

Article

Susceptibility Loci for Type 2 Diabetes in the Ethnically Endogamous Indian Sindhi Population: A Pooled Blood Genome-Wide Association Study

Kanchan V. Pipal^{1,†}, Manju Mamtani^{1,2,†}, Ashwini A. Patel¹, Sujeet G. Jaiswal¹, Manisha T. Jaisinghani¹ and Hemant Kulkarni^{1,2,*}

¹ Lata Medical Research Foundation, Nagpur 440002, India; pipalkanchan17@gmail.com (K.V.P.); manju.mamtani@mnhresearch.com (M.M.); ashwinipatel1993@gmail.com (A.A.P.); sujeetgaiswal@gmail.com (S.G.J.); manishajaisinghani1001@gmail.com (M.T.J.)

² M&H Research, LLC, San Antonio, TX 78249, USA

* Correspondence: hemant.kulkarni@mnhresearch.com

† These authors contributed equally to this work.

Abstract: Type 2 diabetes (T2D) is a complex metabolic derangement that has a strong genetic basis. There is substantial population-specificity in the association of genetic variants with T2D. The Indian urban Sindhi population is at a high risk of T2D. The genetic basis of T2D in this population is unknown. We interrogated 28 pooled whole blood genomes of 1402 participants from the Diabetes In Sindhi Families In Nagpur (DISFIN) study using Illumina's Global Screening Array. From a total of 608,550 biallelic variants, 140 were significantly associated with T2D after adjusting for comorbidities, batch effects, pooling error, kinship status and pooling variation in a random effects multivariable logistic regression framework. Of the 102 well-characterized genes that these variants mapped onto, 70 genes have been previously reported to be associated with T2D to varying degrees with known functional relevance. Excluding open reading frames, intergenic non-coding elements and pseudogenes, our study identified 22 novel candidate genes in the Sindhi population studied. Our study thus points to the potential, interesting candidate genes associated with T2D in an ethnically endogamous population. These candidate genes need to be fully investigated in future studies.

Keywords: type 2 diabetes; genome-wide association study; ethnicity



Citation: Pipal, K.V.; Mamtani, M.; Patel, A.A.; Jaiswal, S.G.; Jaisinghani, M.T.; Kulkarni, H. Susceptibility Loci for Type 2 Diabetes in the Ethnically Endogamous Indian Sindhi Population: A Pooled Blood Genome-Wide Association Study. *Genes* **2022**, *13*, 1298. <https://doi.org/10.3390/genes13081298>

Academic Editor:
Hirokazu Takahashi

Received: 1 July 2022
Accepted: 19 July 2022
Published: 22 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is well established that type 2 diabetes (T2D) has a strong genetic basis. Studies around the world report that T2D as a trait has a heritability ranging from 35–80% indicating that a large proportion of the variability in T2D can be explained by genetics [1–5]. Genome-wide association studies have identified several potential single nucleotide polymorphisms (SNP) that significantly contribute to the genetic basis of T2D [6]. Still, it is recognized that the strength of association of these SNPs with T2D can vary across world populations [7–10]. In that vein, a search for population-specific genetic determinants of T2D still continues.

The Indian Sindhi ethnically endogamous group has historically encountered the challenges of stress and migration. The partition of India and Pakistan in 1947 led to an exodus of Hindu Sindhis residing in the Sindh province (now in Pakistan) to various parts across India. This exodus remains a critical root of the chronic stresses experienced by generations of Indian Sindhis [2,11]. Combined with these stressors, the energy-rich diet and relative lack of physical activity place this population at an enhanced risk of T2D. Despite these well-known environmental determinants of T2D, in the first study of extended and complex Sindhi pedigrees, we observed that the heritability of T2D was 35% [2]. In the era of personalized medicine, this information is important since individual susceptibility to T2D and the individual's response to T2D treatment can both be determined, in part,

by the genetic disposition [12]. To that end, a genome-level interrogation of the potential markers of T2D in the Indian Sindhi population is needed.

We conducted a pooled blood genome-wide association study of T2D in the pedigrees of Indian Sindhi families enrolled in the Diabetes In Sindhi Families In Nagpur (DISFIN) study [2]. This population has a substantial coexistence of metabolic comorbidities such as hypertension, dyslipidemia, general obesity and central obesity. We used innovative statistical approaches to account for these comorbidities and—constrained by the project costs—conducted a pooled whole blood genome-wide association study (GWAS). Here, we report the results of our study that identified interesting leads into the genetic basis of T2D in the Indian Sindhi families.

2. Materials and Methods

2.1. Study Participants

The clinical and genetic data for the present study came from the participants in the DISFIN study. Details of the enrollment protocol, inclusion and exclusion criteria and overall design of the study have been described elsewhere [2]. Briefly, we enrolled endogamous Sindhi families to construct family pedigrees. The pedigrees ranged from simple nuclear ones to complex and extended three generation families. Consenting participants with at least one known patient of T2D in the family were included in the study. Other inclusion criteria were a resident of the study area (Jaripatka, Mecosabag and Khamla areas of Nagpur where the Sindhi ethnic population mostly resides); self-reported Sindhi ethnicity and age ≥ 20 years. Pregnant or lactating women and patients with type 1 diabetes (known or suggested by serum C peptide) were excluded. After investigator-administered, semi-structured interviews and clinical examination, a trained phlebotomist collected blood samples for laboratory assays. Part of the sampled venous blood was stored at -80°C to conduct genetic investigations later. Participant enrollment and blood sample collection took place between 1 March 2016 and 28 February 2017.

2.2. Definitions of Metabolic Conditions

The primary goal of the study was to conduct a genome-wide interrogation in the context of T2. Since metabolic comorbidities such as obesity, hypertension and dyslipidemia commonly coexist with T2D, we also investigated the study participants with respect to these conditions. T2D was defined [13] as self-reported diabetes OR currently on anti-diabetics OR fasting plasma glucose ≥ 126 mg/dL OR random blood glucose ≥ 200 mg/dL OR HbA1c concentration $\geq 6.5\%$. Hypertension was defined [14] as self-reported hypertension OR currently on anti-hypertensives OR systolic blood pressure ≥ 130 mmHg OR diastolic blood pressure ≥ 85 mmHg. Central obesity was defined based on cutoffs for Indian population [15] as a waist circumference ≥ 90 cm for males and ≥ 85 cm for females. Dyslipidemia was defined [16] as presence of any of the following: serum triglycerides ≥ 150 mg/dL and serum high density lipoproteins < 40 mg/dL (for males) or < 50 mg/dL (for females).

2.3. Pool Definitions for GWAS

We conducted a pooled whole blood genome-wide association study. This technique is now well established as an acceptably accurate and inexpensive alternative to individual genome-wide genotyping [17]. To account for the potential contribution of metabolic comorbidities to genome-wide associations, we constructed whole blood pools based on a combination of the presence of T2D, central obesity, hypertension and dyslipidemia. Using these four binary traits we first generated a total of 16 potential combinations. We combined those categories that had a frequency $< 1\%$ with preceding categories. This strategy resulted in a total of 14 adequately represented pools that were based on the metabolic comorbidity profile. To ensure repeatability, we ran the genotyping analyses on duplicate pools. Therefore, we had a total of 28 whole blood pools that were used for genotyping analyses.

2.4. Whole Blood Pool Construction and Genotyping

We aimed to collect at least 3 µg of DNA for each pool from the K-EDTA stored blood samples were thawed gradually to room temperature in a vibration free environment. From these samples, 100 µL of thawed blood was pipetted per sample and added to specific pools defined by the comorbidity profile. From these pooled blood samples, two aliquots of 2 mL each were prepared, thus resulting in 28 pooled samples for GWAS. Whole blood samples were transported to the genotyping laboratory within 6 h on dry ice. All genotyping was carried out at the Genetics Laboratory, MedGenome Labs Ltd. (<https://diagnostics.medgenome.com/>, accessed on 30 June 2022), Bengaluru, India. Genotyping of the pool samples was conducted using Illumina's Infinium Global Screening Array (GSA) platform from harvested genomic DNA (Qiagen, DNeasy Blood and Tissue Kit). A total of 200 ng of genomic DNA was amplified and incubated on day 1. This was followed next day with fragment amplification, precipitation, resuspension, bead chip preparation and hybridization of the sample on the bead chip. Stained samples with red and green channel tags on the bead chip were then imaged and subjected to an auto calling algorithm to generate the B allele frequencies in the pools.

2.5. Statistical Analyses

For each included variant on the GSA array, we had 28 estimations of the B allele frequencies. We accounted for the potential of confounded associations ascribable to the presence of comorbidities, batch effects, intra-replicate correlation, within-pool degree of kinship and the random effects across pools. For this we used the mixed effects logistic regression format to estimate the T statistic from a Wald test to test the significance of association of a given variant with the risk of T2D. Specifically, we used the following regression model to estimate the strength of association:

$$\text{logit}(T2D) = \beta_0 + \beta_s \text{BAF} + \beta_1 \text{HT} + \beta_2 \text{COB} + \beta_3 \text{DYL} + \beta_b \text{BAT} + \beta_r \text{REP} + \beta_k \text{PHI} + \text{RE}(\text{POOL})$$

where, T2D is an indicator variable for presence of T2D; BAF is a continuous variable indicating the B allele frequency; HT, COB and DYL represent the concomitant presence of hypertension, central obesity and dyslipidemia, respectively; BAT is an indicator variable for the chip identifier; REP represents the replicate ID; PHI is the within pool degree of kinship and RE(Pool) represents the random effects across the study pools. The regression coefficients in the equation were used to quantify the differential B allele frequency (β_s), influence of comorbidities (β_1 – β_3), batch effect (β_b), pooling error (β_r) and kinship effect (β_k). All the models used a weighted approach based on pool frequency weights. The rigid Bonferroni correction was used to account for multiple testing and the global type I error rate was thus adjusted to 8.216×10^{-8} . These analyses were conducted using dedicated scripts in R. Manhattan and QQ plots were generated using the qqman library [18] in R. Pooling error was estimated as described by MacGregor et al. [19].

2.6. Functional Relevance of Strongly Associated Variants and Genes

To understand the detected associations from a functional perspective, we used three approaches. First, all the variants and genes were annotated using the DisGeNet database (<https://www.disgenet.org/>, accessed on 30 June 2022) by selecting the diabetes-specific associations. Second, the T2D Knowledge portal (T2DKP) [20] is a meta-analytic summary of genome-wide associations from worldwide populations. We compared the significant associations found in this study with those reported in the T2D Knowledge portal. Third, we also searched the Harmonize database [21] for diabetes-specific gene associations in the context of those observed in the present study. Finally, we constructed a functional gene network of the significantly associated genes using the NetworkAnalyst web resource [22].

3. Results

3.1. Study Participants, Pools and the GSA Variants

The DISFIN study enrolled a total of 1444 participants whose detailed description has been given elsewhere [2]. From this group of participants, blood samples and all laboratory assays were available on 1402 (97%) participants and were included in the present study. The distribution of sociodemographic and clinical features of the included study participants ($n = 1402$) was similar to that in the entire study sample of 1444 participants. The prevalence of T2D, central obesity, hypertension and dyslipidemia in the study sample was 29.82%, 71.98%, 52.93% and 31.03%, respectively.

Whole blood pools were designed based on the presence of four metabolic conditions (T2D, central obesity, hypertension and dyslipidemia). From a potential 16 combinations of these metabolic conditions, we combined infrequent combinations (<1%, highlighted rows in Table 1) giving a total of 14 pools. Further, since these pools were used for genetic analyses in duplicates, we had a total of 28 study pools. The frequency of the 28 constructed whole blood pools is shown in Table 1. From a total of 665,608 variants included in the GSA platform, we included a total of 608,550 variants after excluding them based on sex chromosomes, mitochondria and those with a minor allele frequency (MAF) < 1%. The inclusion criteria for the variants and their distribution across the chromosomes are shown in Figure 1.

Table 1. Description of whole blood pools.

Pool Id	T2D	CO	HTN	DL	<i>n</i>	%
1	No	No	No	No	190	13.55
2	No	No	No	Yes	49	3.50
3	No	No	Yes	No	65	4.64
4	No	No	Yes	Yes	30	2.14
5	No	Yes	No	No	230	16.41
6	No	Yes	No	Yes	90	6.42
7	No	Yes	Yes	No	224	15.98
8	No	Yes	Yes	Yes	106	7.56
9	Yes	No	No	No	22	1.57
	Yes	No	No	Yes	7	0.50
10	Yes	No	Yes	No	19	1.36
	Yes	No	Yes	Yes	11	0.78
11	Yes	Yes	No	No	38	2.71
12	Yes	Yes	No	Yes	34	2.43
13	Yes	Yes	Yes	No	179	12.77
14	Yes	Yes	Yes	Yes	108	7.70

T2D, type 2 diabetes; CO, central obesity; HTN, hypertension; DL, dyslipidemia.

3.2. Pooled Blood and Genotyping Quality Control

The average GenTrain score (a measure of the single nucleotide polymorphism (SNP) calling quality reported by Illumina arrays) across the included variants was 0.8312 (95% confidence interval 0.8309–0.8314). The average estimate of the pooling error was -0.0067 (95% confidence interval was -0.0295 – 0.0162). Thus, both the genotyping and pooling error estimates were in the acceptable range.

3.3. Pooled GWAS Results—Variants

The Manhattan plot (Figure 2A) shows the results of our pooled GWAS study. From the total variants included in this study, 140 were found to be significantly associated with the risk of T2D at the genome level ($p < 8.2163 \times 10^{-8}$). The detailed identification and annotation of the significantly associated variants is provided in Supplementary Tables S1 and S2. Briefly, significant variants were found on all chromosomes, but the majority were located on autosomes 3 and 7 ($n = 13$ each), 2 ($n = 11$) and 1 ($n = 10$). No strand preference was observed (69 and 71 variants on the + and—strand, respectively). The SNP calling quality of

the significantly associated variants was comparable to that of the whole genome (average GenTrain score 0.8264). The average MAF was 25.42% and ranged between 1.01% to 49.86%. A total of 79 variants (56.43%) were within genes with known biological functions. The remaining 61 variants were located within an average of 73,632 bp (95% confidence interval 49,669–97,596 bp) of a gene with known biological function. The Q-Q plot (Figure 2B) showed that there was no systematic inflation ($\lambda = 0.9517$) at the level of the genome—indeed the Q-Q plot remained slightly below the line of expectation for all the variants except those that were significantly associated with T2D.

The top five most significantly associated variants are highlighted in Figure 2A,C. These were: rs1001179 (T→C polymorphism, $P = 9.06 \times 10^{-25}$), rs480948 (A→G polymorphism, $P = 2.36 \times 10^{-24}$), rs360745 (T→C polymorphism, $P = 2.98 \times 10^{-24}$), rs7711236 (T→G polymorphism, $P = 3.87 \times 10^{-24}$) and rs73219073 (T→C polymorphism, $P = 4.62 \times 10^{-24}$). These variants were located on chromosomes 11, 11, 3, 5 and 21, respectively. The top first, second, third and fifth variants were related to the catalase (CAT), mastermind like transcriptional coactivator 2 (MAML2), IQ motif and Sec7 domain ArfGEF 1 (IQSEC1) and PR/SET domain 15 (PRDM15) genes, respectively. Violin plots showing the distribution of these variants in study pools with regard to the presence or absence of T2D are shown in Figure 2C.

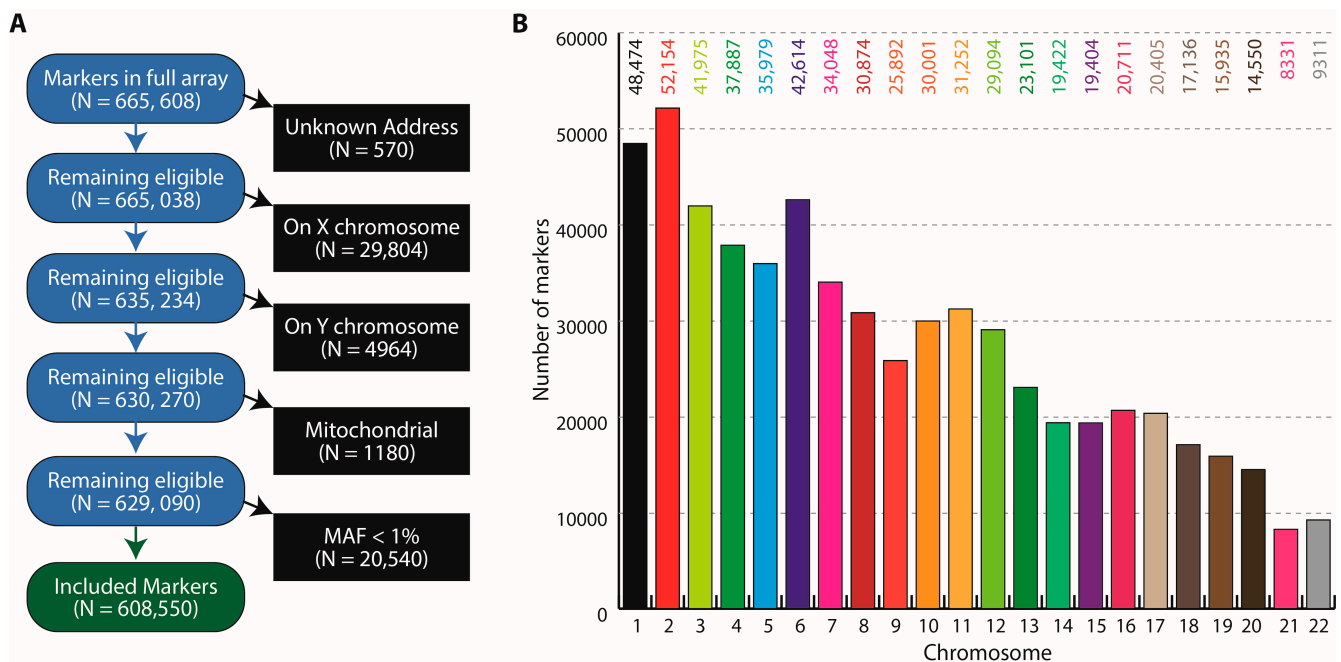


Figure 1. Genome-wide single nucleotide polymorphism (SNP) markers used in the pooled GWAS study. **(A)** Inclusion criteria for markers. Excluded markers are shown in black boxes, candidate markers are shown in blue boxes and the final number of included markers are shown in the green box. **(B)** Distribution of the included SNP markers across the genome. Bars show the number of markers on the indicated chromosome. The numbers are shown at the top of each bar.

3.4. Pooled GWAS Results—Genes

The significantly associated variants ($n = 140$) mapped to a total of 131 nearing or overlapping genes (Figure 3) of which 29 genes represent uncharacterized genes with no official names. We interrogated the relevance of the remaining 102 genes in the context of T2D by searching whether these genes are mentioned in the T2DKP ($n = 4344$ diabetes-related genes), DisGeNET ($n = 2359$ diabetes-related genes) or Harmonizome ($n = 3381$ diabetes-related genes) databases. We observed that nine of the significantly associated genes (ADCY5, MIP, EBF2, MITF, SLC30A8, NCAM1, HDAC9, CSF1 and DNASE1) were found in both T2DKP and in DisGeNET; 13 genes (IQSEC1, LHX2, SOX7, HSD17B12, PKNOX2,

IGSF21, DNAH1, RPTOR, SAXO1, MAGI2, HNRNPAB, B4GALNT4 and ZNF385D) were found in T2DKP only while eight genes (CAT, ERG, CORO2B, ALMS1, TNFRSF11A, RYR2, KAZN and DPYS) were found in DisGeNET only. Further, the Harmonizome database identified another 22 genes (ARHGAP42, CA10, CACNA2D3, CGNL1, CLIC5, DLGAP1, FAT3, FRMD4A, GPR6, HECW1, IQCJ-SCHIP1, MAML2, MBOAT1, NCALD, NRP2, PBRM1, PTPRM, SLC24A3, SORCS2, SPOCK1, SYNDIG1 and TMEM132B) as diabetes-related from those found to be significant in this study. Thus, our study identified a total of 50 genes that were not included in T2DKP, DisGeNET or Harmonizome databases as diabetes related.

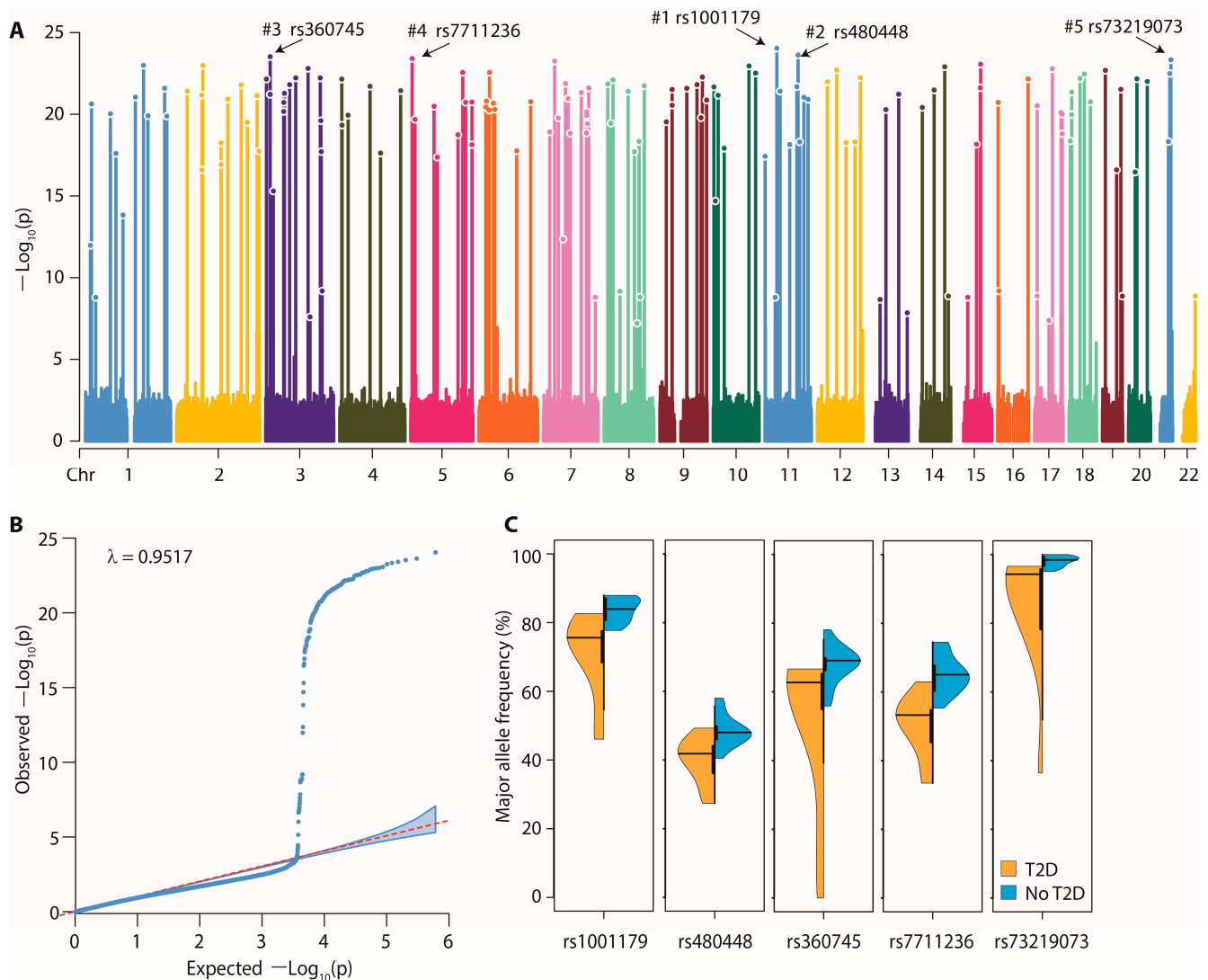


Figure 2. Results of the whole blood pooled genome-wide association study. (A) Manhattan plot. The points and droplines indicate the log-transformed, adjusted and statistically significant p -values. The five topmost significant associations are numbered as 1 through 5 and the corresponding SNP markers are shown at the top of the plot. (B) QQ plot. The plot shows the relationship between observed and expected p -value distribution (with 95% confidence bands). The genomic inflation factor (λ) is shown at the top of the plot. (C) Violin plots for the association of the top five significant SNP markers.

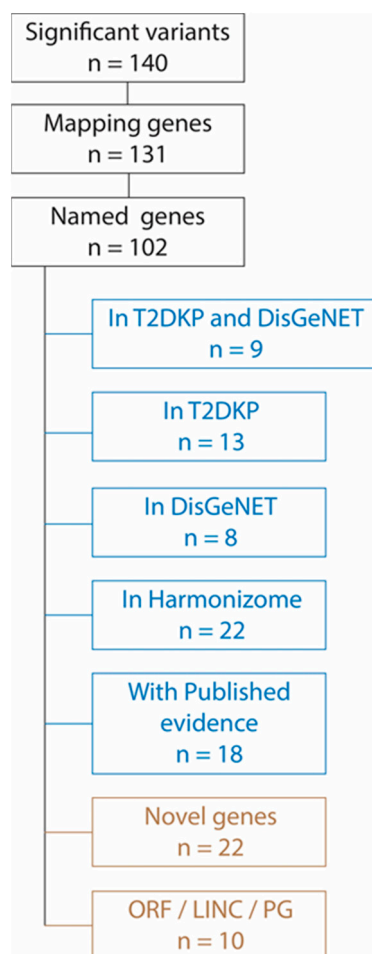


Figure 3. Characterization of the genes associated with T2D. T2DKP, Type 2 Diabetes Knowledge Portal; Dis-GeNET, the Disease–Gene Network database; ORF, open reading frame; LINC, long intergenic non-coding element; PG, pseudogene.

The detailed literature search revealed that 18 of these 50 genes have been associated with T2D previously even if these genes are not included in the T2DKP, DisGeNET and Harmonizome datasets. Published evidence implicated these 18 genes directly through population studies (GBP6, GSTO1, HHIPL1, INTS10, RGS16 and RNU6-679P), indirectly through association with other metabolic conditions (CCDC69, DNAH2, GALNT17, HERC5, KIF5C, MARCH2 and PRDM14) or through association with complications of T2D (CASC15, CCDC107, MIR147A, PFKFB3 and RNF166). Detailed annotation of the published literature in this regard is provided in Supplementary Table S3) [23–41]. From the remaining genes identified to be significantly associated with T2D in this study, there were two open reading frames, three long intergenic non-coding elements and a pseudogene. Thus, we identified a set of 22 genes as novel associations with T2D in the Sindhi population studied (Figure 3). These genes were: ARMH4, CDADC1, CYP2T3P, DRAIC, GNPTG, GUSBP6, HSPA8P13, MYOZ2, NDUFB9P2, PLEKHG1, PRDM15, RN7SKP144, RN7SKP203, RN7SKP250, RNF121, RNU4ATAC7P, RNVU1-6, RPL34P19, SIT1, TMTC1, TXNL1P1 and ZNF765.

3.5. Functional Relevance of the Significantly Associated Genes

Gene functional network analyses (Figure 4) showed that seven distinct functional blocks could be identified from the significantly associated genes. These blocks represented genes associated with the neuronal system functioning (CACNA2D3, CHRNA7, NCALD, ADCY5, RYR2, SLC24A3, SLC 30A8 and MIP); genes associated with Notch

signaling (MAML2 and HDAC9); one gene each representing the diabetes-breast cancer nexus (BRCA2); vitamin C metabolism (GSTO1); mTOR signaling (RPTOR); pyrimidine catabolism (DPYS) and endothelial function (NRP2).

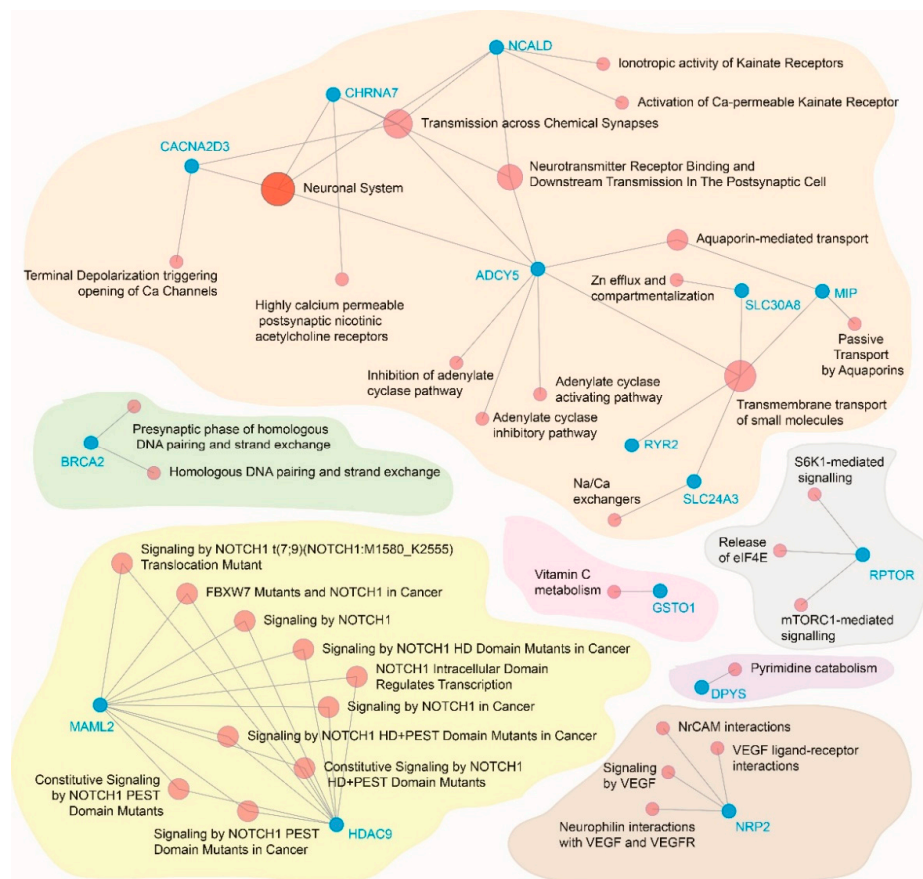


Figure 4. Functional gene network identified from the genes significantly associated with type 2 diabetes. Color coded irregularly shaped polygons represent the functional blocks. Blue nodes represent genes while pink nodes represent functional ontogeny.

4. Discussion

Type 2 diabetes is a complex metabolic disease with a wide web of causation. It is now well known that T2D has a strong genetic basis, but the list of genes implicated in the web of T2D causation is not fully established. This is partly because there is a variation in the strength and specificity of the association (especially in the context of GWAS) of genomic variants across world populations. For example, most of the GWAS studies published relating to T2D have been conducted in the population of European and Asian ancestry with very few studies in African populations. A recent meta-analysis by Chen et al. [42] highlights this point. In this context, it is noteworthy that there are very few GWAS studies on T2D based on Indian populations. Tabassum et al. [43] published a GWAS largely based on the Indo-European and Dravidian population with the strongest association signal around the TMEM163 gene. In the same year, Saxena et al. [44] published a study on the Punjabi Sikh population with a novel signal at the SGCG gene. Recently, a study [45] published on the genetic basis of T2D in three ethnically endogamous Indian populations based in Tamil Nadu, Rajasthan and Andhra Pradesh states (the INDIGENIUS Consortium) also demonstrated substantial population-specificity in the genetic basis of T2D within Indian populations. It is therefore a need to conduct T2D GWAS studies in ethnically coherent subgroups within India. Our study represents an attempt in that direction with a focus on the Indian urban Sindhi ethnicity.

In this pooled whole blood GWAS study, we observed that after adjusting for comorbidities, genetic relatedness, pooling variation and batch effects, a total of 102 well-characterized genes were associated with T2D in the ethnically endogamous Sindhi population. Of these 102 genes, 70 genes had evidence of association consistent with previously published studies and there were 22 novel genes significantly associated with T2D that may indicate a Sindhi population-specificity. The functional relevance and generalization of these associations needs to be investigated in future studies. On the other hand, the functional relevance of the 70 well-characterized genes (shown in Figure 4) extracted some interesting functional blocks. First, the genes associated with the neuronal system were consistent with the fact the brain signaling system is commonly involved before and during the pathogenesis of T2D [46–48]. Notable among these are the ADCY5, SLC30A8 and SLC24A3 genes that are involved in neurohormonal transport and solute (especially sodium, calcium and zinc) transport. In addition, the gene coding type 2 ryanodine receptor (RYR2) regulates calcium release in the endoplasmic reticulum of the pancreatic beta cells and is thus known to be involved in insulin secretion and glycemic homeostasis [49]. Second, the association of the BRCA2 with T2D raises interesting possibilities in this population since simple dietary measures have been shown to significantly improve glucose metabolism in BRCA mutation carriers [50]. Third, the association of the RPTOR gene with T2D is also interesting. This gene partakes in the mTORC1/Raptor signaling pathway and is known to regulate beta-cell maturation and insulin synthesis [51]. Fourth, the role of Notch signaling is essential to beta-cell maturation during the embryonic stage and during adult life [52,53]. We found two genes (MAML2 and HDAC9) that are strategically placed in the Notch signaling pathway and were associated with T2D in this study. Lastly, we also found association of genes implying vitamin C metabolism, pyrimidine metabolism and endothelial function regulation all of which make interesting targets for investigation in future studies.

It is also instructive to consider the potential mechanistic contribution of the top five most significantly associated variants (SAV, shown in Figure 2C). The topmost significantly associated variant (rs1001179) maps to the regulatory moiety of the CAT gene that codes for the catalase protein. This variant influences the transcription factor binding and thereby regulates catalase expression [54]. Concordantly, it has been demonstrated that lower circulating concentration of catalase is significantly associated with the risk of T2D [54,55]. The second most SAV (rs480948) mapped on to the MAML2 gene for which substantial evidence exists supporting its participation in the Notch signaling pathway as described above. Musicant et al. [56] have demonstrated that genomic rearrangements involving the MAML2 gene are associated with an inhibitory influence on the PPAR γ and IGF-1 expression. Similarly, the fifth most SAV (rs73219073 related to the PRDM15 gene) is also critically placed in the genetics of T2D. This gene has been shown to partake in the regulation of PI3K/KRT/mTOR pathway and during the embryonic pancreatic development [57]. Thus, for three of the five SAV, there exists substantial published evidence for potential involvement in the pathophysiology and genetics of T2D. To our knowledge, such supporting evidence is not available for the remaining two SAVs (rs360745 and rs7711236).

There are strengths and weakness in our study. Among strengths, first this is a pooled GWAS that identified most of the known gene–diabetes associations with a few novel associations. The fact that ~70% of the associations observed in this GWAS have been previously reported by other scientific groups in other populations lends an indirect credence to the observations. These observations also have a high plausibility considering the strong biological rationale that supports the association studies. Second, due to the whole pooling blood approach used in the study, we needed 28 blood pools compared to the sample size of 1402 individuals—a genotyping cost reduction by 98% with acceptable pooling error. McGregor et al. [19] have previously shown that this approach is effective in picking up association signals accurately. Arguably, pool construction error may be less important than genotyping error due to the arrays [58]. In resource-limited settings, such

an approach can greatly facilitate the first genetic screens for many other conditions. Future studies can then specifically address the candidate genes identified during this screening. Third, our study sample used a family-based approach. While most of the significantly associated variants had $MAF > 10\%$, the family-based approach implicitly places confidence in the observed associations of variants with lower MAFs.

Conversely, our study has some limitations. First, the pooled GWAS approach is both a strength and a weakness. Presumably, a regular GWAS on all individuals ($n = 1402$) would have led to similar inferences since the estimated pooling error was low and acceptable. However, in the absence a full GWAS study, it is not possible to comment about the robustness of the observed associations. Second, due to the ethnically endogamous and genetically related nature of the study population, it is practically very challenging to design another such cohort for validation purposes. As a result, generalization of the observed associations is neither possible nor desired. We only aimed to find the association signals in the light of those reported from other populations within and outside India. Third, biological explanation for the functional role of the genes is currently not available and cannot be inferred from this study. For an efficient resource utilization, future studies need to specifically investigate the functional role of the genes through transcriptomic studies. We do not have gene expression data on the candidate genes of interest identified by our study, but future studies need to conduct gene expression analyses. We therefore compared the associations observed in this study with those reported in well-established repositories and in other published studies.

5. Conclusions

Without over-interpretation of the data, we conclude that our pooled whole blood GWAS in families of T2D patients uncovered interesting candidate genes for future investigation. While 70% of the observed associations have been previously identified and investigated by other research groups in different populations, we also report 22 novel gene-T2D associations that need to be interrogated more fully in future studies. To that end, our study also points towards the possibility of population-specificity of the observed associations in the urban Sindhi population in India studied here.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13081298/s1>, Table S1: Variants significantly associated with T2D in the pooled whole blood genome-wide association study; Table S2: Nearest genes for the significantly associated variants in the pooled whole blood genome-wide association study; Table S3: Literature support for significantly associated T2D genes not listed in T2DKP, DisGeNET and Harmonizome databases.

Author Contributions: Conceptualization, M.M. and H.K.; methodology, H.K.; R scripting, H.K.; formal analysis, M.M. and H.K.; clinical and laboratory data collection, K.V.P., A.A.P., S.G.J. and M.T.J.; data curation, K.V.P., A.A.P. and H.K.; writing—original draft preparation, K.V.P., M.M. and H.K.; writing—review and editing, M.M. and H.K.; supervision, M.M.; project administration, M.M.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded through the internal funding mechanism of Lata Medical Research Foundation, Nagpur, India. The APC was also funded by Lata Medical Research Foundation, Nagpur, India.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of University of Texas Rio Grande Valley (protocol code 2016-126-07 [940406], dated 30 August 2016) and by the Institutional Ethics Committee of the Lata Medical Research Foundation, Nagpur, India (protocol ID RPC21A, dated 7 May 2016).

Informed Consent Statement: Written, informed consent was obtained from all subjects involved in the DISFIN study.

Data Availability Statement: Lata Medical Research Foundation’s Institutional Ethics Committee (LMRF-IEC) does not allow public data sharing to avoid potential identification. If data is requested for verification of results, we will seek permission from the LMRF-IEC before the requested data can be released. For further clarification of the LMRF-IEC’s data access policy as well as for data access requests, please contact: Dr. Prabir Kumar Das, Member Secretary, Institutional Ethics Committee, Lata Medical Research Foundation, Kinkine Kutir, Vasant Nagar, Nagpur—440022, Ph. No. 91-8805023450, Email: prabir_das23@rediffmail.com.

Acknowledgments: The authors are indebted to Archana Patel, Smita Puppalwar, Shilpa Pawar and Riddhi Tharani for administrative support; to Shantanu Sengupta, Sengupta Hospital and Research Institute, Nagpur, India for the laboratory support; to Neeti Chhangani for financial management; and to Dattaraj Gondane, Harshad Sonavane and Amar Bagde for logistics during the data collection phase.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Rana, S.; Ali, S.; Wani, H.A.; Mushtaq, Q.D.; Sharma, S.; Rehman, M.U. Metabolic syndrome and underlying genetic determinants—A systematic review. *J. Diabetes Metab. Disord.* **2022**, *21*, 1095–1104. [CrossRef] [PubMed]
2. Mamtani, M.; Jaisinghani, M.T.; Jaiswal, S.G.; Pipal, K.V.; Patel, A.A.; Kulkarni, H. Genetic association of anthropometric traits with type 2 diabetes in ethnically endogamous Sindhi families. *PLoS ONE* **2021**, *16*, e0257390. [CrossRef] [PubMed]
3. Asamoah, E.A.; Obirikorang, C.; Acheampong, E.; Annani-Akollor, M.E.; Laing, E.F.; Owiredo, E.W.; Anto, E.O. Heritability and Genetics of Type 2 Diabetes Mellitus in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. *J. Diabetes Res.* **2020**, *2020*, 3198671. [CrossRef]
4. Cuschieri, S. The genetic side of type 2 diabetes—A review. *Diabetes Metab. Syndr.* **2019**, *13*, 2503–2506. [CrossRef] [PubMed]
5. Gu, H.F. Genetic and Epigenetic Studies in Diabetic Kidney Disease. *Front. Genet.* **2019**, *10*, 507. [CrossRef] [PubMed]
6. DeForest, N.; Majithia, A.R. Genetics of Type 2 Diabetes: Implications from Large-Scale Studies. *Curr. Diabetes Rep.* **2022**, *22*, 227–235. [CrossRef] [PubMed]
7. Xia, W.; Chen, N.; Peng, W.; Jia, X.; Yu, Y.; Wu, X.; Gao, H. Systematic Meta-analysis Revealed an Association of PGC-1alpha rs8192678 Polymorphism in Type 2 Diabetes Mellitus. *Dis. Markers* **2019**, *2019*, 2970401. [CrossRef]
8. Chowdhury, R.; Narayan, K.M.; Zabetian, A.; Raj, S.; Tabassum, R. Genetic studies of type 2 diabetes in South Asians: A systematic overview. *Curr. Diabetes Rev.* **2014**, *10*, 258–274. [CrossRef]
9. Fu, J.; Festen, E.A.; Wijmenga, C. Multi-ethnic studies in complex traits. *Hum. Mol. Genet.* **2011**, *20*, R206–R213. [CrossRef]
10. Ukkola, O. Genetic variants of ghrelin in metabolic disorders. *Peptides* **2011**, *32*, 2319–2322. [CrossRef]
11. Carter, M. *Indians and the Colonial Diaspora*; Institute of Southeast Asian Studies: Singapore, 2008.
12. Florez, J.C.; Pearson, E.R. A roadmap to achieve pharmacological precision medicine in diabetes. *Diabetologia* **2022**. [CrossRef] [PubMed]
13. Kumar, R.; Nandhini, L.P.; Kamalanathan, S.; Sahoo, J.; Vivekanadan, M. Evidence for current diagnostic criteria of diabetes mellitus. *World J. Diabetes* **2016**, *7*, 396–405. [CrossRef] [PubMed]
14. Whelton, P.K.; Carey, R.M.; Aronow, W.S.; Casey, D.E., Jr.; Collins, K.J.; Dennison Himmelfarb, C.; DePalma, S.M.; Gidding, S.; Jamerson, K.A.; Jones, D.W.; et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* **2018**, *71*, e116–e135. [CrossRef]
15. Snehaltha, C.; Viswanathan, V.; Ramachandran, A. Cutoff values for normal anthropometric variables in asian Indian adults. *Diabetes Care* **2003**, *26*, 1380–1384. [CrossRef] [PubMed]
16. Sharma, Y.K.; Prakash, N.; Gupta, A. Prevalence of metabolic syndrome as per the NCEP and IDF definitions vis-a-vis severity and duration of psoriasis in a semi-urban Maharashtrian population: A case control study. *Diabetes Metab. Syndr.* **2016**, *10*, S72–S76. [CrossRef]
17. Craig, J.E.; Hewitt, A.W.; McMellon, A.E.; Henders, A.K.; Ma, L.; Wallace, L.; Sharma, S.; Burdon, K.P.; Visscher, P.M.; Montgomery, G.W.; et al. Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res.* **2009**, *19*, 2075–2080. [CrossRef]
18. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 731. [CrossRef]
19. Macgregor, S.; Zhao, Z.Z.; Henders, A.; Nicholas, M.G.; Montgomery, G.W.; Visscher, P.M. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Res.* **2008**, *36*, e35. [CrossRef]
20. Type 2 Diabetes Knowledge Portal. Available online: <https://t2d.hugeamp.org/> (accessed on 30 June 2022).
21. Harmonizome: Integrated Knowledge about Genes and Proteins. Available online: <https://maayanlab.cloud/Harmonizome/> (accessed on 30 June 2022).

22. Network Analyst—Comprehensive Gene Expression Profiling and Network Visual Analytics. Available online: <https://www.networkanalyst.ca/> (accessed on 30 June 2022).
23. Chand, S.; McKnight, A.J.; Shabir, S.; Chan, W.; McCaughan, J.A.; Maxwell, A.P.; Harper, L.; Borrows, R. Analysis of single nucleotide polymorphisms implicate mTOR signalling in the development of new-onset diabetes after transplantation. *BBA Clin.* **2016**, *5*, 41–45. [[CrossRef](#)]
24. Chen, Q.; Hao, W.; Xiao, C.; Wang, R.; Xu, X.; Lu, H.; Chen, W.; Deng, C.X. SIRT6 Is Essential for Adipocyte Differentiation by Regulating Mitotic Clonal Expansion. *Cell Rep.* **2017**, *18*, 3155–3166. [[CrossRef](#)]
25. Del Rio-Moreno, M.; Luque, R.M.; Rangel-Zuniga, O.A.; Alors-Perez, E.; Alcalá-Díaz, J.F.; Roncero-Ramos, I.; Camargo, A.; Gahete, M.D.; Lopez-Miranda, J.; Castano, J.P. Dietary Intervention Modulates the Expression of Splicing Machinery in Cardiovascular Patients at High Risk of Type 2 Diabetes Development: From the CORDIOPREV Study. *Nutrients* **2020**, *12*, 3528. [[CrossRef](#)] [[PubMed](#)]
26. Goo, Y.H.; Son, S.H.; Yechoor, V.K.; Paul, A. Transcriptional Profiling of Foam Cells Reveals Induction of Guanylate-Binding Proteins Following Western Diet Acceleration of Atherosclerosis in the Absence of Global Changes in Inflammation. *J. Am. Heart Assoc.* **2016**, *5*, e002663. [[CrossRef](#)] [[PubMed](#)]
27. Gurung, R.L.; FitzGerald, L.M.; Liu, E.; McComish, B.J.; Kaidonis, G.; Ridge, B.; Hewitt, A.W.; Vote, B.J.; Verma, N.; Craig, J.E.; et al. Identifying Genetic Biomarkers Predicting Response to Anti-Vascular Endothelial Growth Factor Injections in Diabetic Macular Edema. *Int. J. Mol. Sci.* **2022**, *23*, 4042. [[CrossRef](#)]
28. Hebbar, P.; Abubaker, J.A.; Abu-Farha, M.; Alsmadi, O.; Elkum, N.; Alkayal, F.; John, S.E.; Channanath, A.; Iqbal, R.; Pitkaniemi, J.; et al. Genome-wide landscape establishes novel association signals for metabolic traits in the Arab population. *Hum. Genet.* **2021**, *140*, 505–528. [[CrossRef](#)] [[PubMed](#)]
29. Hongbo, M.; Yanjiao, D.; Shuo, W.; Kun, S.; Yanjie, L.; Mengmeng, L. Podocyte RNF166 deficiency alleviates diabetic nephropathy by mitigating mitochondria impairment and apoptosis via regulation of CYLD signal. *Biochem. Biophys. Res. Commun.* **2021**, *545*, 46–53. [[CrossRef](#)]
30. Huang, Q.; Deng, G.; Wei, R.; Wang, Q.; Zou, D.; Wei, J. Comprehensive Identification of Key Genes Involved in Development of Diabetes Mellitus-Related Atherogenesis Using Weighted Gene Correlation Network Analysis. *Front. Cardiovasc. Med.* **2020**, *7*, 580573. [[CrossRef](#)]
31. Jee, D.; Kang, S.; Park, S. Association of Age-Related Cataract Risk with High Polygenetic Risk Scores Involved in Galactose-Related Metabolism and Dietary Interactions. *Lifestyle Genom.* **2022**, *15*, 55–66. [[CrossRef](#)]
32. Li, J.; Zhang, P.; Xia, Y. Study on CCDC69 interfering with the prognosis of patients with breast cancer through PPAR signal pathway. *Eur. J. Histochem. EJH* **2021**, *65*. [[CrossRef](#)]
33. Li, X.Q.; Huang, T.Y. Notoginsenoside R1 alleviates high glucose-induced inflammation and oxidative stress in HUVECs via upregulating miR-147a. *Kaohsiung J. Med. Sci.* **2021**, *37*, 1101–1112. [[CrossRef](#)]
34. Min, J.; Zeng, T.; Roux, M.; Lazar, D.; Chen, L.; Tudzarova, S. The Role of HIF1 α -PFKFB3 Pathway in Diabetic Retinopathy. *J. Clin. Endocrinol. Metab.* **2021**, *106*, 2505–2519. [[CrossRef](#)]
35. Ng, M.C.Y.; Graff, M.; Lu, Y.; Justice, A.E.; Mudgal, P.; Liu, C.T.; Young, K.; Yanek, L.R.; Feitosa, M.F.; Wojczynski, M.K.; et al. Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet.* **2017**, *13*, e1006719. [[CrossRef](#)] [[PubMed](#)]
36. Qu, H.Q.; Qu, J.; Vaccaro, C.; Chang, X.; Mentch, F.; Li, J.; Mafra, F.; Nguyen, K.; Gonzalez, M.; March, M.; et al. Genetic Analysis for Type 1 Diabetes Genes in Juvenile Dermatomyositis Unveils Genetic Disease Overlap. *Rheumatology* **2022**, keac100. [[CrossRef](#)] [[PubMed](#)]
37. Rodriguez-Perez, J.M.; Posadas-Sanchez, R.; Blachman-Braun, R.; Vargas-Alarcon, G.; Posadas-Romero, C.; Rodriguez-Cortes, A.A.; Lopez-Bautista, F.; Tovilla-Zarate, C.A.; Rojas-Toledo, E.X.; Borgonio-Cuadra, V.M.; et al. HHIPL-1 (rs2895811) gene polymorphism is associated with cardiovascular risk factors and cardiometabolic parameters in Mexican patients with myocardial infarction. *Gene* **2018**, *663*, 34–40. [[CrossRef](#)] [[PubMed](#)]
38. Sajuthi, S.P.; Sharma, N.K.; Comeau, M.E.; Chou, J.W.; Bowden, D.W.; Freedman, B.I.; Langefeld, C.D.; Parks, J.S.; Das, S.K. Genetic regulation of adipose tissue transcript expression is involved in modulating serum triglyceride and HDL-cholesterol. *Gene* **2017**, *632*, 50–58. [[CrossRef](#)]
39. Tang, G.Y.; Yu, P.; Zhang, C.; Deng, H.Y.; Lu, M.X.; Le, J.H. The Neuropeptide-Related HERC5/TAC1 Interactions May Be Associated with the Dysregulation of lncRNA GAS5 Expression in Gestational Diabetes Mellitus Exosomes. *Dis. Markers* **2022**, *2022*, 8075285. [[CrossRef](#)]
40. Vivot, K.; Mouille, V.S.; Zarrouki, B.; Tremblay, C.; Mancini, A.D.; Maachi, H.; Ghislain, J.; Poitout, V. The regulator of G-protein signaling RGS16 promotes insulin secretion and β -cell proliferation in rodent and human islets. *Mol. Metab.* **2016**, *5*, 988–996. [[CrossRef](#)]
41. Wang, L.; Lei, L.; Xu, T.; Wang, Y. GSTO1 regulates insulin biosynthesis in pancreatic β cells. *Biochem. Biophys. Res. Commun.* **2020**, *524*, 936–942. [[CrossRef](#)]
42. Chen, J.; Sun, M.; Adeyemo, A.; Pirie, F.; Carstensen, T.; Pomilla, C.; Doumatey, A.P.; Chen, G.; Young, E.H.; Sandhu, M.; et al. Genome-wide association study of type 2 diabetes in Africa. *Diabetologia* **2019**, *62*, 1204–1211. [[CrossRef](#)]

43. Tabassum, R.; Chauhan, G.; Dwivedi, O.P.; Mahajan, A.; Jaiswal, A.; Kaur, I.; Bandesh, K.; Singh, T.; Mathai, B.J.; Pandey, Y.; et al. Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* **2013**, *62*, 977–986. [[CrossRef](#)]
44. Saxena, R.; Saleheen, D.; Been, L.F.; Garavito, M.L.; Braun, T.; Bjornnes, A.; Young, R.; Ho, W.K.; Rasheed, A.; Frossard, P.; et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **2013**, *62*, 1746–1755. [[CrossRef](#)]
45. Venkatesan, V.; Lopez-Alvarenga, J.C.; Arya, R.; Ramu, D.; Koshy, T.; Ravichandran, U.; Ponnala, A.R.; Sharma, S.K.; Lodha, S.; Sharma, K.K.; et al. Burden of Type 2 Diabetes and Associated Cardiometabolic Traits and Their Heritability Estimates in Endogamous Ethnic Groups of India: Findings From the INDIGENIUS Consortium. *Front. Endocrinol.* **2022**, *13*, 847692. [[CrossRef](#)] [[PubMed](#)]
46. Shpakov, A.O.; Derkach, K.V.; Berstein, L.M. Brain signaling systems in the Type 2 diabetes and metabolic syndrome: Promising target to treat and prevent these diseases. *Future Sci. OA* **2015**, *1*, FSO25. [[CrossRef](#)] [[PubMed](#)]
47. Hao, K.; Di Narzo, A.F.; Ho, L.; Luo, W.; Li, S.; Chen, R.; Li, T.; Dubner, L.; Pasinetti, G.M. Shared genetic etiology underlying Alzheimer's disease and type 2 diabetes. *Mol. Asp. Med.* **2015**, *43–44*, 66–76. [[CrossRef](#)] [[PubMed](#)]
48. Rahimi, M.; Vinciguerra, M.; Daghighi, M.; Ozcan, B.; Akbarkhanzadeh, V.; Sheedfar, F.; Amini, M.; Mazza, T.; Paziienza, V.; Motazacker, M.M.; et al. Age-related obesity and type 2 diabetes dysregulate neuronal associated genes and proteins in humans. *Oncotarget* **2015**, *6*, 29818–29832. [[CrossRef](#)] [[PubMed](#)]
49. Santulli, G.; Pagano, G.; Sardu, C.; Xie, W.; Reiken, S.; D'Ascia, S.L.; Cannone, M.; Marziliano, N.; Trimarco, B.; Guise, T.A.; et al. Calcium release channel RyR2 regulates insulin release and glucose homeostasis. *J. Clin. Investig.* **2015**, *125*, 1968–1978. [[CrossRef](#)] [[PubMed](#)]
50. Bruno, E.; Manoukian, S.; Venturelli, E.; Oliverio, A.; Rovera, F.; Iula, G.; Morelli, D.; Peissel, B.; Azzolini, J.; Roveda, E.; et al. Adherence to Mediterranean Diet and Metabolic Syndrome in BRCA Mutation Carriers. *Integr. Cancer Ther.* **2018**, *17*, 153–160. [[CrossRef](#)]
51. Yin, Q.; Ni, Q.; Wang, Y.; Zhang, H.; Li, W.; Nie, A.; Wang, S.; Gu, Y.; Wang, Q.; Ning, G. Raptor determines β -cell identity and plasticity independent of hyperglycemia in mice. *Nat. Commun.* **2020**, *11*, 2538. [[CrossRef](#)]
52. Cras-Meneur, C.; Conlon, M.; Zhang, Y.; Pasca Di Magliano, M.; Bernal-Mizrachi, E. Early pancreatic islet fate and maturation is controlled through RBP-Jkappa. *Sci. Rep.* **2016**, *6*, 26874. [[CrossRef](#)]
53. Bonal, C.; Herrera, P.L. Understanding the extrinsic and intrinsic signals involved in pancreas and β -cell development: From endoderm to β cells. *Curr. Opin. Organ Transplant.* **2007**, *12*, 40–48. [[CrossRef](#)]
54. Goth, L.; Nagy, T.; Kosa, Z.; Fejes, Z.; Bhattoa, H.P.; Paragh, G.; Kaplar, M. Effects of rs769217 and rs1001179 polymorphisms of catalase gene on blood catalase, carbohydrate and lipid biomarkers in diabetes mellitus. *Free Radic. Res.* **2012**, *46*, 1249–1257. [[CrossRef](#)]
55. Flekac, M.; Skrha, J.; Hilgertova, J.; Lacinova, Z.; Jarolimkova, M. Gene polymorphisms of superoxide dismutases and catalase in diabetes mellitus. *BMC Med. Genet.* **2008**, *9*, 30. [[CrossRef](#)] [[PubMed](#)]
56. Musicant, A.M.; Parag-Sharma, K.; Gong, W.; Sengupta, M.; Chatterjee, A.; Henry, E.C.; Tsai, Y.H.; Hayward, M.C.; Sheth, S.; Betancourt, R.; et al. CRTC1/MAML2 directs a PGC-1alpha-IGF-1 circuit that confers vulnerability to PPARgamma inhibition. *Cell Rep.* **2021**, *34*, 108768. [[CrossRef](#)] [[PubMed](#)]
57. White, P.; May, C.L.; Lamounier, R.N.; Brestelli, J.E.; Kaestner, K.H. Defining pancreatic endocrine precursors and their descendants. *Diabetes* **2008**, *57*, 654–668. [[CrossRef](#)] [[PubMed](#)]
58. Macgregor, S. Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur. J. Hum. Genet.* **2007**, *15*, 501–504. [[CrossRef](#)] [[PubMed](#)]