age at diagnosis PRS in Asian Indians (DMDSC) and Europeans (GoSHARE) used in the study cohort. In this multicentre study, we focused on inter-ancestry differences in the genetics of age at onset of T2D that might influence ethnic-ancestry differences in health outcomes in general and T2D.

Results

A total of 34,001 participants with T2D were included for this study after-quality control filtering: 8,295 T2D patients of Asian Indians ancestry from a large tertiary diabetes centre, Dr Mohan's Diabetes Specialties Centre (DMDSC), 6,999 patients of European ancestry from Genetics of Diabetes Audit and Research Tayside and Scotland (GoDARTS)[16], 4,155 patients of European ancestry from Scottish Health Research Register (SHARE)[17] and 14,552 T2D participants from United Kingdom Biobank (UKBB). 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 (Supplement Fig 1). A detailed illustration of the study design is presented in Fig 1. The population characteristics of the

cohorts are described in Table 1. Notably, we observed the average age of diagnosis of T2D in the Asian Indians was 40 years, whereas in the white Europeans, the mean age of diagnosis was 60 years.

SNP-based heritability

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

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

The genome-wide association analyses of age at diagnosis of type 2 diabetes were conducted for each cohort separately using a linear mixed model as implemented in BOLT-LMM. We then conducted trans-ancestry meta-analyses by utilising Haplotype Reference Consortium (HRC) imputed data up to 26.2 million SNPs directly genotyped or successfully imputed at high quality across all the data sets. Association statistics and P values for heterogeneity were combined using fixed-effects meta-analyses as implemented in METAL [18]. Our meta-analysis revealed two previously known T2D loci at chromosome 10 q25.2 near Transcription factor 7-like 2 (rs79603146, *TCF7L2*, $P < 2.4 \times 10^{-12}$, 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, $P < 2.29 \times 10^{-8}$, Beta -0.053; SE 0.01; P-het =0.007) associated with T2D age-diagnosis (Fig 2). The allelic frequency of CDKAL1 is more common in DMDSC Asian Indian cohort (MAF = 0.26) compared to Caucasians (MAF =0.18) The lead SNPs at *TCF7L2* and *CDKAL1* (Fig 3A and 4A) 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 it being . Interestingly the effect size of the variants was much lower in the cohorts of European descent. Ethnic-specific meta-analysis results are presented in supplement tables.

Stratification of Type 2 Diabetes by Age Onset

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.

Based on the South Indian mean age of diagnosis (Table 1), the study participants in both ethnicities were stratified into the early-diagnosis T2D group (20-55 years) (Fig 3B & D) and late diagnosis T2D (diagnosed after 55 years) (Fig 3C & E). 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 3B, D & 4B, D). These variants have very little effect on age of diagnosis in those with diabetes diagnosed after 55 years of age in either ethnicity (Fig 3C, E& 4C, E).

The role of other T2D variants in Age of diagnosis

We identified several previously reported T2D variants as suggestive signals ($P < 1 \times 10^{-5}$) in this age onset of T2D trans-ethnic meta-analyses (Supplement Table). In particular, the risk variant nearby *SEC24B* at chromosome location 4q25 (rs76170449, $P< 1.79\times10^{-7}$) is also associated with cardiovascular traits, and 3p24.3 *ZNF385D* (rs17011243, $P< 1.13\times10^{-5}$) 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 < 3.40 × 10^{-6}) and 17q21.2 (rs684214, *MLX*, P< 2.40 ×10^{-6}) with no difference in their effect estimates between two distinct ethnic groups (Supplement Table). We replicated previously reported South Asian T2D genome-wide signals[19–22] with suggestive evidence or a nominal association for age onset 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 Asian Indian cohorts

In the meta-analysis of only the Asian Indian cohorts, we also found an additional novel genome-wide signal at chromosome 10 q26.12 near the *WDR11* region (rs3011366, P < 2.46 ×10^{-8}, Beta 1.44, SE 0.25). However, this variant was not associated the age-of-onset in the European cohorts (Table 2). 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[23]. The Conditional analyses conducted in South Indian ancestry (Table 3), 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

182 Indian Cohorts. The regional plot for independent association of *TCF7L2* and *CDKAL1* is

183 shown in Fig 5.

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[24]. Several other SNPs reached suggestive significance for age onset of T2D, and the direction of effect was consistent across all cohorts of European descent *(Supplementary Table).*

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

Polygenic Risk Scores (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[25]. To investigate whether different genetic variants shared between ethnicities were conferring the risk of the onset of T2D: first, we constructed a PRS for all genome wide significant SNPs associated with age of diagnosis using the DMDSC 1 cohort of the Asian Indian genetic data (n=5801) and validated this using the DMDSC 2 of Asian Indian data which contained no overlapping participants; next we assessed the performance of this South Asian GRS in the European cohort (GoSHARE) (Fig 6). The PRS replicated strongly between the Asian Indian cohorts. On the other hand, the Asian 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 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 Asian 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. 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 Asian 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 Asian Indians that were very rare in the European cohort. Our findings emphasize and support our recently reported finding that Asian Indians have greater genetic beta-cell dysfunction compared to Europeans[26].

The role of beta cell function as a driver for the early age onset of T2D in Asian Indians is well supported from our ethnic specific TCF7L2 and CDKAL1 signals. In addition, WDR 11 has previously been associated with fasting glucose[27], but not type 2 diabetes susceptibility per se.

One of the strengths of this study is that we address the lack of transferability and consistency of underlying age onset of T2D genetics between two ethnic groups. To date, this is the first study that demonstrates the genome wide PRS of age at onset T2D in Asian Indians. Overall, our polygenic risk scores findings derived from Asian Indian based GWAS can be useful for the population specific studies and it cannot be used to predict in Europeans where the genetics of T2D age onset itself is very distinct and unique between two ethnic backgrounds. We also noted that transferability of PRS across different ethnic groups demands careful evaluation.

One of the limitations in our study is the modest sample size of 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 Asian Indian population and South Asian living in UK; thus, findings from the present study merits further validation in an independent larger discovery cohort. The biological interpretation of the significant Asian Indian T2D polygenic effects reported here needs further validation using an independent South Asian Indian cohort.

In conclusion, our study demonstrated the association of several previously established loci in European T2D GWAS for age onset 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 Asian Indians demonstrates the importance of further study of the genetic architecture of age of onset of type 2 diabetes in this ancestral group.

Materials and Methods

Ethics statement

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 Research Ethics Committees approved the study.

Study participants

We included participants from four independent cohorts: Dr Mohan's Diabetes Specialties Centre (DMDSC), Chennai, India, Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS), Genetics of Scottish Health Research Register (GoSHARE) and the United Kingdom Biobank (UKBB). DMDSC is a diabetes centre of single speciality hospitals and clinics established in 1991 in Chennai, Southern India, which includes currently there are 50 clinics in various locations across 10 states in India[28]. A total of 500,000 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. 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[16]. GoSHARE currently comprised of a biobank of around 74,000 individuals across NHS Fife, and NHS Tayside[17]. 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 2006-2010 at age 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.

Phenotyping – Age at Diagnosis of Type 2 Diabetes

We included 8,295 type 2 diabetes patients from DMDSC cohort of Asian Indians, whose first clinical visit was within one year of diagnosis. Diabetes is diagnosed by general practitioners using WHO criteria[29] 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 GADA positive. We selected the study participants in the GoDARTS, and GoSHARE based on the following inclusion criteria, aged between 20 and 80 years, the type 2 diabetes status diagnosed by general practitioners, and the clinical measurements as per WHO guidelines and recorded in the Scottish Care Information-Diabetes system. 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 Controls, and Imputations

DMDSC genotyping was conducted for approximately 5,801 patients with type 2 diabetes by Illumina using the global screening arrays version (GSA v1.0); for the remaining participants (N=2,494) were genotyped on 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. We excluded the SNPs with less than 97 % call rate and HWE p-value less than $1e^{-6}$ (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[30].

Genotyping of GoDARTS and SHARE cohorts 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 $p< 1 \times 10^{-6}$. Following sample and SNP quality control, Haplotype Reference Consortium (HRC) panel for both populations using the Michigan Imputation Server[31]. We retained all imputed SNPs with imputation information score > 0.4 and removed monomorphic markers.

The genome-wide genotyping for 488,377 participants in the UKBB was performed using custom-designed genotyping arrays including UK Biobank Axiom and UK BiLEVE Axiom Affymetrix array. We selected European and South Asian individuals in the UKBB based on PCA and self-reported ethnic background. A total of 14,552 participants fulfilled the phenotype criteria in this study. UK biobank genotyping, QC, PCA and imputation protocol are described elsewhere[32]. For the current study, we restricted the analyses to the UK Biobank participants who were in the full release imputed genomics datasets. We selected

319 individuals of Asian Indians and European descents based on the principal component
320 analysis (PCA) and the self-reported ancestry information.

321 **Ethnic-Specific Meta-analysis of GWAS**

322 Genome-wide association analyses were performed independently for each cohort using
323 an additive model while adjusting for sex (Fig 1). We estimated allelic effects using the
324 using a linear mixed model as implemented in BOLT-LMM version 2.3.2[33] which
325 accounts for relatedness and any population stratification and SNPTESTv2.5 in each
326 cohort accordingly. We performed a meta-analysis based on ancestry: Asian Indians
327 specific analyses include the DMDSC cohort, a unique Asian Indians representative data,
328 and South Asians in the UKBB and the European specific analyses include the GoDARTS,
329 GoSHARE, and White Europeans in the UKBB. There was no evidence of population
330 stratification in meta-analysis (genomic inflation factor, $\lambda = 1.007$). We performed the
331 meta-analyses using a fixed-effect method in METAL software, which assumes the effect
332 allele is the same for each study within an ancestry. We then combined the summary
333 statistics of GWAS from all the study populations. SNPs with imputation quality score <
334 0.40 were excluded from the analysis. Heterogeneity across these studies was assessed
335 by the I^2 (low to high) and Cochran's Q statistics as reported by METAL. The Forest plot
336 was generated using the metafor package[34]. We annotated the genetic variants using
337 the University of California Santa Cruz (UCSC) Genome resource18 based on the
338 Genome Reference Consortium Human genome build 37.

339 **Conditional analysis**

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

342 **SNP-based heritability**

343 We used the summary statistics data from the Asian Indians and Europeans specific meta-
344 analyses to estimate the SNP-based heritability in a liability scale using Linkage
345 Disequilibrium Score Regression (LDSC) software[35].

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

347 For the Polygenic risk scores (PRS), we considered summary statistics of GoSHARE and
348 DMDSC samples. The PRSice tool generates the scores by the weighted sum of the risk
349 allele carried by individuals based on effect estimate. We removed DNA polymorphisms
350 with ambiguous strands (A/T or C/G) from the score derivation. SNPs were clumped to a
351 more significant SNP in LD ($r2 \geq 0.10$) within a 500 kb window. The PRS calculation
352 considered several p-value thresholds (0.001, 0.05, and 0.1).

Supporting Information

Acknowledgements

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. 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.

Author Contributions

Study concept and design: CNAP.

Data collection and access: GoDARTS, GoSHARE, UKBB - CNAP, AD, ERP

Data collection and access: DMDSC - MV, ARM, RV

Data processing: MKS, SS, SL, NS.

Data analysis: SS.

Drafting of the manuscript: SS and CNAP.

Critical revision of the manuscript: All authors involved in the critical revision and approved the final version for publication.

Conflict of Interest / Disclosures:

Role of the Funder/Sponsor:
The funders had no role in design and conduct of the study; data collection and analysis, preparation of the manuscript or approval of the manuscript and decision to publish this work.

References

1. Kooner J. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43: 984–989.
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: 41–47. doi:10.1038/nature18642
3. Cho Y. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2011;44: 67–72.
4. Mohan V, Rao GH. Type 2 Diabetes in South Asians: Epidemiology, Risk Factors & Prevention. Type 2 Diabetes in South Asians: Epidemiology, Risk Factors & Prevention. 2007. doi:10.5005/jp/books/10994
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. Nature Genetics. 2012. doi:10.1038/ng.2383
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. Nature Genetics. 2010;42: 579–589. doi:10.1038/ng.609
7. Saeedi P, Salpea P, Karuranga S, Petersohn I, Malanda B, Gregg EW, et al. Mortality attributable to diabetes in 20â€“79Â years old adults, 2019 estimates: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. 2020 [cited 22 Jun 2021]. doi:10.1016/j.diabres.2020.108086
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: 1175–1178. doi:10.1007/s00125-006-0219-2
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: 1441–1448. doi:10.2337/DC14-2814
10. Shah A, Kanaya AM. Diabetes and associated complications in the South Asian population. Current Cardiology Reports. Current Medicine Group LLC 1; 2014. p. 476. doi:10.1007/s11886-014-0476-5
11. Sattar et al Vascular Outcomes by Age of Diabetes Diagnosis. 2019. doi:10.1161/CIRCULATIONAHA.118.037885
12. 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. Nature Genetics. 2018;50: 1505–1513. doi:10.1038/s41588-018-0241-6
13. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature Genetics. 2014;46: 234–244. doi:10.1038/ng.2897
14. Spracklen CN, Horikoshi M, Jin Kim Y, Lin K, Bragg F, Moon S, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals Genetic discovery from association analyses. Nature. [cited 17 Jun 2021]. doi:10.1038/s41586-020-2263-3

15. 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. Nature Communications. 2018. doi:10.1038/s41467-018-04951-w
16. 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). International Journal of Epidemiology. 2018;47: 380–381j. doi:10.1093/ije/dyx140
17. 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: e013351. doi:10.1136/BMJOPEN-2016-013351
18. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. BIOINFORMATICS APPLICATIONS NOTE. 2010;26: 2190–2191. doi:10.1093/bioinformatics/btq340
19. Chambers JC, Abbott J, Zhang W, Turro E, Scott WR, Tan ST, et al. The South Asian genome. PLoS ONE. 2014;9: 102645. doi:10.1371/journal.pone.0102645
20. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature Genetics. 2014;46: 234–244. doi:10.1038/ng.2897
21. 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 & Informatics. 2020;18: 1–12. doi:10.5808/GI.2020.18.3.E27
22. Chidambaram M, Liju S, Saboo • Banshi, Sathyavani K, Viswanathan V, Pankratz N, et al. Replication of genome-wide association signals in Asian Indians with early-onset type 2 diabetes. Acta Diabetologica. 53. doi:10.1007/s00592-016-0889-2
23. Boehnke M, Mccarthy MI, Genet N. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk HHS Public Access Author manuscript. Nat Genet. 2010;42: 105–116. doi:10.1038/ng.520
24. Manning AK, Hivert M-F, 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. Nature Genetics. 2012. doi:10.1038/ng.2274
25. Choi SW, Shin T, Mak H, O'reilly PF. A guide to performing Polygenic Risk Score analyses. doi:10.1101/416545
26. 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 [cited 1 Apr 2022]. doi:10.1007/S00125-022-05671-Z
27. Boehnke M, Mccarthy MI, Genet N. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk HHS Public Access Author manuscript. Nat Genet. 2010;42: 105–116. doi:10.1038/ng.520
28. 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. Journal of Diabetes Science and Technology. SAGE Publications Inc.; 2011. pp. 543–552. doi:10.1177/193229681100500309
29. definition and diagnosis of diabetes mellitus and intermediate hyperglycemia RepoRt of a WHo/IDf ConsultatIon. 2006.

30. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016;48: 1279–1283. doi:10.1038/ng.3643
31. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nature Genetics. 2016;48: 1284–1287. doi:10.1038/ng.3656
32. Bycroft C, Freeman C, Petkova D, Band G, Elliott L, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv. 2017; 166298. doi:10.1101/166298
33. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics. 2015;47: 284–290. doi:10.1038/ng.3190
34. Viechtbauer W. Conducting meta-analyses in R with the metafor. Journal of Statistical Software. 2010;36: 1–48. doi:10.18637/jss.v036.i03
35. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genomewide association studies. Nature GeNetics VOLUME. 2015;47. doi:10.1038/ng.3211

Table 1. Characteristics of the study population.

Study Cohort	N	Age at Diagnosis/ Onset mean age in years ± SD	BMI	Gender Male: Female (%)	Ethnicity
DMDSC 1	5,801	41.7±10.17	26.3	61.5;38.5	Asian Indians
DMDSC 2	2,494	41.3±9.77	26.3	58.17;41.8	Asian Indians
UKBB	808	50.1±11.2	28.7	73.7;26.3	South Asians
GoDARTS	6,999	59±10.12	31	54.9;44.8	Europeans
SHARE	4,155	58.2±10.5	31.6	53.5;46.4	Europeans
UKBB	13,744	54.6±10.6	32	69.5;36.3	Europeans

494

Table 2. Summary statistics of the most significant SNPs from meta-analysis

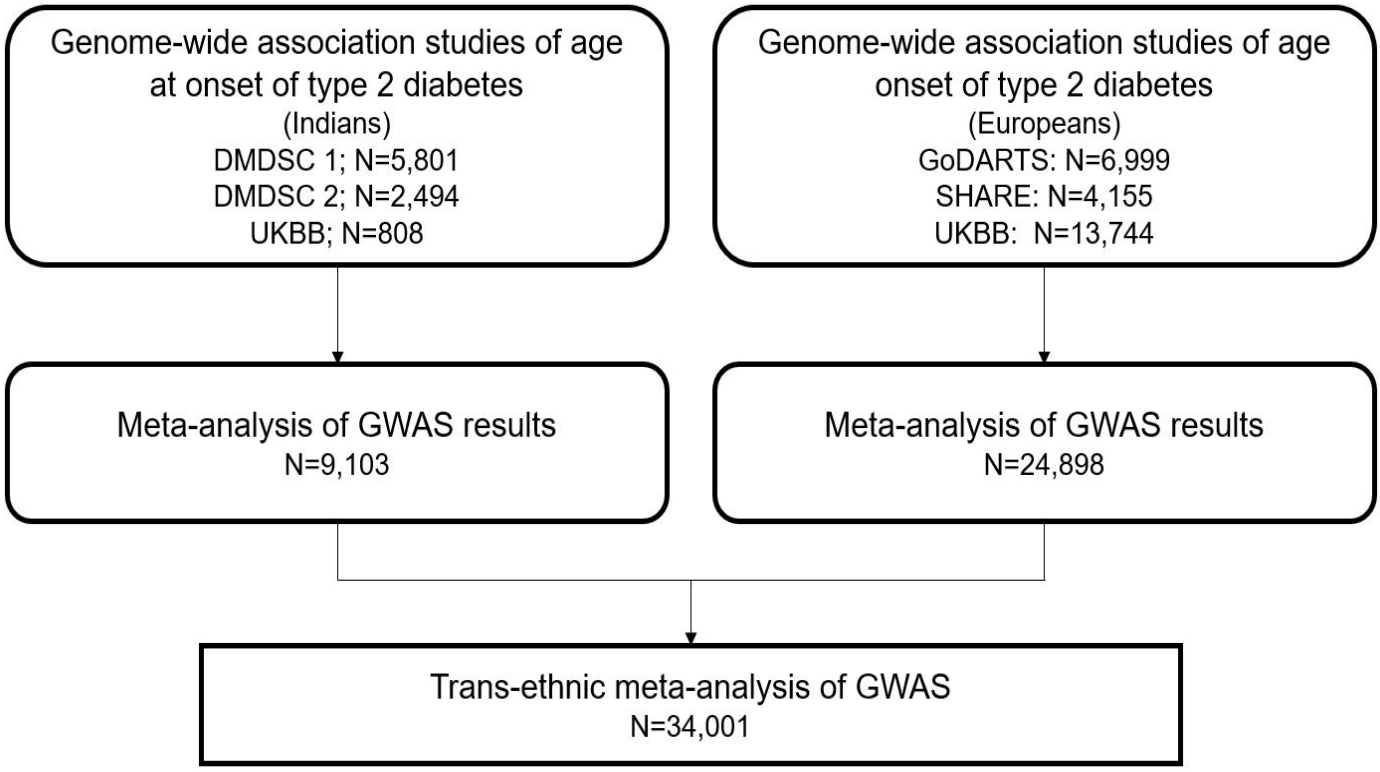
SNP	CHR	POS	EA/NEA	Study COHORT	BETA	SE	EAF	P value	Het Pval	Gene
rs7903146	10	114758349	T/C	DMDSC data freeze 1	-1.26	0.20	0.35	1×10^{-10}		
				DMDSC data freeze 2	-0.40	0.36	0.34	2.7×10^{-03}		
				UKBB(SAS)	-0.33	0.37	0.35	0.3		
				META (SAS)	-0.72	0.32	0.34	1.08×10^{-08}	*0.02*	*TCF7L2*
				GoDARTS	-0.02	0.17	0.34	0.11		
				SHARE	-0.90	0.26	0.32	0.0008		
				UKBB (Europeans)	-0.35	0.08	0.35	1.3×10^{-05}		
				META (Europeans)	-0.02	0.017	0.34	0.004	*2.72e-06*	
				Trans-ethnic Meta-Analysis	-0.43	0.02	0.35	2.4×10^{-12}	*0.01*	
rs9368219	6	20674691		DMDSC data freeze 1	-1.20	0.21	0.25	4.3×10^{-08}		
				DMDSC data freeze 2	-0.38	0.38	0.25	6.1×10^{-02}		
				UKBB(SAS)	-0.59	0.40	0.27	0.14		
				META (SAS)	-0.92	0.17	0.26	6.59×10^{-08}	*0.002*	*CDKAL1*
				GoDARTS	-0.03	0.02	0.18	0.004		
				SHARE	-0.8	0.30	0.19	0.007		
				UKBB (Europeans)	-0.17	0.09	0.17	0.07		

				META (Europeans)	-0.046	0.02	0.19	0.03	*0.02*	
				Trans-ethnic Meta-Analysis	-0.05	0.23	0.21	2.29×10^{-08}	*0.007*	
rs3011366	10	122554701	G/A	DMDSC data freeze 1	1.35	0.32	0.10	3.1×10^{-05}		*WDR11*
				DMDSC data freeze 2	1.02	0.57	0.10	0.07		
				UKBB(SAS)	2.44	0.68	0.08	3.4×10^{-04}		
				META (South Asians)	1.44	0.25	0.09	2.4×10^{-08}	*0.25*	
				GoDARTS	0.21	0.11	0.01	0.36		
				SHARE	-0.31	1.09	0.01	0.77		
				UKBB (Europeans)	-0.21	0.41	0.008	0.6		
				META (Europeans)	-0.09	0.08	0.01	0.2	*0.85*	
				Trans-Ethnic Meta-Analysis	0.21	0.07	0.01	0.005	*0.0001*	

509
510 **Table 3. Summary statistics of the novel variants after conditioning on index SNPs**

SNP	CHR	POS	EA/NEA	BETA	SE	P value	EAF	Gene
rs570193324	10	114925065	T/C	9.8	2.3	3.2e-05	*0.002*	*TCF7L2*
rs143316471	06	20693278	T/C	-5.3	2.3	0.0054	*0.003*	*CDKAL1*

511



512

Fig 1. Study design.

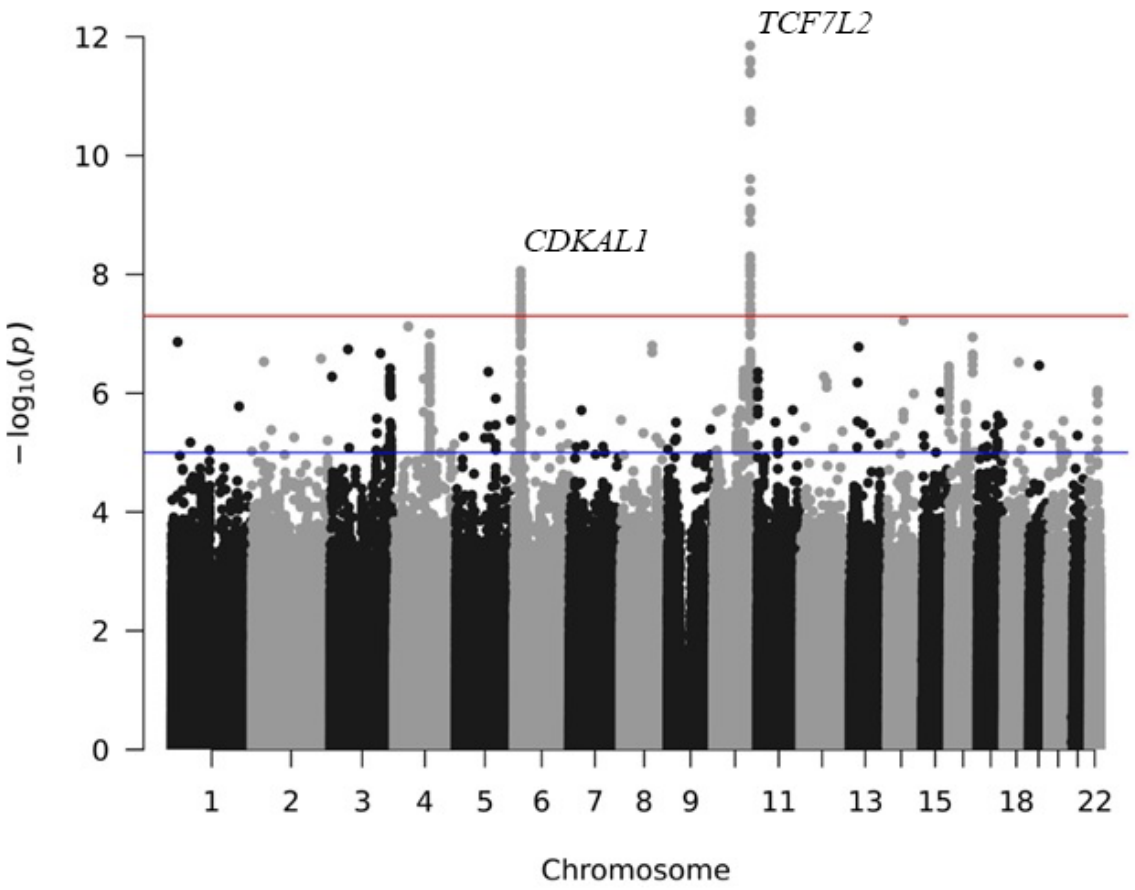


Fig 2. Manhattan plot showing the P-value of association tests for SNPs with Age at diagnosis of T2D in a Trans ethnic meta-analysis of GWAS. Two horizontal lines from the bottom indicate the suggestive ($P<5\times10^{-5}$) and genome-wide significance threshold ($P<5\times10^{-8}$), respectively. The X-axis represents the physical position and the 22 autosomal chromosomes; Y-axis represents the negative logarithm of association p-value. Each dot on the plot represents millions of imputed SNPs across the whole genome.

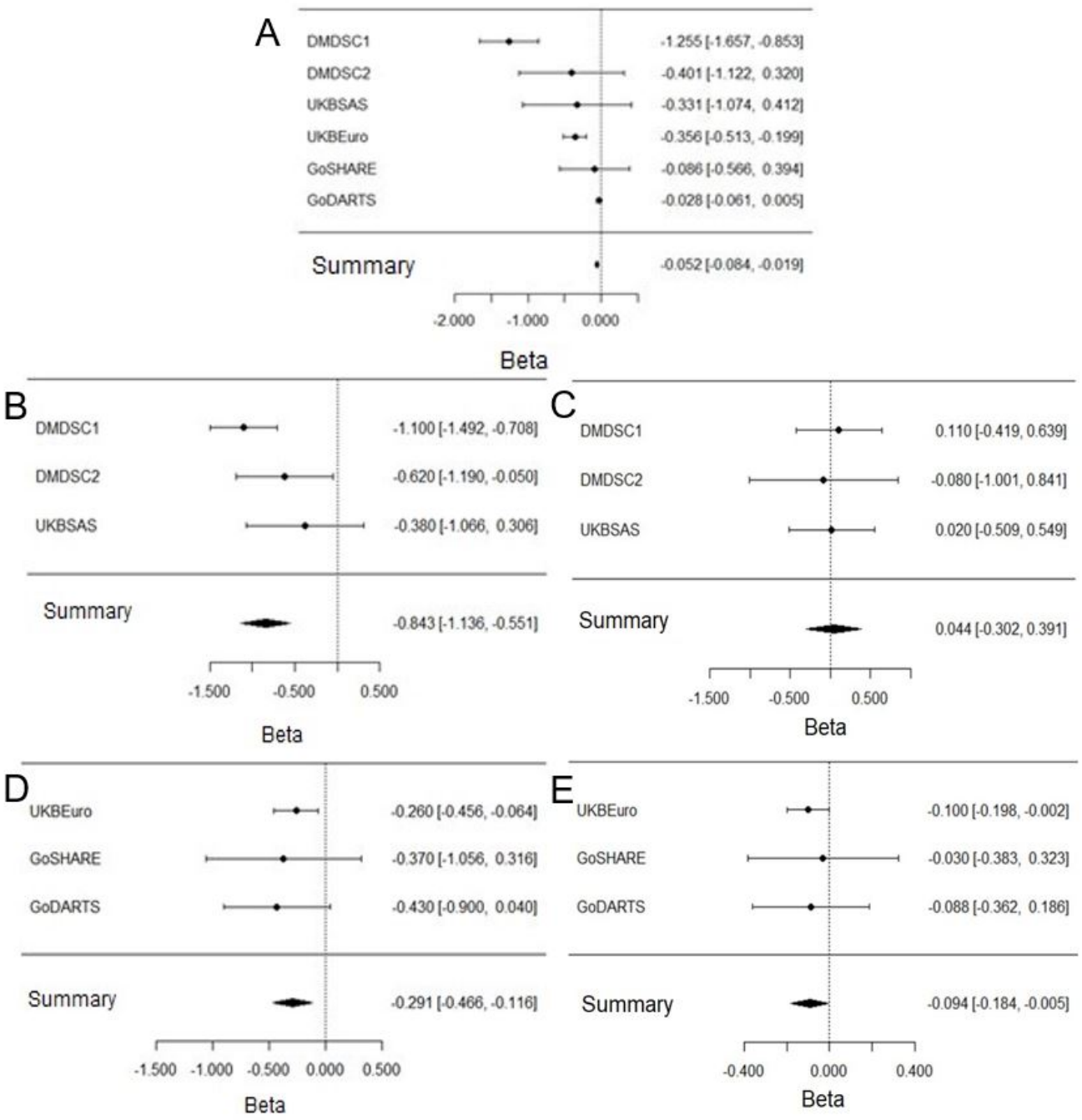


Fig 3. Forest plot for the top significant SNP rs7903146 near *TCF7L2* (Effect allele - T) **in individuals with T2D**. DMDSC Dr. Mohan Diabetes speciality Clinic**,** UKBSAS UK Biobank South Asians. GoDARTS, Genetics of Diabetes Audit and Research in Tayside Scotland; GoSHARE, Genetics of Scottish Health Research Register; UKBB Euro, United Kingdom Biobank Europeans**.** A) Overall meta-analysis of GWAS of Age onset of T2D; B) South Asians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); C) South Asians with later onset of T2D (Age at diagnosis between 20 – 55 years); D) Caucasians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); E) Caucasians with later onset of T2D

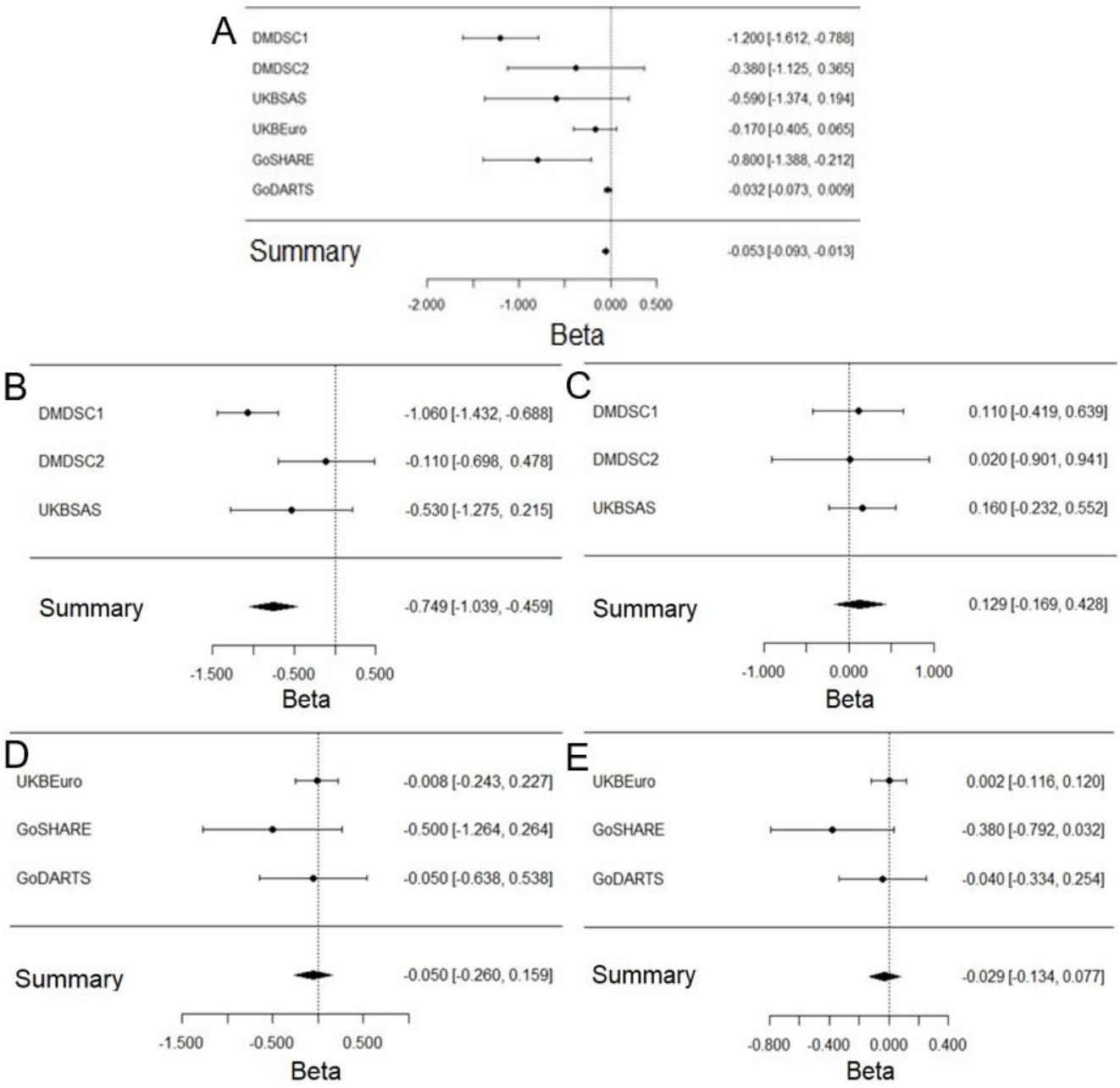


Fig 4. Forest plot for the top significant SNP rs9368219 near *CDKAL1* (Effect allele - T) **in individuals with T2D**. DMDSC Dr. Mohan Diabetes speciality Clinic**,** UKBSAS UK Biobank South Asians. GoDARTS, Genetics of Diabetes Audit and Research in Tayside Scotland; GoSHARE, Genetics of Scottish Health Research Register; UKBB Euro, United Kingdom Biobank Europeans**.** A) Overall meta-analysis of GWAS of Age onset of T2D; B) South Asians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); C) South Asians with later onset of T2D (Age at diagnosis between 20 – 55 years); D) Caucasians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); E) Caucasians with later onset of T2D

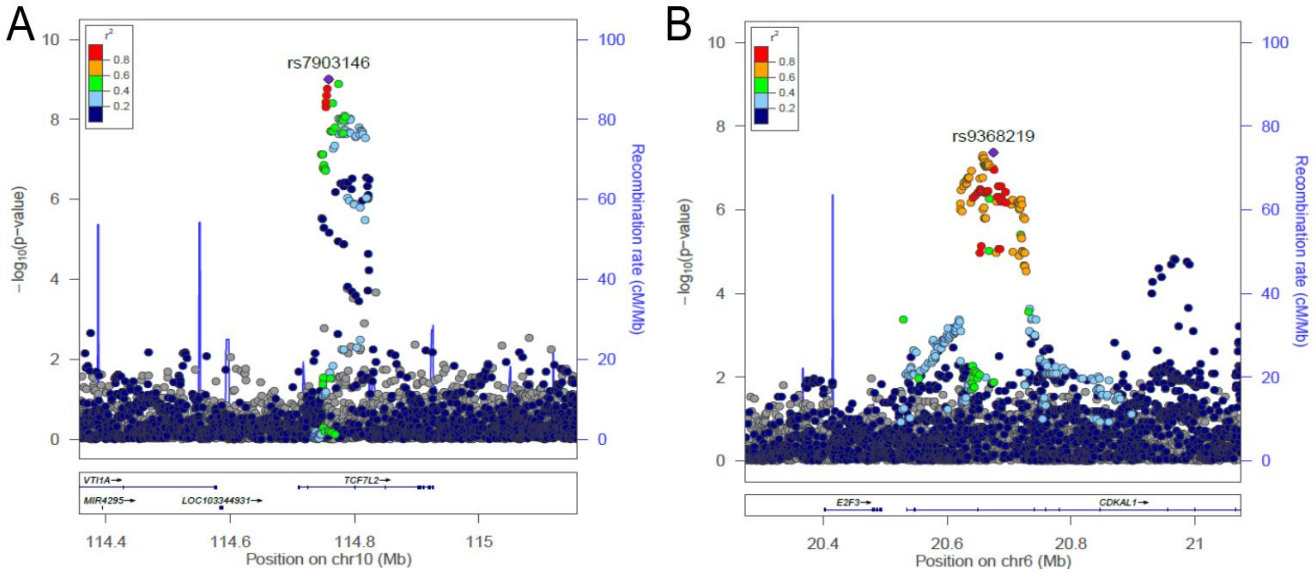


Fig 5. Regional association plots of top significant SNPs. Associations at *TCF7L2* (A) and *CDKAL1* (B) in the South Asian meta–genome-wide association studies. Chr, chromosome; cM/Mb, centimorgan/megabase (genomic location in reference build 37 [Hg19]).

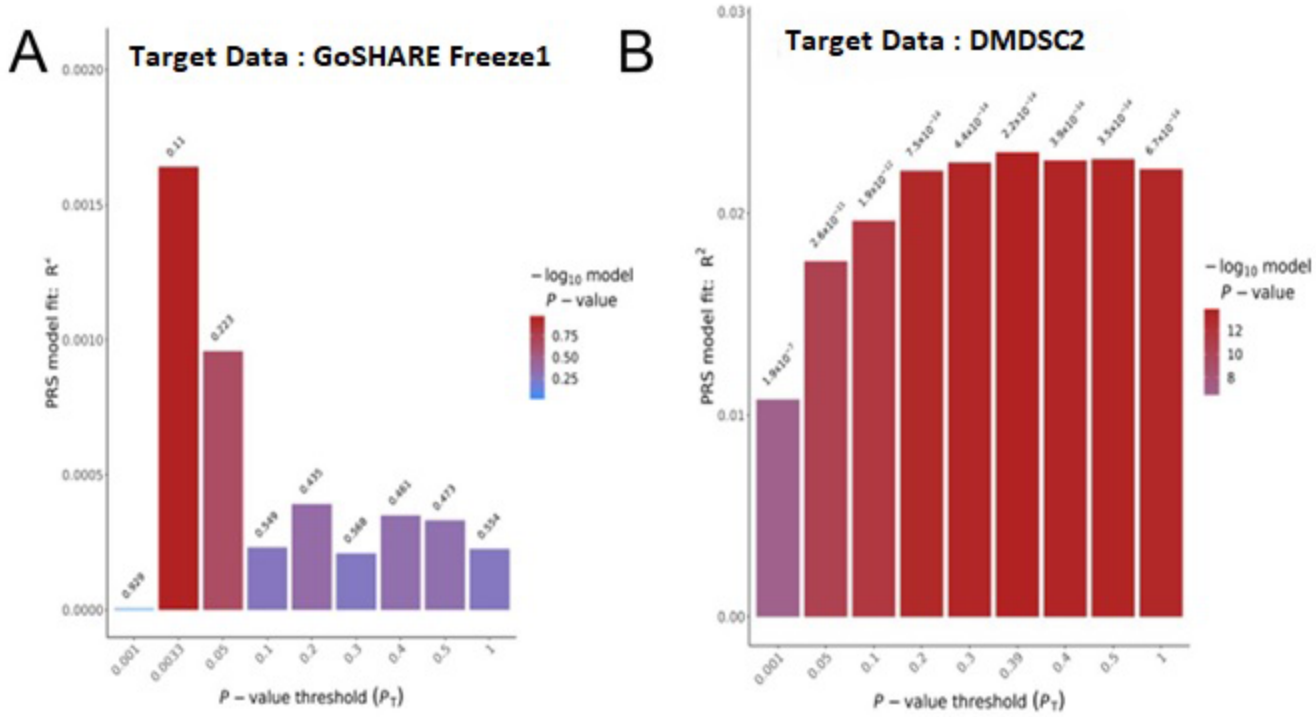


Fig 6. Performance of South Indian PRS in European population. PRS generated using Asian Indian summary data from DMDSC cohort 1 and tested in European samples (GoSHARE) and South Indian independent cohort (DMDSC 2) for polygenic risk prediction of age of onset.

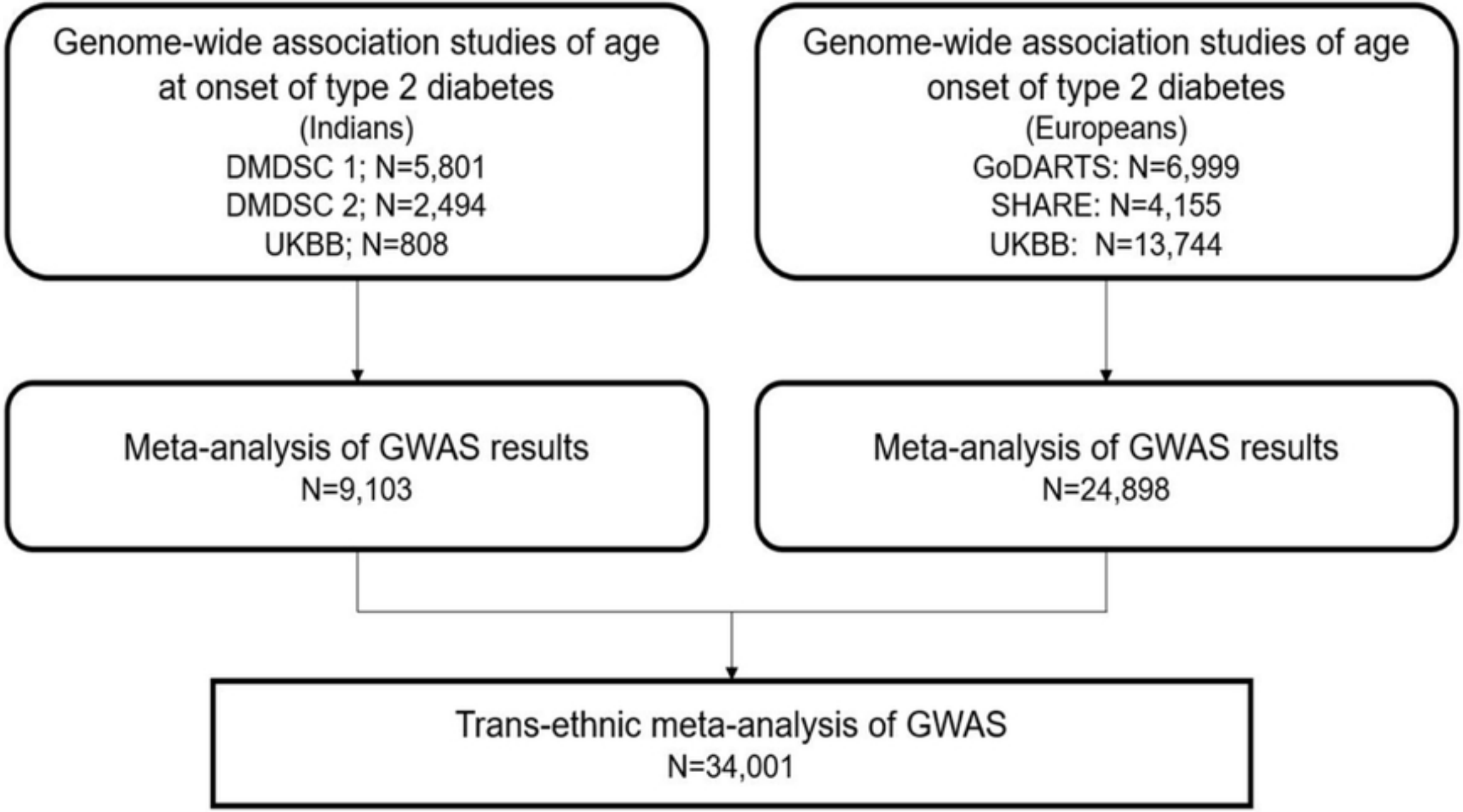


Figure1_StudyDesign

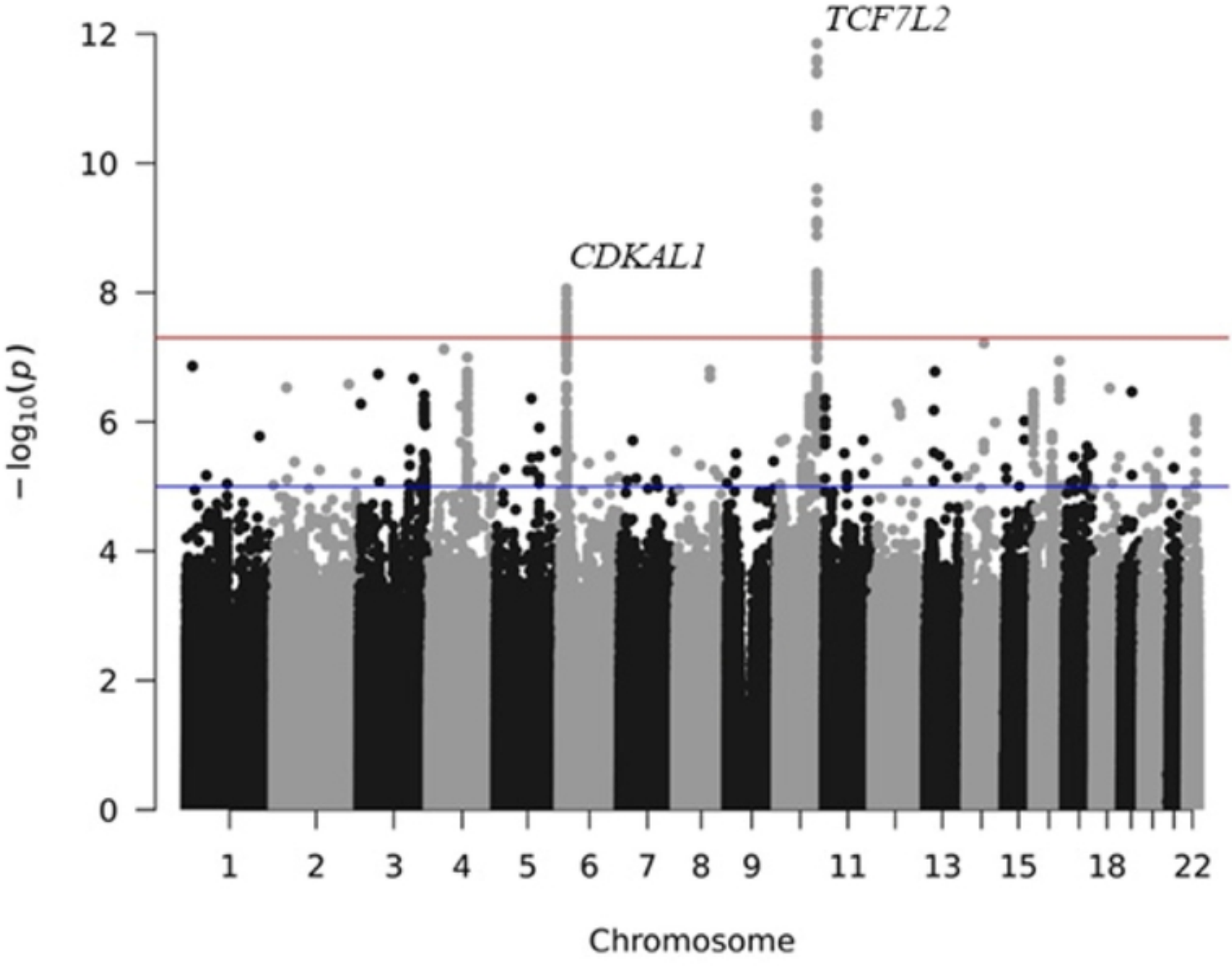


Figure2_Manhattangwas

Figure3_Forest plot

Figure4_Forest plot

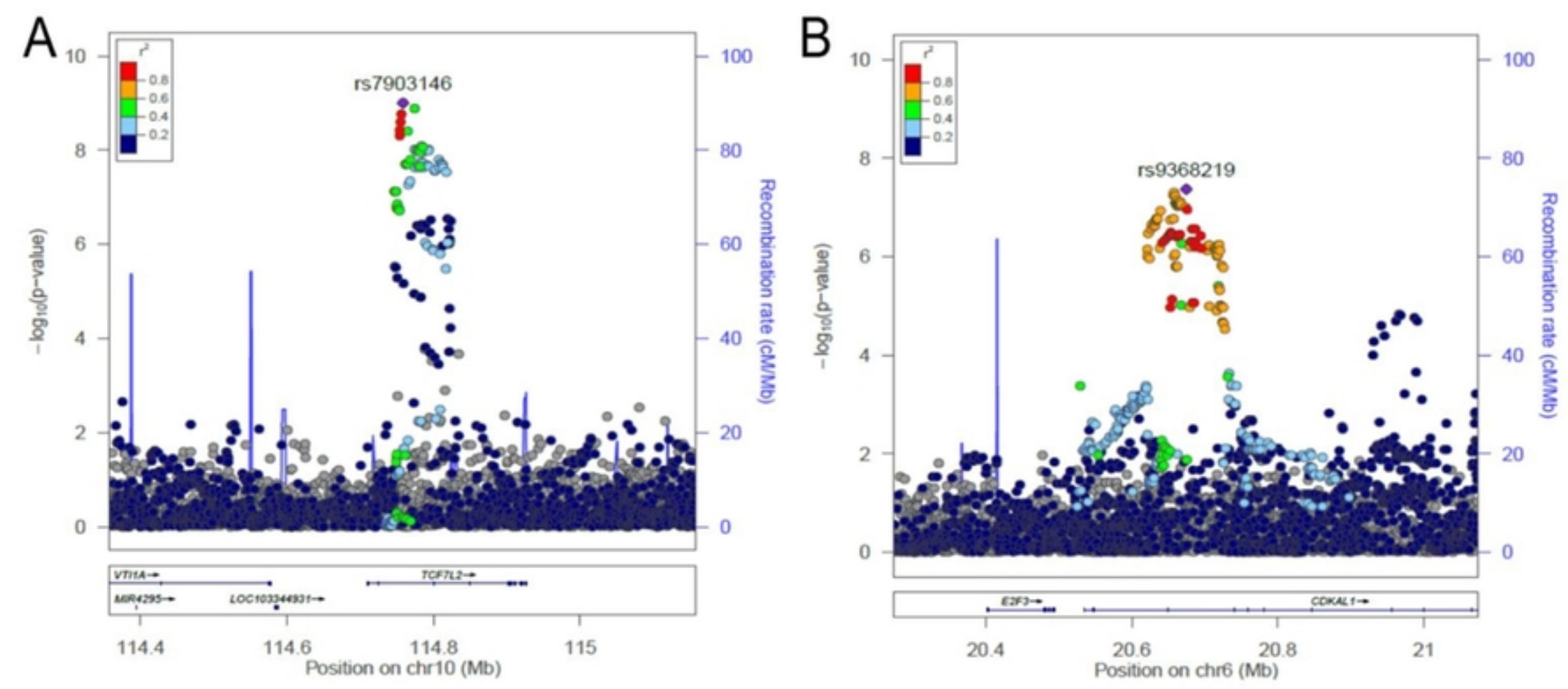


Figure5_Regional

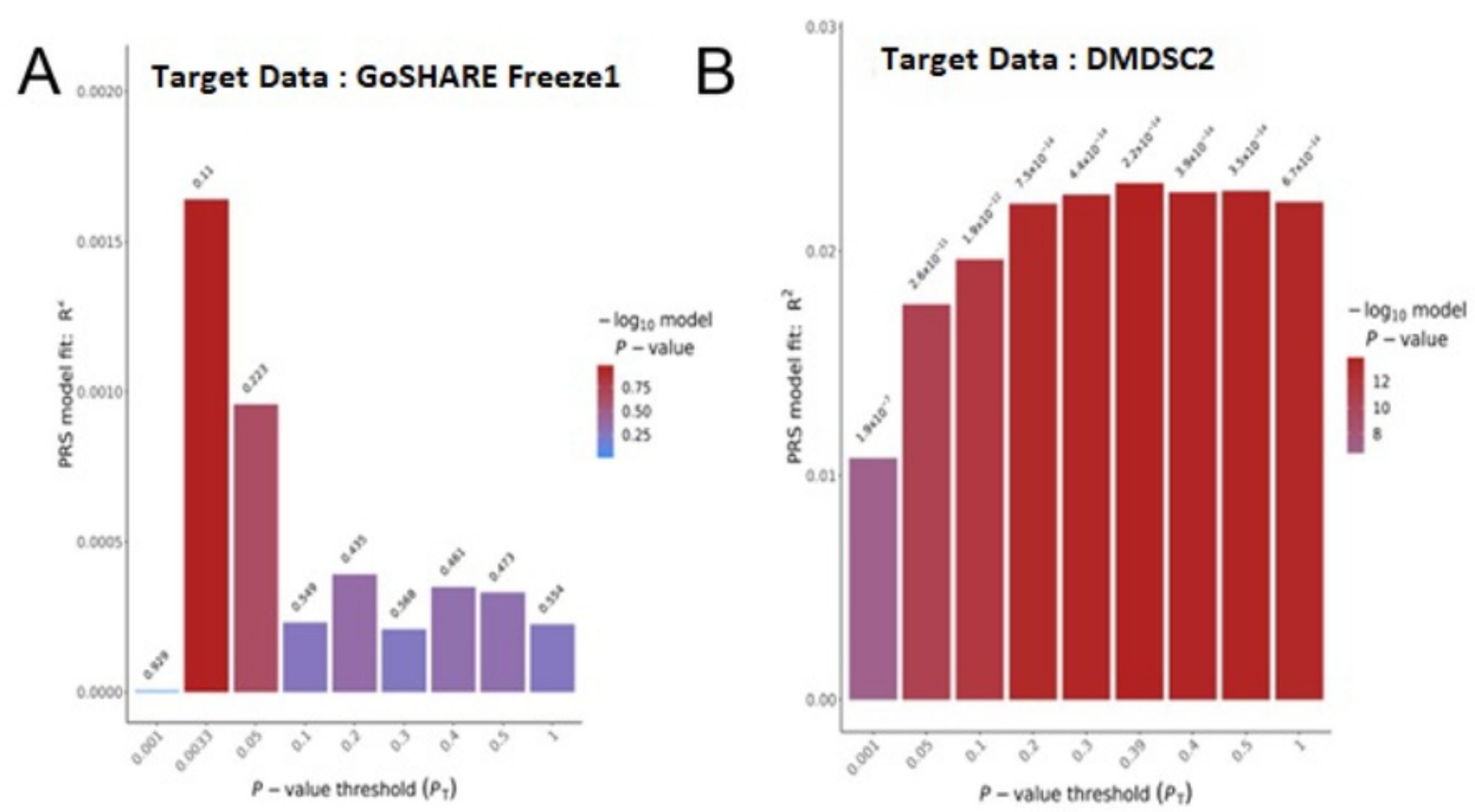

A
Target Data : GoSHARE Freeze1
PRS model fit: R^2
0.0020
0.0015
0.0010
0.0005
0.0000
0.11
0.223
0.829
0.349
0.435
0.368
0.461
0.473
0.324
0.001
0.0013
0.05
0.1
0.2
0.3
0.4
0.5
1
P – value threshold (P_T)
– $\log_{10}$ model P – value
0.75
0.50
0.25
B
Target Data : DMDSC2
PRS model fit: R^2
0.03
0.02
0.01
0.00
1.9×10^{-7}
2.6×10^{-11}
1.9×10^{-12}
7.5×10^{-14}
4.4×10^{-14}
2.2×10^{-14}
3.9×10^{-14}
3.5×10^{-14}
6.7×10^{-14}
0.001
0.05
0.1
0.2
0.3
0.39
0.4
0.5
1
P – value threshold (P_T)
– $\log_{10}$ model P – value
12
10
8

Figure6_PRS