GMR
The Original

GENOMIC VARIANTS ASSOCIATED WITH EARLY-ONSET TYPE 2 DIABETES IN SOUTH ASIAN POPULATIONS

Dr. Hariballav Mahapatra[1*], Shikha[2], Nagendra Babu Rajaboina[3], Dr.Niraj Lodha[4]

[1]Diabetologist, Department of Diabetology, Sevayan Diabetes Centre, Puri, Shree Hospitals, Bhubaneswar, India 751014
ORCID ID: https://orcid.org/0000-0001-9845-0509 Email I'd: nullandzero@gmail.com
[2]PhD Department of Microbiology, Maharshi Dayanand University, 124001 ORCID ID: https://orcid.org/0009-0008-3850-2678
Email ID: narwalshikha70@gmail.com
[3]Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh
Email: rnagendrababu@kluniversity.in
[4]Consultant Physician, Department of Internal medicine and Diabetes, Lodha Hospital and Research Centre, Pali, Rajasthan
Email ID: nirajlodh2001@gmail.com ORCID ID: 0009-0009-9134-176X

ABSTRACT

Type 2 diabetes mellitus (T2D) is more prevalent and of higher burden earlier in South Asian populations, but the underlying genetic mechanisms of early susceptibility to the condition are not fully understood. Open genomic databases offer the possibility to find variants that may have an impact on the manifestation of the disease in the past. The present study aimed to identify genomic variants associated with T2D in South Asians and prioritize those with potential early-onset relevance using a composite scoring framework. A secondary analysis was conducted using variants obtained from the Type 2 Diabetes Knowledge Portal based on a large-scale association dataset. Following initial filtering within the portal, the top 500 variants were analyzed. Variants were evaluated using statistical significance, effect size, functional annotation, and gene mapping. A composite score integrating −log10(p-value), absolute effect size, and functional impact was used to classify variants, with the upper quartile designated as early-onset-associated. Of the 500 variants analyzed, 293 were genome-wide significant, and 125 were classified as early-onset-associated. Strong association signals were concentrated at the TCF7L2 locus, with additional contributions from *CDKAL1*, *IGF2BP2*, *HHEX*, *ADCY5*, *PPARG*, *UBE2E2*, and *SLC30A8*. Most variants were non-coding, indicating a predominantly regulatory architecture, with selected coding variants suggesting functional relevance. These findings highlight key susceptibility loci in South Asians and demonstrate that composite prioritization can identify variants with potential relevance to early-onset T2D, providing targets for future validation.

KEYWORDS: type 2 diabetes, South Asian population, early-onset diabetes, genomic variants, TCF7L2

1. INTRODUCTION

Type 2 diabetes mellitus (T2D) is a chronic metabolic condition that is linked to incessant hyperglycemia due to the combination of insulin resistance and insulin-secreting dysfunction. It is one of the most important health challenges in the world, and its rates are growing rapidly both in developed and developing countries (Ahmad et al., 2022). The morbidity and mortality rates are linked to the disease, partly because of the long-term complications, which are cardiovascular disease, neuropathy, nephropathy, and retinopathy. At the physiological level, T2D is the result of the complicated impacts of the pancreatic beta-cell dysfunction, the disturbed glucose metabolism, and the dysregulation of systemic metabolism (Galicia-Garcia et al., 2020). T2D has increased significantly in the past decades globally because of demographic shifts, urbanization, and lifestyle changes. According to epidemiological studies, there is a sharp rise in incidence and prevalence, especially in the countries of low and middle income, where the healthcare systems tend to be severely limited (Chen et al., 2025). Diseases are formed by both environmental and genetic factors, the role of which in this case is played by lifestyle determinants: diet, physical inactivity, and obesity (Dendup et al., 2018). Nevertheless, these environmental factors cannot explain inter-individual differences in the susceptibility to the disease, which means that there is a significant genetic factor (Tremblay & Hamet, 2019). T2D has a complex genetic structure, and a number of loci determine various biological processes. Genome-wide association studies (GWAS) have reported several variants that are risk factors and linked to T2D, with many located outside of the coding sequences of proteins, implying a role in regulation and not direct and exact protein-coding actions (Kreienkamp et al., 2023). These variants can also affect the expression of genes, secretion of insulin and glucose homeostasis, which has led to heterogeneity of the disease. Moreover, recent studies have also shown the importance of genetic and epigenetic processes in the connection of T2D with cardiovascular complications, which emphasizes the systemic effects of genetic vulnerability (De Rosa et al., 2018). More recent progress also highlights that genetic drivers play a role in heterogeneity in the pathophysiology of a disease, meaning that T2D is not a disease per se but a continuum of similar related metabolic states (Suzuki et al., 2024). South Asians, among other groups, have a significantly disproportionate risk of developing T2D at relatively younger ages and lower body mass indices than in other ethnic groups. It is explained by the fact that, in addition to genetic predisposition, such factors as dietary habits and lack of physical activity contribute to this increased vulnerability (Jenum et al., 2019). The premature development of T2D among the population is particularly worrying because it is linked to a prolonged

period of exposure to the disease and the probability of complications. It has been shown by clinical studies that early-onset T2D patients have a higher risk of developing microvascular complications and have a more aggressive disease course (Huang et al., 2019). T2D has now become a unique clinical and epidemiological problem, with diagnosis at an earlier age, and in many instances, a more significant genetic influence. Recent sources demonstrate that cases, which have an onset in earlier stages, are more likely to have more severe metabolic disturbances and worse long-term outcomes than later-onset disease (Strati et al., 2024). Furthermore, there is an increasing amount of evidence that early-onset T2D can be characterized by unique biological processes and genetic determinants, which once again justify the necessity of specific study (Misra et al., 2023). In spite of these revelations, the majority of large-scale genetic studies have not specifically addressed early-onset disease or did not provide specific phenotypic information (age at diagnosis).

In spite of the significant advancements in the study of genetic variants in relation to T2D, there are still a number of gaps. Most of the studies are very skewed in terms of their population of Europeans, and therefore, findings cannot be generalized to other ethnicities. Specifically, the research on South Asian populations is relatively low, though they have an enormous disease burden. Besides, the lack of age-specific data in most publicly available datasets limits the direct examination of early-onset T2D. This drawback renders the need to adopt alternative analytical methods that could be used to entail possible relevance to early-onset disease based on variant features, including effect size, statistical significance, and functional annotation. The current research intends to explore the genomic variations related to T2D among South Asians based on the publicly available genetic association information. The composite prioritization framework is used to determine variants that have the possibility of being relevant to early-onset disease using several dimensions of genetic evidence. This research aims to improve the knowledge about the genetic structure of T2D and emphasize the variants that can lead to the earlier occurrence of the disease by concentrating on population-specific data and using a systematic approach to the analysis.

2. METHODOLOGY

2.1 Research Design

The research was based on a secondary data analysis design based on the publicly available data on genomic association. This was a cross-sectional and analytical method as it aimed at identifying and ranking genetic variants that are related to T2D among South Asian populations. A quantitative design was used to assess the associations at the level of variants and to obtain a composite score of prioritization that indicates possible applicability to early-onset disease. Statistical computing tools were used in the analysis, which provided a level of reproducibility and consistency in all the steps.

2.2 Data Source

The T2D data in this paper were obtained as a result of a large-scale genetic association study of T2D in South Asian and European populations (Loh et al., 2022). The data were retrieved in the Type 2 Diabetes Knowledge Portal, which is a free-of-charge platform that collects and packages genetic information about diabetes and metabolic characteristics (Costanzo et al., 2023). The preliminary screening and rudimentary analyses were carried out in the portal to find pertinent variants related to T2D among South Asian populations. On this, the 500 variants with the highest association significance were picked and downloaded to be analyzed further. Only the versions that matched the South Asian lineage and T2D phenotype were utilized in the current study.

2.3 Data Preparation and Cleaning

The data was loaded into the analysis environment and analyzed in terms of completeness and consistency. Redundant records were eliminated, and variables were converted into the right data type. Important variables such as effect size (beta), p-value, allele frequency and functional annotation were maintained. Field variants that had missing or invalid values in key fields were not included to improve the reliability of the analysis. Formatting characters in the gene annotations were filtered out so that the groups and interpretation could be properly done.

2.4 Inclusion Criteria

The variants were included in the analysis provided they were annotated with South Asian ancestry, related to the T2D phenotype, and included valid statistical values, such as beta (β) and p-value. Also, variants that had functional annotation information and gene mapping were kept. Additional screening on the basis of allele frequency or genomic position was not done, and all the eligible variants in the dataset were evaluated exhaustively.

2.5 Statistical Analysis

The evaluation of the association signals was done using reported p-values and the evaluation of the effect sizes using R. Variants were classified in genome-wide significance ($p < 5\times10^{-8}$) and suggestive ($p < 1\times10^{-5}$) groups. The descriptive statistics were calculated to describe the distribution of the variants according to these categories. The effect sizes were evaluated in terms of absolute beta (β) values to understand the strength of association without taking into consideration the direction.

2.6 Functional Annotation and Gene Mapping

Variants were categorized based on the functional effects, such as intronic, intergenic, regulatory, and coding. Mapping was done at the level of genes using the nearest annotation of the gene in the dataset. This allowed variants to be aggregated on the locus level and the detection of genes with a dense set of association signals.

2.7 Early-Onset Prioritization Framework

A composite scoring system was created to rank variants that had the potential to be relevant to the early-onset T2D. The score was a combination of statistical significance, expressed in the form of -log 10 (p-value), effect size expressed as the absolute beta value, and functional annotation expressed as a categorical weight score representing the predicted biological impact of each variant.

All the items were scaled down and summed up to produce a total prioritization score. The variations in the upper quartile of the score distribution were further divided into early-onset-associated and the rest were general T2D variants. This allowed the systematic detection of more impactful variants without using explicit age-based stratification.

3. RESULTS

3.1 Overview of Variant Dataset and Filtering

Five hundred T2DM-related genomic variants among South Asian individuals were studied. Data cleaning did not involve any form of elimination of variants, implying a complete and well-organized dataset. Out of these, 293 variants were found to pass the genome-wide significance threshold ($p < 5 \times 10^{-8}$), and all the 500 variants were found to pass the suggestive significance threshold ($p < 1 \times 10^{-5}$). The composite prioritization framework was used to rank 125 variants (25%), which were classified as early-onset-associated, and the remaining 375 variants were classified as general T2D variants (Table 1).

Table 1. Summary of variant filtering and classification in the South Asian type 2 diabetes dataset

Metric	Value
Total variants	500
Genome-wide significant ($p < 5 \times 10^{-8}$)	293
Suggestive ($p < 1 \times 10^{-5}$)	500
Early-onset associated	125
General T2D	375

This distribution suggests that a large fraction of the statistically strong associations in the dataset are present, as well as a well-identified set of prioritized variants to be relevant in the early-onset.

3.2 Genome-wide Association Patterns

The volcano plot (Figure 1) shows the distribution of the associations of variants in terms of their effect sizes and the statistical significance. There was also a large range of variants with moderate-to-high effect sizes and extremely low p-values, particularly around effect sizes of about 0.2. The variants that are associated with early-onset are overrepresented in highly significant signals with greater effect sizes, providing stronger genetic effects.

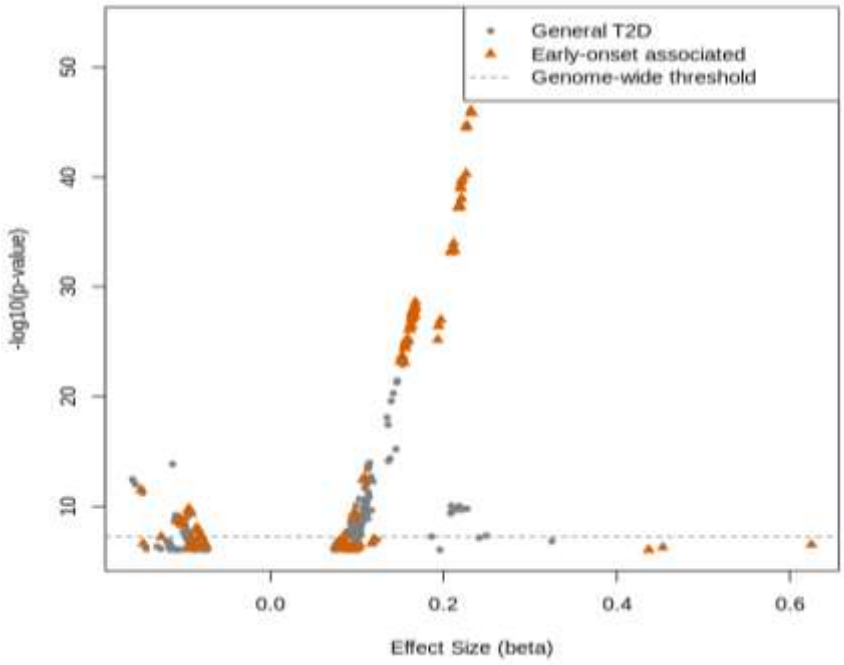

Figure 1. Volcano plot of genomic variants associated with T2D in South Asian populations

A Manhattan plot was used to analyze the genome-wide distribution of the association signals (Figure 2). The peaks of association were discrete, and there was a very high density of them on chromosome 10. Instead of being evenly distributed, association signals are concentrated in major genome loci, which are regions of high genetic impact.

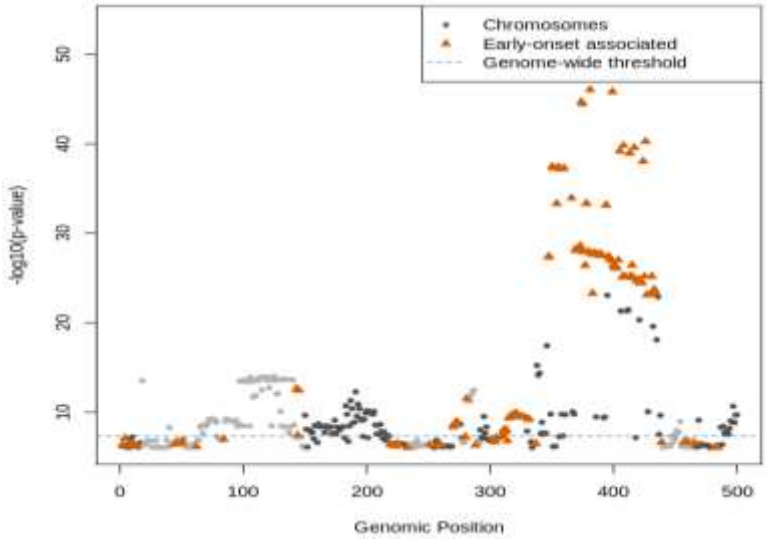

Figure 2. Manhattan plot showing genome-wide distribution of T2D-associated variants in South Asian populations

3.3 Identification of Top Early-Onset-Associated Variants

Table 2 displays the upper-ranking variants of the early-onset-related ones, according to the composite scoring model. Such variants combine statistical significance, effect size and functional annotation.

Table 2. Top early-onset-associated genomic variants in South Asian type 2 diabetes

Variant ID	rsID	Gene	Effect (β)	p-value	−log10(p)	Functional Class	EO Score
10:114758349:C:T	rs7903146	*TCF7L2*	0.2436	2.98×10^{-54}	53.53	Intron	1.309
10:114754784:T:C	rs35198068	*TCF7L2*	0.2424	1.42×10^{-53}	52.85	Intron	1.293
10:114754071:T:C	rs34872471	*TCF7L2*	0.2419	3.16×10^{-53}	52.50	Intron	1.284
10:114773926:C:T	rs12244851	*TCF7L2*	0.2341	7.06×10^{-50}	49.15	Intron	1.200
10:114754088:T:C	rs7901695	*TCF7L2*	0.2317	3.52×10^{-49}	48.45	Intron	1.181
8:118184783:C:T	rs13266634	*SLC30A8*	−0.1018	4.12×10^{-9}	8.38	Missense	1.101
8:41519462:A:G	rs515071	*ANK1*	0.0931	5.19×10^{-7}	6.29	Splice region	1.041
17:58308517:T:C	rs570707875	*USP32*	0.6250	2.96×10^{-7}	6.53	Intron	1.010

The most significant variants were highly concentrated in the *TCF7L2* locus, which revealed several highly significant signals with consistent effect sizes. Moreover, *SLC30A8*, a gene related to insulin secretion, and *ANK1*, which are related to metabolic regulation, were also found in the prioritized set among variants. It is noteworthy that there is a missense variant of *SLC30A8*, which implies that there could be functional implications in addition to the regulatory ones.

3.4 Functional Characterization of Variants

Figure 3 indicates the distribution of functional consequences of the variants. Most variants were either intronic or intergenic, and a smaller percentage of variants belonged to either the coding or regulatory regions. The vast majority of variants appear in the non-coding regions, indicating that the genetic architecture is largely regulatory.

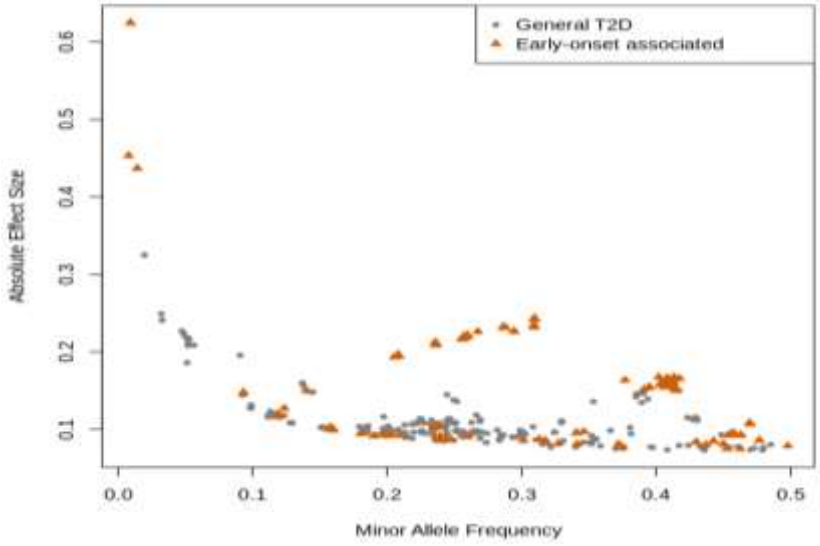

Figure 3. Distribution of functional consequences among type 2 diabetes–associated variants

Although of low frequency, variants of codings like missense mutations were observed in high-ranking early-onset-associated variants, which suggests their possibility of involvement in disease mechanisms.

3.5 Genetic Architecture of Early-Onset Variants

Figure 4 shows the dependence between the effect size and the allele frequency. The variants with less strength of allele frequency appeared to have greater effect sizes, whereas more frequent variants were usually less affected.

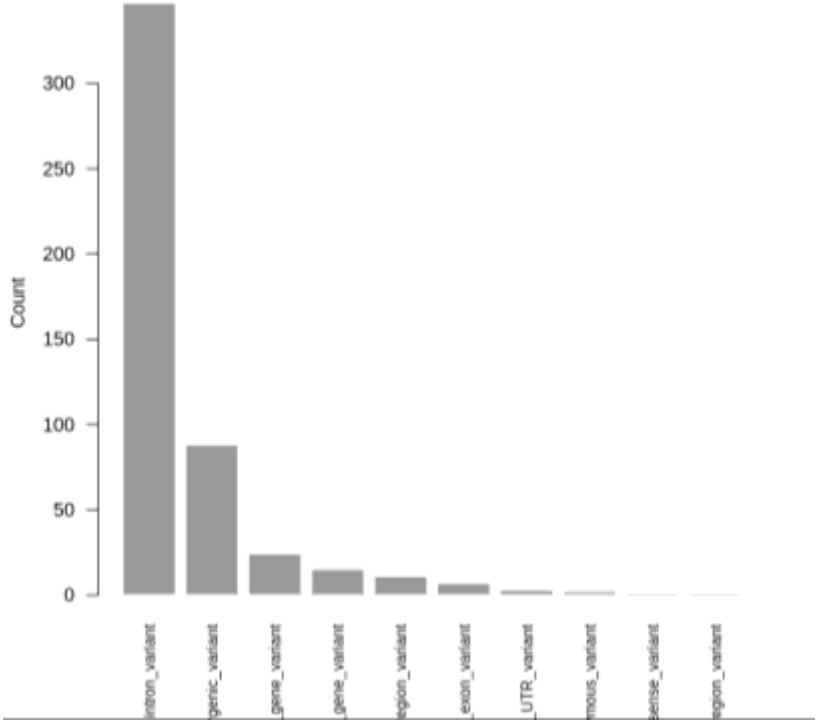

Figure 4. Relationship between effect size and minor allele frequency of type 2 diabetes–associated variants

Variants that are early-onset are likely to have larger effect sizes, which is in favor of a stronger genetic contribution. Figure 5 summarizes classification of variants into early-onset-associated and general.

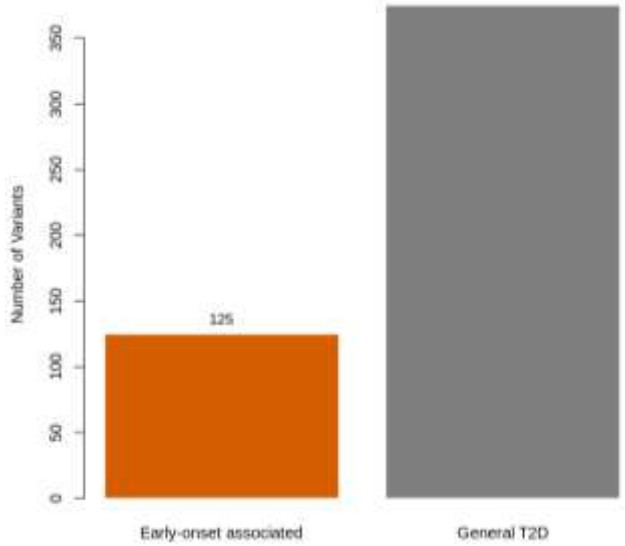

Figure 5. Classification of variants into early-onset-associated and general type 2 diabetes groups

3.6 Gene-Level Distribution of Association Signals

Table 3 provides a summary of the distribution of variants at the gene level. A number of loci had a high density of associated variants with *TCF7L2* (99 variants), and then *CDKAL1*, *IGF2BP2*, and *HHEX*.

Table 3. Gene-level summary of variant distribution in South Asian type 2 diabetes

Gene	Variant Count
TCF7L2	99
CDKAL1	68
IGF2BP2	53
HHEX	31

ADCY5	20
PPARG	20
UBE2E2	20
DGKB	17
HNF4A	14
ANK1	11

These genes are established factors of T2D predisposition, especially in the pathways of beta-cell functioning, insulin release, and glucose metabolism. The occurrence of variants in these loci being clustered further confirms their key involvement in the genetic determinism of T2D among South Asians.

4. DISCUSSION

The described findings indicate that the genetic architecture of T2D in the South Asian populations is highly non-random, and clustering of the association signals is highly clustered to a specific few key susceptibility loci. The large ratio of genome-wide significant variants suggests that the dataset that is selected is concentrated on strong genetic signals and not weak associations. Under this model, the group of variants that have been categorized as early-onset-associated always had a more significant statistical value, larger effect sizes, or increased functional significance, implying that such variants can be part of a more impactful aspect of the genetic environment. This trend suggests that predisposition in early-life is not conditioned by completely different loci; however, a subgroup of the already-established T2DM-associated variants exhibits enhanced genetic impacts or functional significance. The prioritization strategy, thus, seems to be effective in ranking the variants according to their relative biological and statistical impact, allowing for the identification of the signals that might be more important for the earlier disease onset in the wider range of T2D genetics.

One of the main findings was the strong prevalence of *TCF7L2*, not only in the ranked version table, but also in the level of gene summary. This is biologically possible as *TCF7L2* is among the most replicated T2D susceptibility genes and has a long-established glucose homeostasis, beta-cell biology, and incretin-related pathway role. Its significance in this analysis discussion, hence, contributes more to the validity of the identified signal and does not necessarily indicate bias in the dataset, especially since other much more significant variants were concentrated in the same locus (del Bosque-Plata et al., 2021). This relevance to South Asian populations is further supported by evidence demonstrating that *TCF7L2* gene variants significantly predispose individuals to the development of T2D, particularly in metabolically high-risk groups, reinforcing the broader biological importance of this locus (Katsoulis et al., 2018). Other than *TCF7L2*, other genes that were detected during the locus-level summary are highly consistent with prior documentation of T2D. The second most represented locus in the dataset, *CDKAL1*, was also found to be involved in the pancreatic beta-cell dysfunction and diabetic complications, which is highly in line with a disease model where impaired insulin secretion is a primary pathogenic mechanism (Ghosh et al., 2022). On the same note, *IGF2BP2* has been linked to metabolic dysregulation and has been suggested to be a significant mediator of the cellular processes involved in T2D, which accounts for its display at the top of the list in the present analysis (Cao et al., 2021). The *HHEX* enrichment should also be mentioned, with the variants of this gene being previously associated with the T2D risk and believed to affect the pathways that are pertinent to the development of the pancreas and its control over the glucose levels (Galavi et al., 2019). The fact that *ADCY5*, *PPARG*, and *UBE2E2* are also the most common loci also indicates the biological coherence of findings. *ADCY5* variants were also found to be associated with glycemic traits and impaired glucose regulation, which confirms that this locus is pathogenic in disease vulnerability among related metabolic phenotypes (Lin et al., 2020). *PPARG* is one of the major metabolic genes in adipocyte differentiation and insulin sensitivity, and the previous meta-analytic data indicate its role in T2D predisposition (Li et al., 2019). The *UBE2E2* is especially relevant to the early prioritization context as it was established that, in cases of altered expression in pancreatic beta-cells, the mass of the beta-cells can decrease and lead to the deficiency of glucose tolerance, which would serve as a reasonable mechanistic explanation of stronger or earlier diabetic phenotypes (Sakurai et al., 2024). An important functional dimension is also added to the interpretation by the list of variants ranked on top. Although the majority of the variants of higher priorities were intronic, the presence of a missense variant in *SLC30A8* and a splice-region variant in ANK1 suggests that coding or near-coding changes can also help to provide substantial value to higher-priority signals. This is applicable since non-coding variation is the predominant type of T2D genetics, but variations with stronger functional implications can be especially informative in case one is trying to prioritize loci based on their relevance in early-onset. That way, the current results align with the general evidence that the high-throughput genomic techniques are becoming more illuminating to the multidimensional regulatory and coding impacts of diabetes susceptibility (Dziewulska et al., 2018).

The framework of early-onset prioritization applied here does not provide the direct effects of age-at-diagnosis, but the findings are, nevertheless, consistent with the current body of literature on earlier and more severe manifestations of T2D in South Asians. Recent studies have indicated that the genetic aetiology of early-onset and progression-related T2D among South Asians is stronger inherited elements than can be evident based on the traditional definitions of cases by adult onset (Hodgson et al., 2025a). Similarly, South India-based studies have shown that the already known GWAS loci continue to be pertinent in the early onset of T2D, which confirms the assumption that the variants of diabetes previously identified might have a different degree of relevance in younger patients (Liju et al., 2020). The fact that early-onset-related variants in the current study were directionally related to have occupied regions of greater significance and, in other cases, greater effect size, is consequently directionally consistent with the larger literature.

Meanwhile, the consequences of such findings are to be taken with the necessary precaution. The central implication does not consist in the fact that certain variants were found to result in early-onset T2D, but in the fact that a consistent scoring

method can identify variants and loci, which should be further investigated in age-stratified cohorts. This kind of prioritization can assist in narrowing down the targets of candidate studies to pursue functional studies and enhance genetic risk modeling in South Asians and future endeavors to differentiate between early-progressive and more typical adult-onset T2D. The robust manifestation of loci like *HNF4A* and *CDKAL1* in recent South Asian exome-wide studies also indicates that integrating summary-statistic prioritization with coding-variant studies can be particularly enlightening in future studies (Hodgson et al., 2025b).

There are a number of restrictions that should be recognized. To begin with, the analysis was performed with a pre-defined set of variants retrieved through the database, restricting the possibility to conduct the exploration of the genome as a whole and finding new loci other than those contained in the set. Second, the clinical data at the individual level, such as the age of diagnosis, treatment history, and the presence or absence of complications, were not provided, which could not allow a direct assessment of the onset of the disease early and could only be interpreted in terms of suspected associations. Third, early-onset-associated variants were categorized using a composite scoring strategy that combined effect size, statistical significance and functional annotation. Though useful as a prioritisation strategy, this approach is not a replacement for a clinically defined early-onset phenotype. Further research is expected to confirm such results with larger cohorts of South Asians with extensive phenotypic data, especially age-of-onset data, to directly deal with early-onset disease. Whole-genome or exome sequencing data with functional validation would also enhance the process of identifying biologically relevant variants and help understand whether the prioritized loci are uniquely linked to the susceptibility in early life or are representative of the overall T2D risk.

5. CONCLUSION

The genetic variation in relation to type 2 diabetes in South Asian populations was centered on the few well-established areas of susceptibility loci, with *TCF7L2* being the most dominant signal and other smaller contributions by *CDKAL1*, *IGF2BP2*, *HHEX*, *ADCY5*, *PPARG*, *UBE2E2*, and *SLC30A8*. The results show that variants that were ranked using the composite framework were those variants that were more statistically supported, larger in magnitude, or functionally relevant than the larger set of associated variants. This justifies the usefulness of a systematic prioritization strategy to recognize loci of possible interest in early-onset disease in situations where there are no direct data on age-at-diagnosis. The domination of intronic and intergenic types also indicates that regulatory processes continue to play a key role in the genetic pathophysiology of T2D in this group, and the occurrence of functional signals in the form of selected coding and splice-region variants. Even though the analysis does not demonstrate the direct early-onset causality, it gives a narrowed collection of candidate variants and genes to be validated in the future. Subsequent research based on larger cohorts of South Asians with detailed phenotypic measures should be necessary to determine whether these prioritized loci are uniquely associated with early-onset susceptibility or represent a more general risk of T2DM.

REFERENCES

1. Ahmad, E., Lim, S., Lamptey, R., Webb, D. R., & Davies, M. J. (2022). Type 2 diabetes. *The Lancet*, *400*(10365), 1803-1820.
2. Cao, J., Yan, W., Ma, X., Huang, H., & Yan, H. (2021). Insulin-like growth factor 2 mRNA-binding protein 2—A potential link between type 2 diabetes mellitus and cancer. *The journal of clinical endocrinology & metabolism*, *106*(10), 2807-2818.
3. Chen, K., Zhang, H., Wu, N., Li, B., Li, S., & Mu, Y. (2025). Global risk factors, epidemiology, and disease burden of type 2 diabetes. *Science China Life Sciences*, 1-12.
4. Costanzo, M. C., von Grotthuss, M., Massung, J., Jang, D., Caulkins, L., Koesterer, R., ... & Abecasis, G. R. (2023). The Type 2 Diabetes Knowledge Portal: An open-access genetic resource dedicated to type 2 diabetes and related traits. *Cell metabolism*, *35*(4), 695-710.
5. De Rosa, S., Arcidiacono, B., Chiefari, E., Brunetti, A., Indolfi, C., & Foti, D. P. (2018). Type 2 diabetes mellitus and cardiovascular disease: genetic and epigenetic links. *Frontiers in endocrinology*, *9*, 2.
6. del Bosque-Plata, L., Martínez-Martínez, E., Espinoza-Camacho, M. Á., & Gragnoli, C. (2021). The role of TCF7L2 in type 2 diabetes. *Diabetes*, *70*(6), 1220-1228.
7. Dendup, T., Feng, X., Clingan, S., & Astell-Burt, T. (2018). Environmental risk factors for developing type 2 diabetes mellitus: a systematic review. *International journal of environmental research and public health*, *15*(1), 78.
8. Dziewulska, A., Dobosz, A. M., & Dobrzyn, A. (2018). High-throughput approaches onto uncover (Epi) genomic architecture of type 2 diabetes. *Genes*, *9*(8), 374.
9. Galavi, H., Mollashahee-Kohkan, F., Saravani, R., Sargazi, S., Noorzehi, N., & Shahraki, H. (2019). HHEX gene polymorphisms and type 2 diabetes mellitus: A case-control report from Iran. *Journal of cellular biochemistry*, *120*(10), 16445-16451.
10. Galicia-Garcia, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K. B., ... & Martín, C. (2020). Pathophysiology of type 2 diabetes mellitus. *International journal of molecular sciences*, *21*(17), 6275.
11. Ghosh, C., Das, N., Saha, S., Kundu, T., Sircar, D., & Roy, P. (2022). Involvement of Cdkal1 in the etiology of type 2 diabetes mellitus and microvascular diabetic complications: a review. *Journal of Diabetes & Metabolic Disorders*, *21*(1), 991-1001.
12. Hodgson, S., Bui, V., Bigossi, M., Stow, D., Maroteau, C., Williamson, A., ... & Siddiqui, M. K. (2025b). Exome-wide association study in 54,698 south Asians identifies novel type 2 diabetes associations with RNF19A, HNF4A, and dissects role of coding variants in GP2 and CDKAL1. *medRxiv*, 2025-09.
13. Hodgson, S., Williamson, A., Bigossi, M., Stow, D., Jacobs, B. M., Samuel, M., ... & Finer, S. (2025a). Genetic basis of early onset and progression of type 2 diabetes in South Asians. *Nature Medicine*, *31*(1), 323-331.

14. Huang, J. X., Liao, Y. F., & Li, Y. M. (2019). Clinical features and microvascular complications risk factors of early-onset type 2 diabetes mellitus. *Current Medical Science*, *39*(5), 754-758.
15. Jenum, A. K., Brekke, I., Mdala, I., Muilwijk, M., Ramachandran, A., Kjøllesdal, M., ... & van Valkengoed, I. G. (2019). Effects of dietary and physical activity interventions on the risk of type 2 diabetes in South Asians: meta-analysis of individual participant data from randomised controlled trials. *Diabetologia*, *62*(8), 1337-1348.
16. Katsoulis, K., Paschou, S. A., Hatzi, E., Tigas, S., Georgiou, I., & Tsatsoulis, A. (2018). TCF7L2 gene variants predispose to the development of type 2 diabetes mellitus among individuals with metabolic syndrome. *Hormones*, *17*(3), 359-365.
17. Kreienkamp, R. J., Voight, B. F., Gloyn, A. L., & Udler, M. S. (2023). Genetics of type 2 diabetes. *Diabetes in America [Internet]*.
18. Li, J., Niu, X., Li, J., & Wang, Q. (2019). Association of PPARG gene polymorphisms Pro12Ala with type 2 diabetes mellitus: a meta-analysis. *Current Diabetes Reviews*, *15*(4), 277-283.
19. Liju, S., Chidambaram, M., Mohan, V., & Radha, V. (2020). Impact of type 2 diabetes variants identified through genome-wide association studies in early-onset type 2 diabetes from South Indian population. *Genomics & informatics*, *18*(3), e27.
20. Lin, R., Yuan, Z., Zhang, C., Ju, H., Sun, Y., Huang, N., ... & Jin, L. (2020). Common genetic variants in ADCY5 and gestational glycemic traits. *PloS one*, *15*(3), e0230032.
21. Loh, M., Zhang, W., Ng, H. K., Schmid, K., Lamri, A., Tong, L., ... & Chambers, J. C. (2022). Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Communications Biology*, *5*(1), 329.
22. Misra, S., Ke, C., Srinivasan, S., Goyal, A., Nyriyenda, M. J., Florez, J. C., ... & Luk, A. (2023). Current insights and emerging trends in early-onset type 2 diabetes. *The Lancet Diabetes & Endocrinology*, *11*(10), 768-782.
23. Sakurai, Y., Kubota, N., Takamoto, I., Wada, N., Aihara, M., Hayashi, T., ... & Yamauchi, T. (2024). Overexpression of UBE2E2 in mouse pancreatic β-cells leads to glucose intolerance via reduction of β-cell mass. *Diabetes*, *73*(3), 474-489.
24. Strati, M., Moustaki, M., Psaltopoulou, T., Vryonidou, A., & Paschou, S. A. (2024). Early onset type 2 diabetes mellitus: an update. *Endocrine*, *85*(3), 965-978.
25. Suzuki, K., Hatzikotoulas, K., Southam, L., Taylor, H. J., Yin, X., Lorenz, K. M., ... & Kamanu, F. K. (2024). Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature*, *627*(8003), 347-357.
26. Tremblay, J., & Hamet, P. (2019). Environmental and genetic contributions to diabetes. *Metabolism*, *100*, 153952.