

Supplementary Online Content

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

Sundararajan Srinivasan¹, Samuel Liju², Natarajan Sathish², Moneeza K Siddiqui¹, Ranjit Mohan Anjana², Ewan R. Pearson¹, Alexander S.F. Doney¹, Viswanathan Mohan MD², Venkatesan Radha², 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 +44 1382 383155

Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, United Kingdom.

Supplementary Tables 5

Supplementary Figures 13

Supplementary Table S1. 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	South Indians
DMDSC 2	2,494	41.3±9.77	26.3	58.17;41.8	South 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
GoSHARE	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

Supplementary Table S2. STROBE Statements and Checklist

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Checklist	Recommendation
Title and abstract	1	×	(a) Indicate the study's design with a commonly used term in the title or the abstract
		×	(b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction			
Background/rationale	2	×	Explain the scientific background and rationale for the investigation being reported
Objectives	3	×	State specific objectives, including any prespecified hypotheses
Methods			
Study design	4	×	Present key elements of study design early in the paper
Setting	5	×	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	×	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
			(b) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	×	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	×	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	×	Describe any efforts to address potential sources of bias
Study size	10	×	Explain how the study size was arrived at
Quantitative variables	11	×	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	×	(a) Describe all statistical methods, including those used to control for confounding
		×	(b) Describe any methods used to examine subgroups and interactions
		×	(c) Explain how missing data were addressed
		NA	(d) If applicable, explain how loss to follow-up was addressed
		×	(e) Describe any sensitivity analyses
Results			
Participants	13*	×	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		×	(b) Give reasons for non-participation at each stage
		×	(c) Consider use of a flow diagram
Descriptive data	14*	×	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders

		NA	(b) Indicate number of participants with missing data for each variable of interest
		NA	(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	NA	Report numbers of outcome events or summary measures over time
Main results	16	×	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		×	(b) Report category boundaries when continuous variables were categorized
		×	(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	×	Report other analyses done—eg, analyses of subgroups and interactions, and sensitivity analyses
Discussion			
Key results	18	×	Summarise key results with reference to study objectives
Limitations	19	×	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	×	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	×	Discuss the generalisability (external validity) of the study results
Other information Funding	22	×	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

Supplementary Table S3. Summary statistics of the novel variants after conditioning on index SNPs

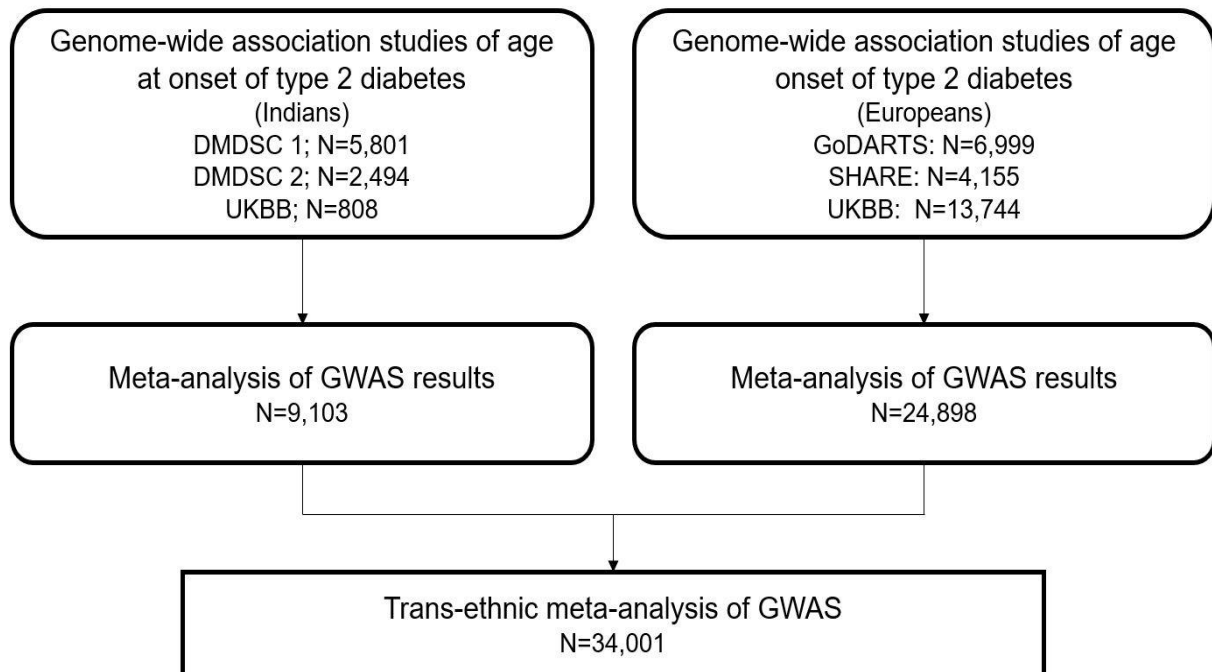
SNP	CHR	POS	EA/NEA	BETA	SE	P value	EAf	Gene
rs570193324	10	114925065	T/C	9.8	2.3	3.2×10^{-05}	0.002	TCF7L2
rs143316471	06	20693278	T/C	-5.3	2.3	0.0054	0.003	CDKALI

Supplementary Table S4. Summary statistics of the novel variants in DMDSC cohort after adjusting for BMI

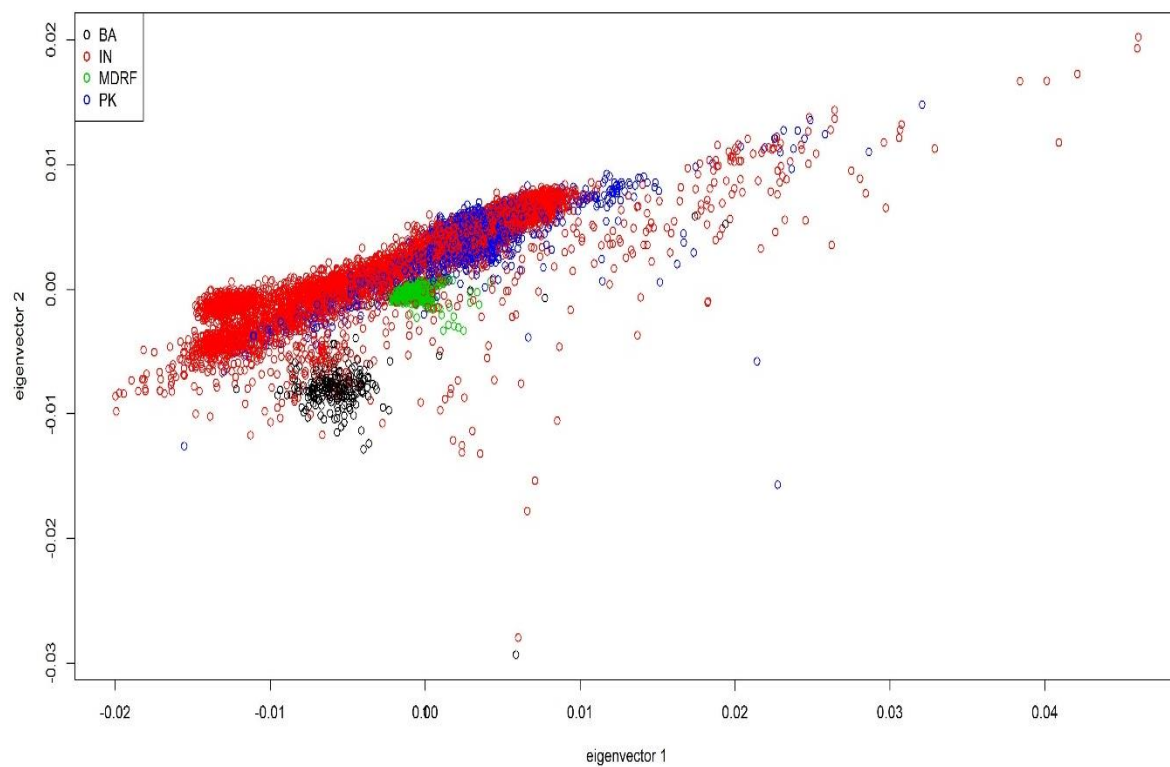
SNP	Study Cohort	Sample Size N	BETA (unadjusted)	BETA (adjusted BMI)	SE	EAf	P value (unadjusted)	P value (Adjusted BMI)
rs7903146 (TCF7L2)	DMDSC data freeze 1	5801	-1.26	-1.26	0.20	0.35	1×10^{-10}	1.0×10^{-09}
rs9368219 (CDKALI)	DMDSC data freeze 1	5801	-1.20	-1.18	0.21	0.25	4.3×10^{-08}	4.0×10^{-08}

Supplementary Table S5. Proportion of individuals in DMSC under different age groups used in this study

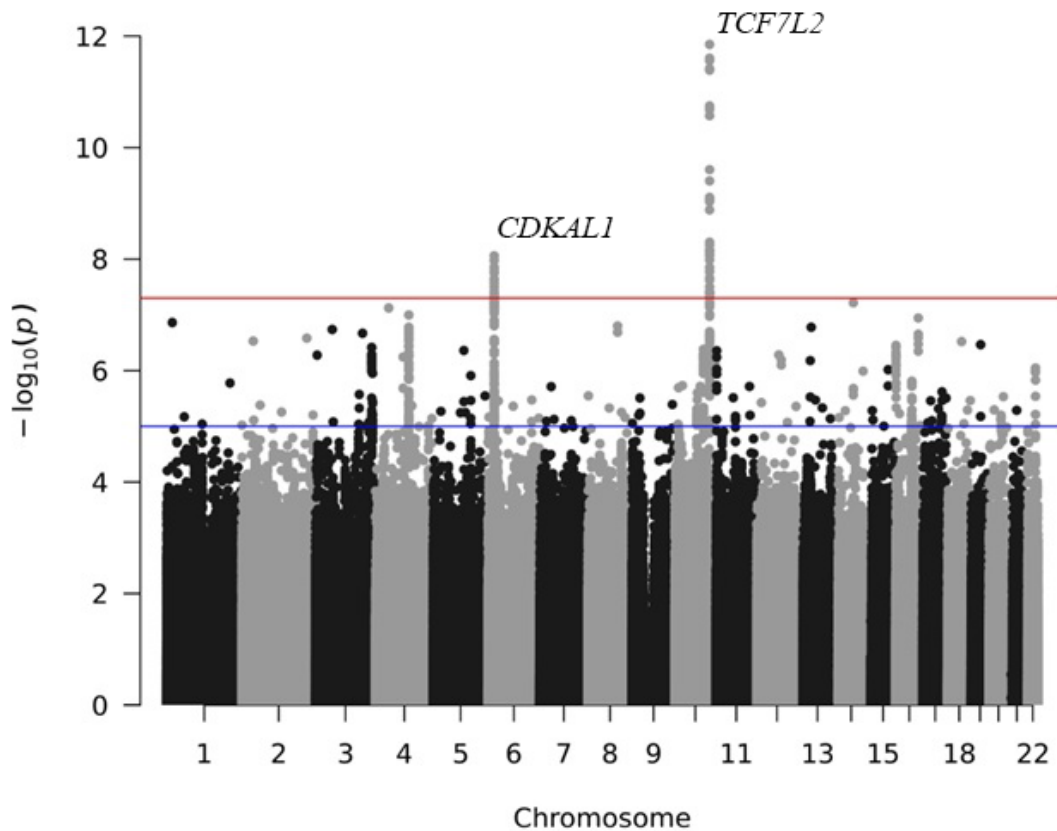
Age group T2D (Years)	DMDSC 1	DMDSC 2
< 20	0	0
>20 & < 30	912 (15.8%)	460 (18.5%)
> 30 & < 40	1582 (27.2%)	657 (26.3%)
> 40 & < 55	2697 (46.5%)	985 (39.5%)
> 55	610 (10.5%)	392 (15.7%)



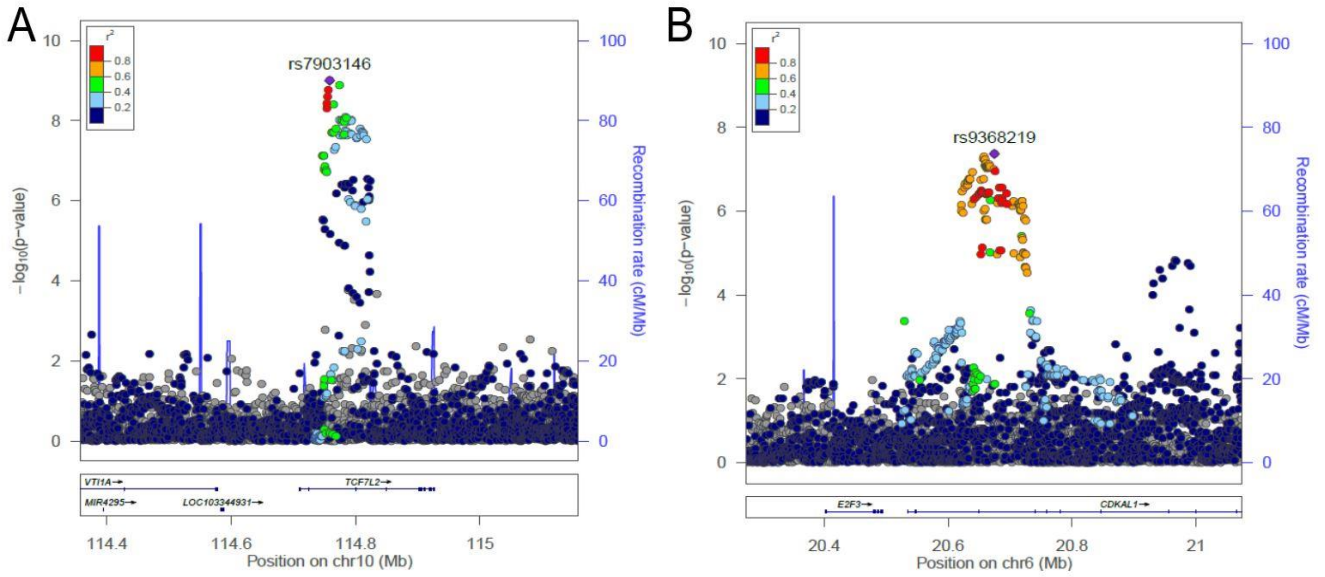
Supplementary Figure S1. Study Design.



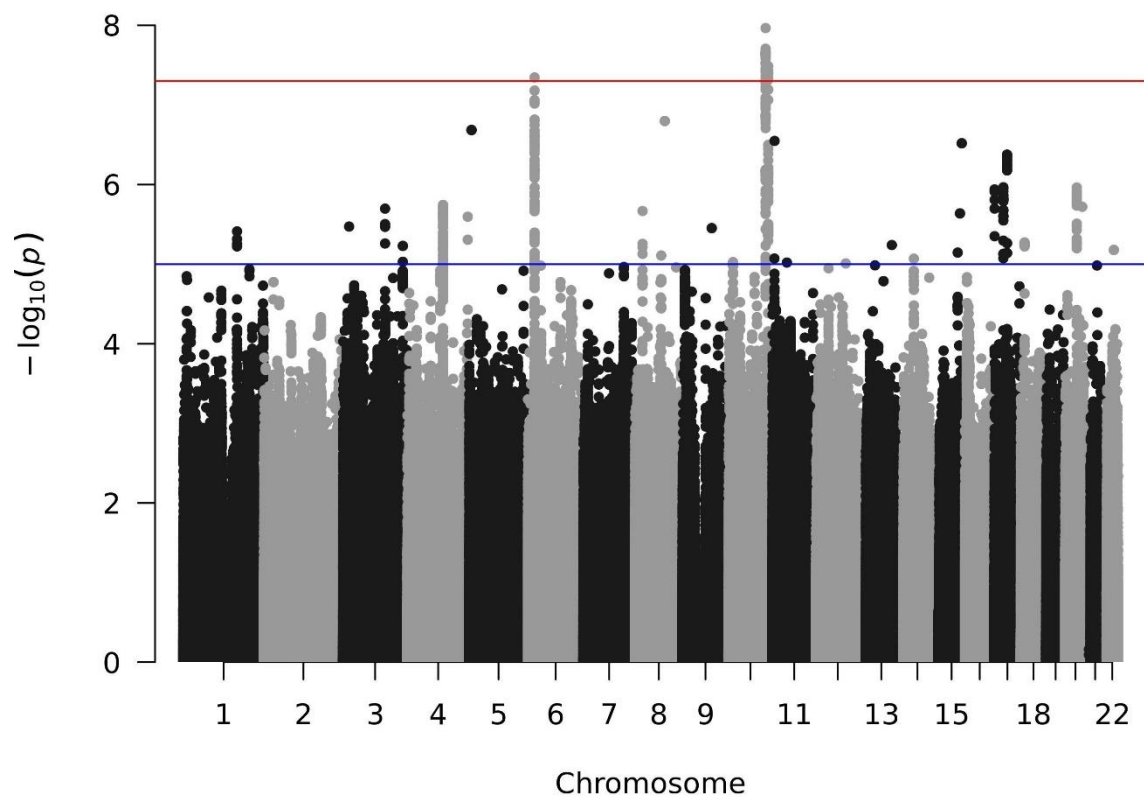
Supplementary Figure S 2. PCA with the UKKB South Asians and South Indians from DMSDC (Formerly MDRF). BA Bangladeshi; IN Indians & PK Pakistani in UK biobank.



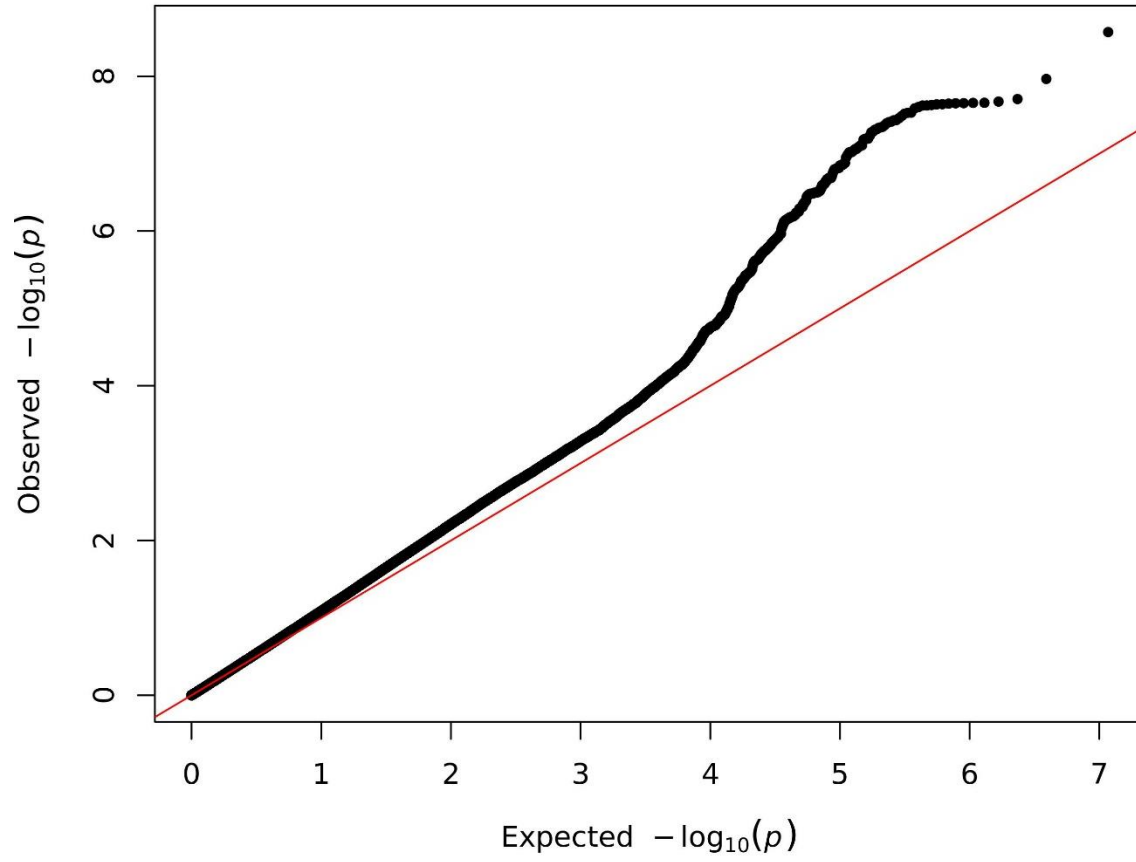
Supplementary Figure S 3. Manhattan plot showing the P-value of association tests for SNPs with Age at diagnosis of T2D in a Trans ethnic meta-analysis of GWAS. Two horizontal lines from the bottom indicate the suggestive ($P < 5 \times 10^{-5}$) and genome-wide significance threshold ($P < 5 \times 10^{-8}$), 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.



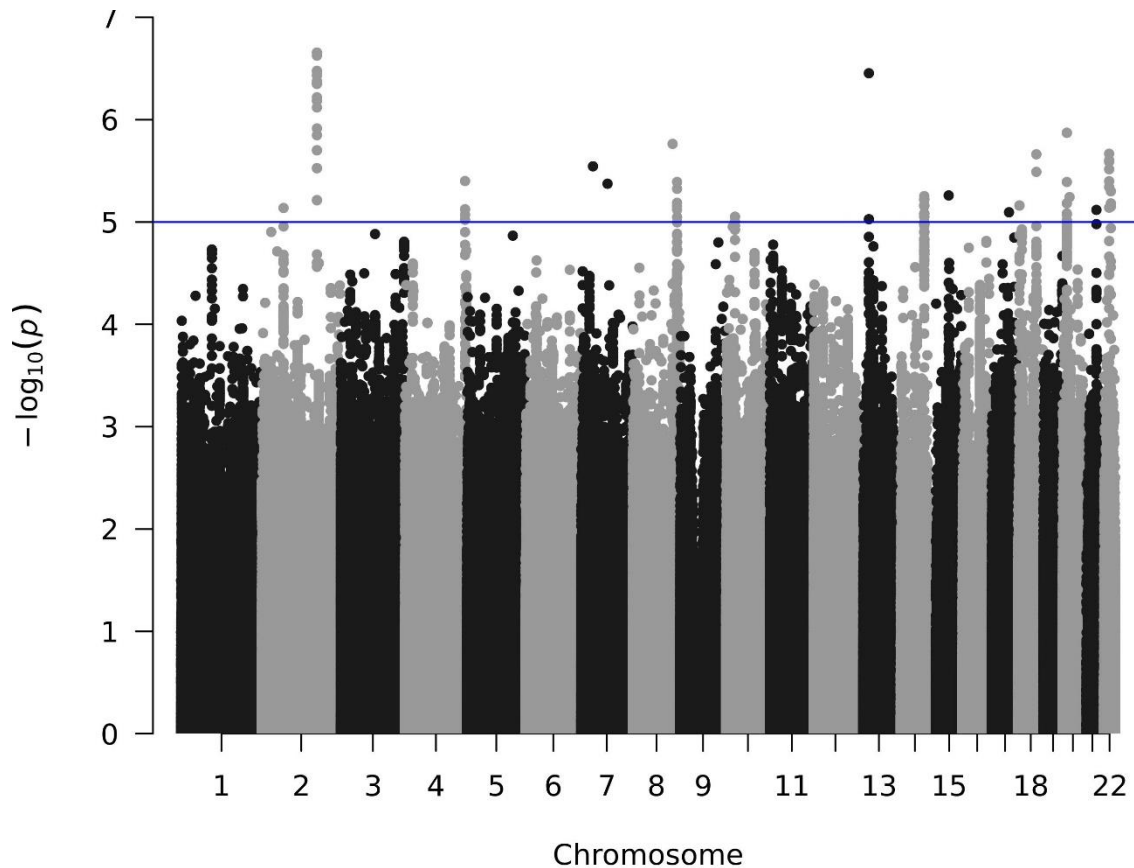
Supplementary Figure S 4. Regional association plots of top significant SNPs. Associations at TCF7L2 (A) and CDKAL1 (B) in the South Asian meta-genome-wide association studies. Chr, chromosome; cM/Mb, centimorgan/megabase (genomic location in reference build 37 [Hg19]).



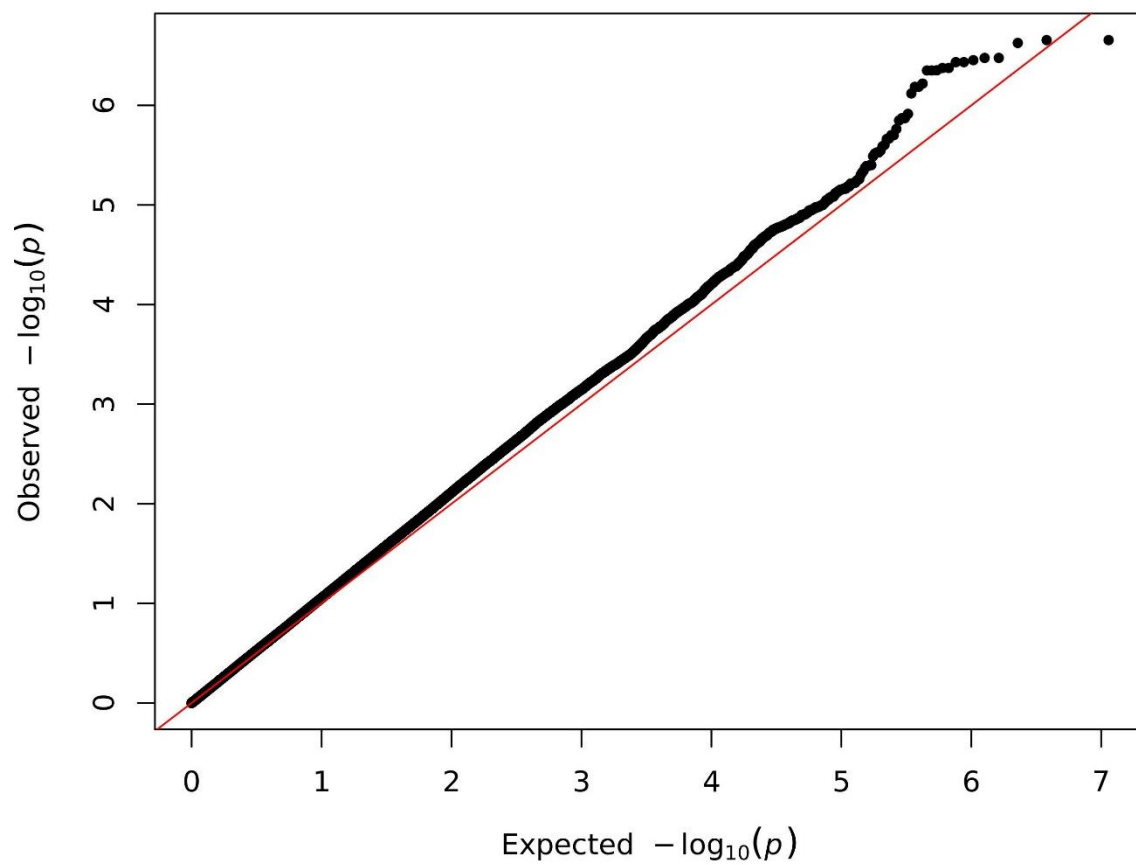
Supplementary Figure S 5. Manhattan plot of genome-wide meta-analysis of South Asians. It is showing the P-value of association tests for SNPs with possible Age at Onset of T2D in overall Trans ethnic meta-analysis of GWAS. Two horizontal lines from the bottom indicate the suggestive ($P < 5 \times 10^{-5}$) and genome-wide significance threshold ($P < 5 \times 10^{-8}$), 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. Top hits mapped to nearby genes are labelled.



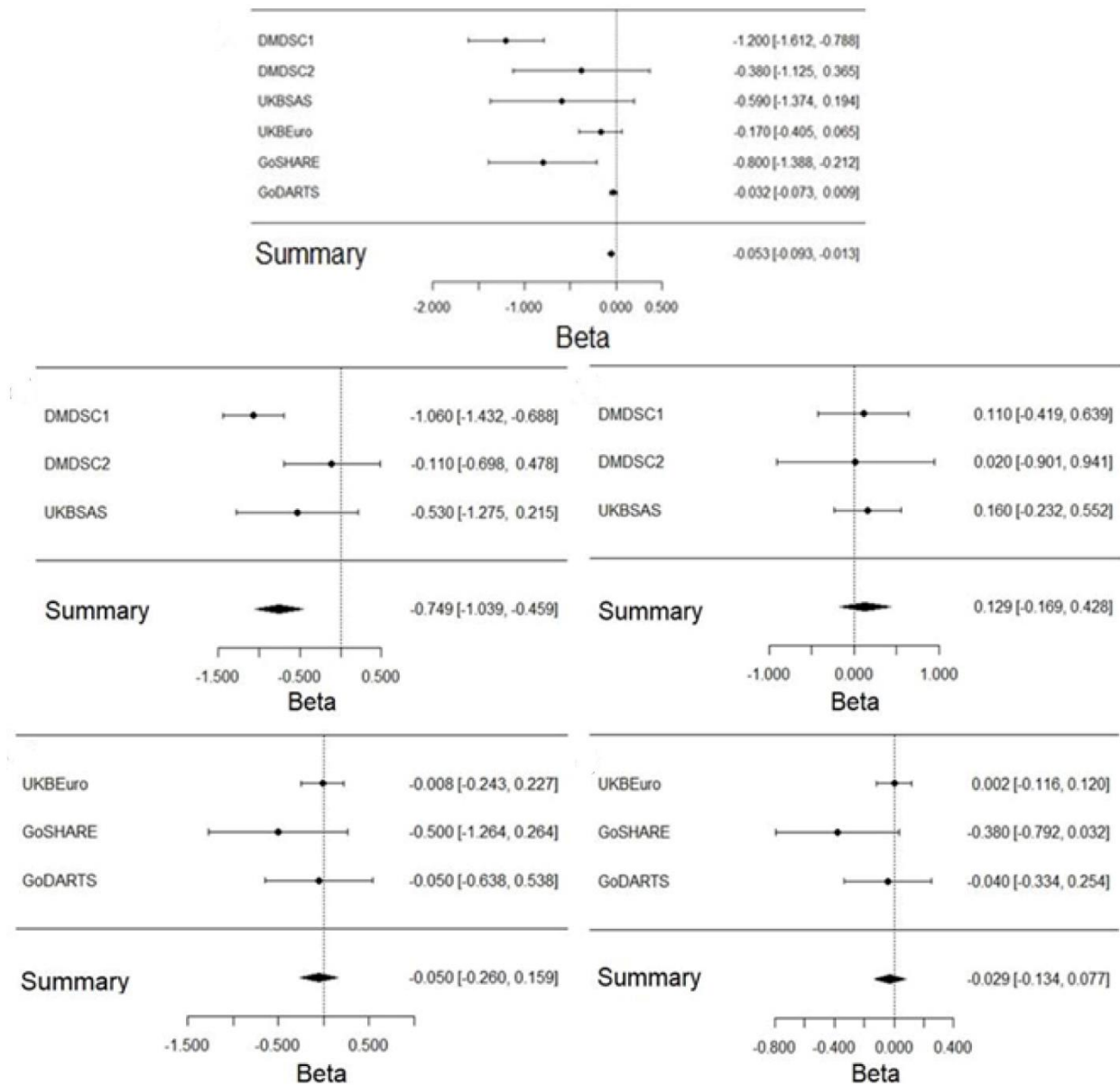
Supplementary Figure S 6. Quantile-Quantile plot of genome-wide trans-ethnic meta-analysis of South Asians.



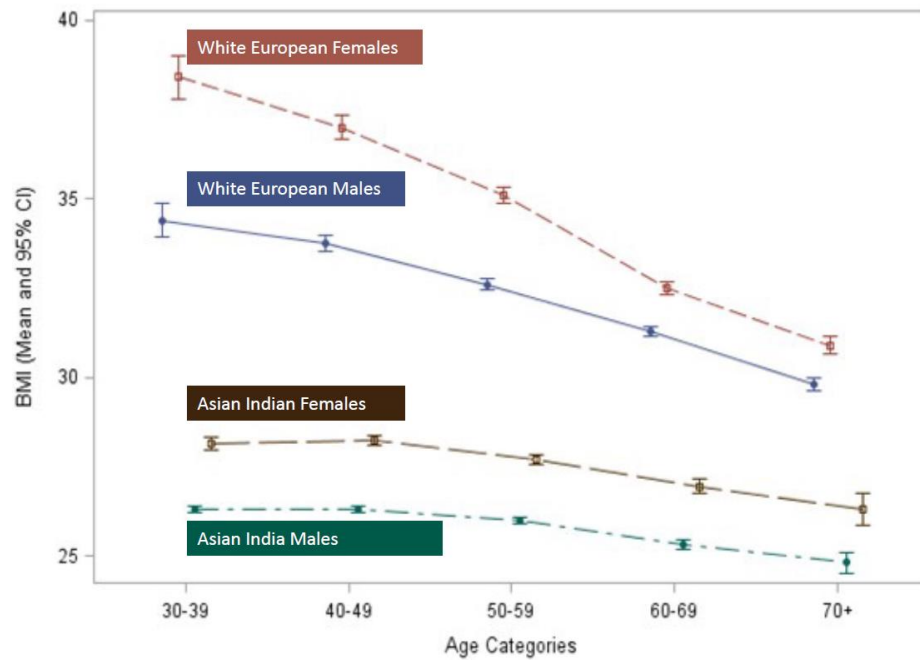
Supplementary Figure S 7. Manhattan plot of genome-wide trans meta-analysis of Europeans. It is showing the P-value of association tests for SNPs with possible age at onset of T2D in overall trans ethnic meta-analysis of GWAS. The horizontal lines from the bottom indicate the suggestive ($P < 5 \times 10^{-5}$). 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. Top hits mapped to nearby genes are labelled.



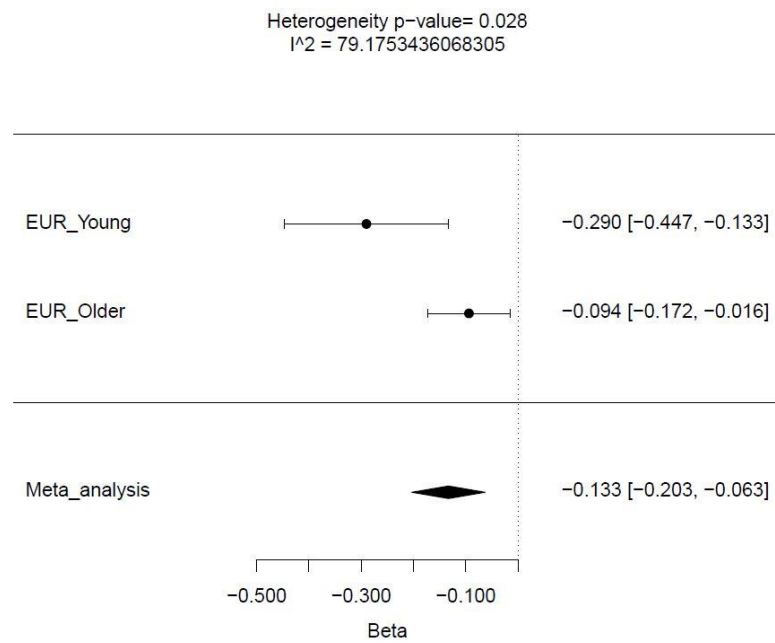
Supplementary Figure S 8. Quantile-Quantile plot of genome-wide trans-ethnic meta-analysis of Europeans



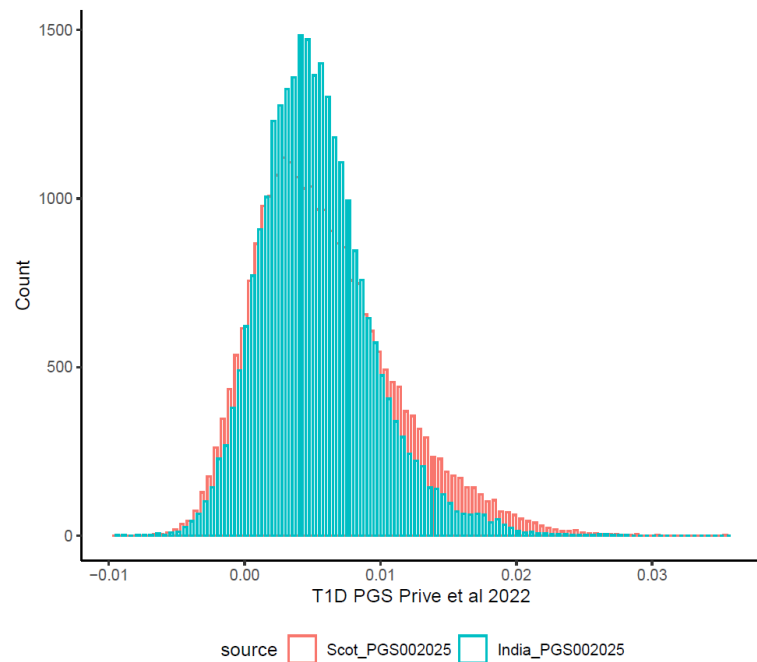
Supplementary Figure S 9. Forest plot for the top significant SNP *rs9368219* near *CDKAL1* (Effect allele - T) in DMDSC Dr. Mohan Diabetes speciality Clinic, UKBSAS UK Biobank South Asians. GoDARTS, Genetics of Diabetes Audit and Research in Tayside Scotland; GoSHARE, Genetics of Scottish Health Research Register; UKBB Euro, United Kingdom Biobank Europeans. A) Overall meta-analysis of GWAS of Age onset of T2D; B) South Asian Indians with earlier onset of T2D (Age at diagnosis between 20 – 55 years); C) South Asian Indians with later onset of T2D (Age at diagnosis over 55 years); D) Europeans with earlier onset of T2D (Age at diagnosis between 20 – 55 years); E) Europeans with later onset of T2D (Age at diagnosis over 55 years)



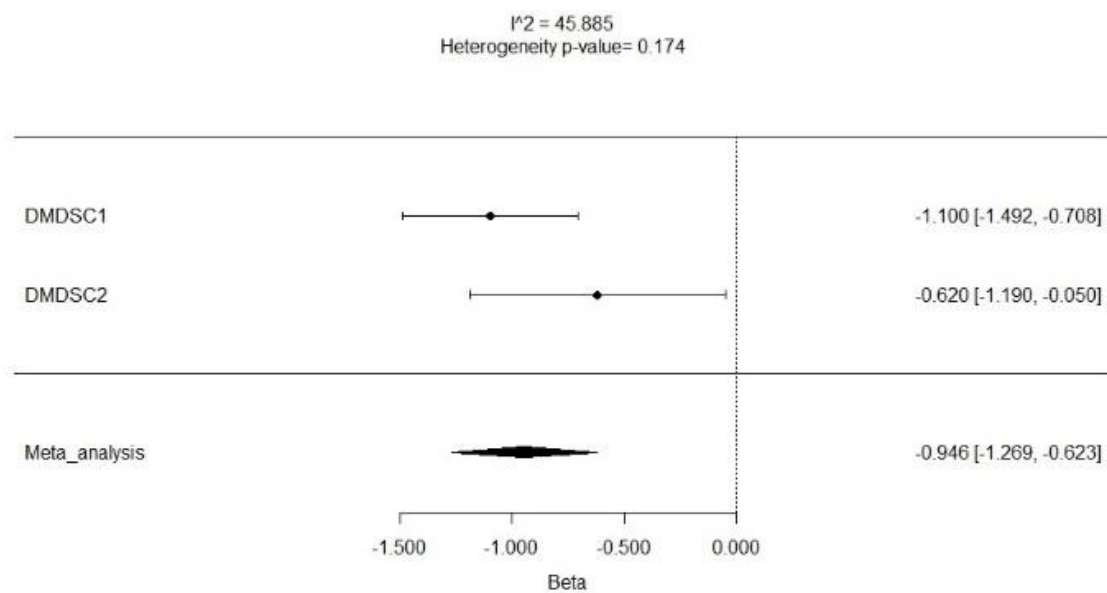
Supplementary Figure S 10. Females in squares and males in solid circles. White European males in blue solid circles with solid lines, females in red squares with dashed lines, South Indian males in green circles with dot dashed lines and females in brown squares with long dashed lines.



Supplementary Figure S 11. Forest plot for the SNP rs7903146 near TCF7L2 (Effect allele - T) Overall meta-analysis of Caucasians with earlier onset of T2D (Age at diagnosis between 20 – 55 years and later onset of T2D (Age at diagnosis over 55 years).



Supplementary Figure S 12. T1D Polygenic Score derived from PGS catalogue *Scottish in red, DMDSC in blue*



Supplementary Figure S 13. South Asian specific Meta Analysis of TCF7L2 without UKBB SAS