DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Single-nucleotide polymorphism-based genetic risk estimate on Hungarian general and Roma population for type 2 diabetes mellitus

by Nardos Abebe Werissa

Supervisor: Róza Ádány, MD, PhD, DSc

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF HEALTH SCIENCES

DEBRECEN, 2021

Table of contents

List of abbreviations

ADRA2A - Adrenoceptor alpha 2a

BMI - Body mass index

C2CD4B - C2 calcium dependent domain containing 4B

CDKAL1 - CDK5 regulatory subunit associated protein 1-like 1

CDKN2A/B - Cyclin dependent kinase inhibitor 2A/2B

CVD - Cardiovascular disease

CI - Confidence interval

DGKB–TMEM195 - Diacylglycerol kinase beta

DNA - Deoxyribonucleic acid

EDTA - Ethylenediaminetetraacetic acid

EOT2DM - Early-onset of type 2 diabetes mellitus

EU - European Union

FADS1 - Fatty acid desaturase 1

FG - Fasting glucose

G6PC2 - Glucose-6-phosphatase catalytic subunit 2

GCKR - Glucokinase regulatory protein

GIPR - Gastric inhibitory polypeptide receptor

GLIS3 - GLIS family zinc finger 3

GP - General Practitioner

GPMSSP - General Practitioners' Morbidity Sentinel Stations Program

GRS - Genetic risk score

GWAS - Genome wide association study

HDL-C - High-density lipoprotein cholesterol

HHEX - Hematopoietically expressed homeobox protein

HuGE - Human genome

HWE - Hardy−Weinberg equilibrium

IDF - International Diabetes Federation

KCNJ11 - Potassium voltage-gated channel subfamily J member 11

LD - Linkage disequilibrium

MAF - Mutation Analysis Core Facility

MTNR1B - Melatonin receptor 1B

ORs - Odds ratios

PreDM - Prediabetes

PROX1 - Prospero homeobox protein 1

SNPs – Single-nucleotide polymorphisms

T2DM - Type 2 diabetes mellitus

TCF7L2 - Transcription factor 7-like 2

TG - Triglyceride

UOT2DM - Usual-onset of type 2 diabetes mellitus

wGRS - Genetic risk score (weighted)

WHO - World Health Organization

1. Introduction

Diabetes mellitus (DM), more commonly referred as “diabetes” is characterized by chronically elevated glucose concentration in the blood resulting from deficiency in production of insulin by pancreas or inadequate sensitivity of cells to the action of insulin. There are two major forms of DM: type 1 diabetes mellitus (T1DM) in which the pancreas fails to produce the insulin and type 2 diabetes mellitus (T2DM) which results from the body’s inability to adequately respond to insulin. T1DM is mostly seen in children and adolescents and the later form occurs most commonly in adults [1].

Although T2DM is generally considered as a disease of the older age groups, during the last decade a large number of publications have drawn attention to the emergence of diabetes at an increasingly early age. The review on these studies by Lascar et al. emphasizes that “the prevalence of T2DM in adolescents and young adults is dramatically increasing…. raising the possibility of a future public health catastrophe” [2]. Early onset of T2DM (EOT2DM) is defined as diagnosis at below 45 years of age [3] and countries – among them China [4] and the United Kingdom [5] - documented an increase in the incidence and consequently the prevalence of EOT2DM. Moreover, Japan [6] and Taiwan [7], reported that more than 50% of diabetes cases in children and adolescents are T2DM.

T2DM is among the four major non communicable diseases by the World Health Organization (WHO); globally, about one in 11 adults have diabetes mellitus, 90% of whom have T2DM [8]. T2DM and its complications contribute massively to the mortality and disability globally [9]. In a more recent study it was reported that in 2019 diabetes caused about 4.2 million adult deaths throughout the world and approximately 11.3% of all the global deaths are associated with it [10].

Diabetes, next to HIV/AIDS, has the second biggest negative effect on reducing health adjusted life expectancy across the globe [11]. Patients with T2DM, particularly with poor glycemic control, are highly prone to associated comorbidities including hypertension, dyslipidemia, nonalcoholic fatty liver disease, renal failure, microvascular and macrovascular complications [12-14]. T2DM and its complications significantly affect the quality of life and exert a major burden on individuals, economy, and the healthcare system [12, 15, 16]. In fact, its impact depends on the onset age of the disease, EOT2DM has a more aggressive disease phenotype, leading to premature development of complications [2, 3, 17, 18]. For instance, macrovascular and microvascular complications [19], cardiovascular disease and microalbuminuria [20] are more common among diabetes patients with EOT2DM compared with the usual onset of T2DM (UOT2DM) (defined as diagnosis at ≥45 years of age). In general, individuals with EOT2DM have significantly poorer metabolic profiles than individuals with UOT2DM [21].

T2DM results from complex interplay between multiple genes, epigenetic and environmental/lifestyle factors [22, 23].

Several lifestyle factors are known to play important role in the development of T2DM. These are physical inactivity, sedentary lifestyle, obesity, stress and depression, disturbed sleep, cigarette smoking, high alcohol intake and energy-dense diets [23-25]. Environmental toxins, noise, increased exposure to residential traffic, and fine airborne particulate matter may also contribute to the development of T2DM [24, 26].

There is a strong inheritable genetic connection with T2DM; the risk of developing T2DM is nearly two to three fold if a person has a single diabetic parent and five to six fold if both parents are diabetic compared to the risk of a person with non-diabetic parents [27]. Studies of twins suggest

that T2DM might be linked with genetics and the concordance rates were estimated to be 34%-76% for monozygotic and 16%-37% for dizygotic twins [28, 29]. In addition, few studies also uncovered the existence of genetic factors on the age of onset of T2DM [30-32]. 81% of children and young people with T2DM had a positive family history (70% with first-degree relatives: 17% both parents affected, 50% mother alone, 23% father alone, and 10% sibling alone affected; and 11% with second-degree relatives) in the United Kingdom [30].

Over the past two decades, genome-wide association studies (GWAS), candidate gene studies and linkage studies, not only have discovered more than 150 single nucleotide polymorphisms (SNPs) in different genes to play important role in the development of T2DM [33] but also, have uncovered several SNPs that influence the age of onset for T2DM [34-44]. In 1998, the first candidate gene, PPARG which encodes the nuclear receptor PPAR-ɤ was identified to reproducibly associated with T2DM in Finnish population [45]. CAPN10 gene on chromosome 10, which encodes calpain-like cysteine protease family, calpain-10 (CAPN10) was the first T2DM susceptibility gene to be identified in early 2000s through linkage studies [46]. In 2007, after the development of new genotyping technology, several novel gene variants (e.g., TCF7L2, MTNR1B, CDKAL1, HHEX, SLC30A8) were discovered to be associated with the development of T2DM. Their associations were confirmed and replicated in many GWAS studies on multiple populations [47-49] and also observed on ethnically diverse populations [50, 51]. The majority of these SNPs exert their effect on the disease risk through deficient insulin secretion and few of them through insulin resistance [52].

Almost all these discovered SNPs separately have modest effects on the risk of T2DM (odds ratios ≤1.4) [53]; thus, they are just one at a time cannot be informative for the estimation of risk of T2DM. Summarizing the effects of SNPs into genetic risk scores (GRSs, unweighted and

weighted) gives an opportunity to examine the combined effect of these genetic factors on an outcome [54]. Genetic risk score modelling at the population level provides an opportunity to assess the degree of genetic load between different population groups among them ethnicities and can shed light on how it varies across population groups. GRS modeling also helps to explore the effect of genetics on the age of onset for different diseases.

Currently, very limited number of studies are available to explore the genetic susceptibility of T2DM in populations with non-European origin [48, 55-58], and none of these studies were carried out on Roma population. In addition, so far, no study was assessed the impact of genetic factors on the age of onset for T2DM in Hungarian population. Knowing the genetic background of T2DM development among the two populations could help the identification of groups for interventions targeting T2DM prevention. It may also help the development of tools for the stratification and estimating the risk of earlier onset of T2DM on the Hungarian population.

1.1. Roma population and T2DM among them

With an estimated population of 10-12 million, the Roma are Europe's largest and the most vulnerable ethnic group [59]. Approximately six million of Roma live in the European Union [60]. Roma arrived in the Balkans from North India in the Xth century and then migrated to Europe in three migration waves [61]. Currently, this minority group is clustered in the Central and Eastern European countries, largely in Bulgaria, North Macedonia, Hungary, Slovakia and Romania [62]. Nowadays, Roma population are becoming the target population for ethnic-based studies, however, only a limited number of them have explored their genetic risk for different traits or phenotypes. A huge number of studies have demonstrated that the Roma suffer from poor health [63], unhealthy living conditions [64], low life expectancy [65], severely limited access to health

services [66, 67], and discrimination [68], which are closely linked to a low level of education, a high rate of unemployment, and their low socio-economic status in general [69].

Higher prevalence of prediabetes (PreDM) - defined as a fasting blood glucose level above the normal but below the diabetic threshold, i.e., between 5.6 and 6.9 mmol/L [70] - and T2DM was shown in a previous study which compared PreDM and T2DM between Hungarian Roma and Hungarian general population (27.09% vs. 15.56%; $p<0.001$) [71]. In other studies, the higher prevalence of T2DM among Roma compared to the general population (of Caucasian origin) in Serbia (11.1% vs. 6.7%) [72] and in Slovakia (30% vs. 10%; $p<0.001$) [73] was also reported. A 25% of higher prevalence of T2DM in Roma population compared to the Czech majority population was also reported by the government of the Czech Republic [58].

However, the latest review by Nunes et al. on publications related to the prevalence of diabetes mellitus in the Roma population [74] concludes that "none of the previous studies reached the standards regarding representative samples and number of cases for a conclusive result" the researchers suggested an increased prevalence of diabetes in Roma compared with the majority populations and the authors also raised a possible genetic risk to T2DM among Roma known to have Asian origin by accepting the theory of the increased genetic susceptibility to T2DM in different Asian (Japanese, Chinese and Indian) populations [74, 75].

Based on the shorter life expectancy and the higher prevalence of metabolic syndrome among Roma, Simko et al created the so called "thrifty genes'' theory supposing that during the course of many generation long migration from India to Europe, they suffered with food insufficiency and in order to withstand this deficiency they might have developed adaptive metabolic and genetic changes [76]. After their arrival to Europe, the somewhat better food accessibility together with

abruptly reduced physical activity has resulted in the development of metabolic syndrome and consequently increased T2DM and cardiovascular mortality. This hypothesis is supported by findings showing the significantly higher prevalence of metabolic syndrome [71, 77], as well as increased CVD risk [78-81] and significantly higher mortality [82, 83] among Roma. In addition the higher prevalence of T2DM [58] and genetically modified disturbances in other cardiometabolic traits [84-86] were also detected in the Roma populations in Europe.

2. Aims

The aims of our study were:

1. To investigate whether higher prevalence of PreDM and T2DM among Roma is due to inheritable and/or other factors.
2. To compare the risk allele frequencies between the Roma and Hungarian general populations.
3. To estimate and compare the risk allele load in the Roma and Hungarian general populations using the GRS modelling approach based on 16 SNPs related to T2DM.
4. To evaluate the joint effect of T2DM associated 23 SNPs using GRS on the age of onset for T2DM in the Hungarian population.

3. Materials and methods

All the data used in this dissertation are from previously created databases.

3.1. Study design

The current study consists of data assembled from previous three surveys involving 1168 individuals representative of Hungarian T2DM population (case population) [87, 88], 1783 individuals representative of Hungarian general population [89, 90] and 1260 individuals representative of Roma living in segregated colonies in North-East Hungary, where they mainly concentrated [90, 91]. The study flowchart is shown in Figure 1.

3.2. Samples

3.2.1. Sample representative for Hungarian T2DM population

The study subjects as T2DM population were obtained from a survey (Survey 1) based on the framework of General Practitioners' Morbidity Sentinel Stations Program (GPMSSP) in 2005. GPMSSP was established in 1998 jointly by the School of Public Health in the University of Debrecen and the National Public Health and Medical Officer Service to monitor the prevalence and incidence of chronic non communicable diseases of high public health importance in Hungary [87]. The source population consisted of 138,088 persons registered in the GPMSSP framework and the case population (n=1324) was randomly selected from 15,944 T2DM patients registered by the seventy-two participating general practitioners (GPs). A total of 1168 (response rate of 88.2%) representative of Hungarian T2DM patients were included in this survey [87, 88]. Physical examinations (weight, height, waist circumference, and blood pressure) were carried out by the GPs; and blood samples (native and EDTA-anticoagulated) for laboratory investigation (fasting glucose, HDL-C and triglyceride) and DNA isolation were collected by GPs as well. Information

on sociodemographic characteristics and self-assessed health status were obtained using a self-administered questionnaire [87]. Within this program a total of 1168 DNA samples were obtained.

The sample of T2DM case population was categorized into 3 groups based on the age of onset for T2DM:

1. ≤49 years, n=191
2. 50-59 years, n=340
3. ≥60 years, n=350

3.2.2. Sample representative for Hungarian general population

A cross-sectional study (Survey 2) based on the framework GPMSSP was carried out to estimate the prevalence of metabolic syndrome among Hungarians in 2006 [89]. The source population of this study consisted of all individuals aged 20-69 years, registered by fifty-nine participating GPs from eight counties. 1999 participants were selected randomly from the file of the residents of the catchment area. From this survey 1783 participants (91% response rate; 36 participants were excluded due to lacking blood sample or questionnaire-based data) with full record and DNA samples were involved in our study. The selected sample is representative for the Hungarian adult population aged 20-69 years in terms of geographic, age and sex distribution. GPs recorded relevant medical history, performed physical examination such as weight, height, waist circumference and blood pressure measurements; and collected venous blood samples (native and EDTA-anticoagulated) for laboratory measurements (fasting glucose, HDL-cholesterol and triglyceride) and genotype investigations.

The samples of the Hungarian general population were divided into 3 subpopulations based on the proposal of the experts committee on diagnosis and classification of diabetes mellitus [92].

The three subpopulations were:

1. Subjects with normal FG level: FG<5.6 mmol/L, n=1197
2. Prediabetic subjects: FG between 5.6 and 6.9 mmol/L, n=108
3. T2DM patients: any person who had FG level of 7 mmol/L or higher and/or was under antidiabetic treatment, n=110

The sample of Hungarian general population further categorized in to 5 groups based on the GRS values:

- GRS<4, n=91
- GRS =4, n=286
- GRS=6, n=469
- GRS=8, n=379
- GRS>8, n=190

3.2.3. Sample representative for Roma population

Using stratified multistep sampling technique, participants were selected from two counties (Hajdú-Bihar and Szabolcs-Szatmár-Bereg) of North-East Hungary, where majority of Roma colonies are accumulated. Segregated colonies with more than 100 inhabitants were considered as the study base, resulting in 64 eligible colonies (Survey 2 in 2015). From these colonies, 40 colonies (25 from Hajdú-Bihar county and 15 from Szabolcs-Szatmár-Bereg county) were randomly selected. First, using GPs' validated household lists, 25 households were randomly chosen from each colony. Then, adults 20-64 years were identified, and one person was selected by random table from each household. From the 25GPs, only 22 GPs (3 GPs refused to participate) in Hajdú-Bihar county (22X25 persons) and each of the invited 15 GPs in Szabolcs-Szatmár-Bereg county (15X25 persons) became involved, thus the final sample consisted of 925 people [90]. From

the 925 people, 725 individuals were committed to participate in the study (response rate 78.4%). As part of the health survey, interviewer-assisted questionnaires were used to collect data on sociodemographic factors, and self-assessed health status. Medical histories were recorded by general practitioners, and each participant went through a physical examination (weight, height, waist circumference, blood pressure measurements). Venous blood samples (native and EDTA-anticoagulated) were taken for laboratory analysis (glucose, triglyceride, HDL-cholesterol levels) and genotype investigations [90].

Additional samples were obtained in the framework (Survey 3) of the Public Health Focused Model Program for Organizing Primary Care Services in 2013 [91, 93]. This program aimed at reducing social inequalities in health through primary healthcare reform. The program encompassed the two most disadvantaged regions of Hungary, i.e., Northern Hungary and the Northern Great Plain. In these regions, 4 primary care clusters (totally involving 24 GPs' practices) were established in Hajdú-Bihar, Borsod-Abaúj-Zemplén, Jász-Nagykun-Szolnok and Heves counties. The sampling method of the study participants was quite similar to that of explained above [90]. Within this framework further 535 samples from the Roma population dwelling in North-East of Hungary were collected, totally making the Roma sample 1260.

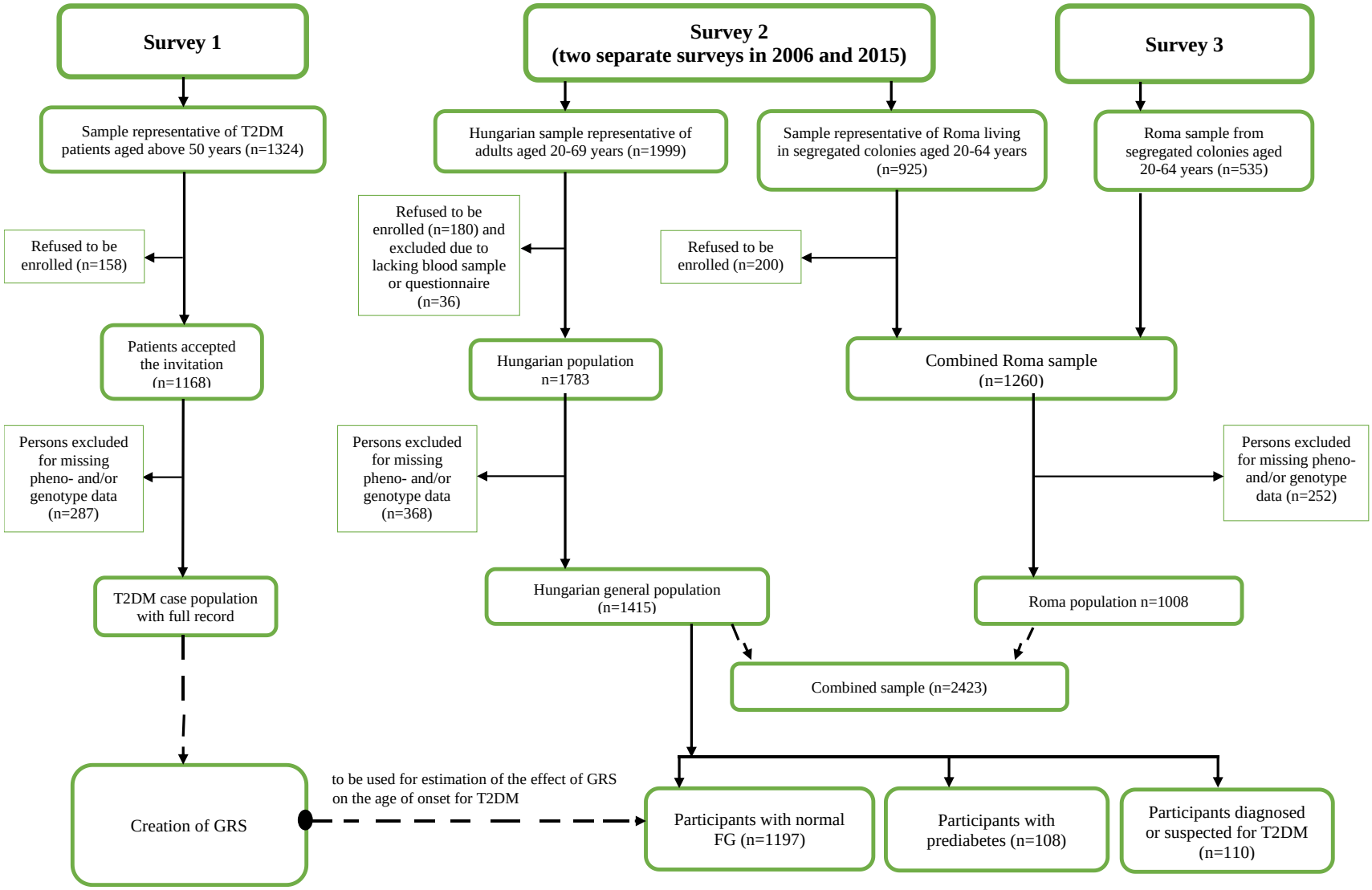


Figure 1. Flowchart showing the processes of sample selection, stratification, and creation of subgroups of the Hungarian general population.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national ethical committees and with the 1964 Helsinki declaration and its later amendments. The above-described studies were approved by the Ethical Committee of the University of Debrecen, Medical Health Sciences Centre (reference No. 2462-2006 and 2699-2007) and by the Ethical Committee of the Hungarian Scientific Council on Health (reference Nos. NKFP/1/0003/2005; 8907-O/2011-EKU and TUKEB 48495-2/2014/EKU).

3.3. DNA extraction

DNA was isolated from EDTA-anticoagulated blood samples using a MagNA Pure LC system (Roche Diagnostics, Basel, Switzerland) with a MagNA Pure LC DNA Isolation Kit–Large Volume according to the manufacturer's instructions. DNA extraction was done by the technical assistant Zsuzsa Edit Tóth.

3.4. SNP selection

A systematic literature search using online databases (PubMed, HuGE Navigator and Ensembl) was conducted to identify the SNPs that were found to be associated with T2DM. During the SNP selection process, previously published meta-analysis results (reported as odds ratios) were considered to be of high priority (for comparing risk allele load between Roma and Hungarian general population) and additional SNPs that are associated with FG level were identified for evaluation of the effect of genetic factors on the age of onset for T2DM (see the list of SNPs selected with references in Table 1). The SNP selections was done by me (Nardos Abebe Werissa)

3.5. Genotyping

The search resulted in the identification of 23 SNPs, (of which16 SNPs with meta-analysis odds ratio results) that were genotyped by the service provider (Mutation Analysis Core Facility (MAF) of the Karolinska University Hospital, Sweden). Genotyping was performed on a MassARRAY

platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry. Validation, concordance analysis and quality control were conducted by the MAF, according to their protocols.

MassARRAY SNP Genotyping combines the benefits of a simple and accurate primer extension chemistry of the iPLEX assay with state-of-the-art matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry to quickly and cost effectively characterize genotypes with the highest levels of reproducibility (>99% call rates with >99.7% accuracy on validated assay)[94].

Table 1. List of susceptible SNPs considered in the genetic risk score computation with their genes, effect alleles and their effect on T2DM

SNP	Gene	Effect on T2DM	Effect allele	Average effect size (OR)	References
SNPs identified with meta-analysis odds ratio results					
rs7903146	TCF7L2	Transcription factor 7-like 2 encodes a high mobility group box-containing transcription factor which involved in Wnt signaling pathway[95]. This Wnt signaling pathway plays significant role in the islet cell proliferation and differentiation in the pancreas. The rs7903146 is associated with impaired beta-cell function [96] and reduced insulin secretion [97] but not with insulin resistance and enhance the rate of hepatic glucose production [98].The rs7903146 polymorphism subvert the Wnt signaling pathway and impedes the insulin secretion, and finally ends up with progression of T2DM [99, 100].	T	1.40	[49, 101-106]
rs10811661	CDKN2A/B	Cyclin dependent kinase inhibitor 2A/2B inhibits the activity of CDK4 and CDK6 which involved in the pancreatic beta cell function and regeneration. The rs10811661 polymorphism influences pancreatic beta cell proliferation in pancreatic islets and mass, and further results in the development of diabetes [107].	T	1.22	[49, 106, 108-111]
rs10946398	CDKAL1	CDK5 regulatory subunit associated protein 1 like 1 encodes a protein that inhibits the activation of cyclin-dependent kinase 5 (CDK5). CDK5 play a role in the loss of beta cell function in the pancreatic beta cell. Genetic defect in CDKAL1 affect the development of T2DM [112, 113]	C	1.18	[108, 114, 115]
rs1111875	HHEX	Hematopoietically expressed homeobox protein encodes a member of the homeobox family of transcription factors involved in Wnt signaling pathway. Variants in this gene could also play in insulin degradation or insulin sensitivity and beta cell dysfunction [116, 117]	C	1.16	[118, 119]
rs5219	KCNJ11	Potassium inwardly rectifying channel, subfamily J, member 11 encodes the subunit protein of KATP (Kir6.2) and is highly expressed in the pancreas. Mutation in the KCNJ11 E23K gene affects sensitivity of the ion channel to ATP and makes the channel consume more ATP. Finally, insulin release is damaged and the and increase the risk of T2DM [120]	T	1.13	[121-124]
rs11671664	GIPR	Gastric inhibitory polypeptide receptor is expressed in pancreatic islets and adipocytes, and linked with insulin resistance and T2DM, stimulation of glucose-stimulated insulin secretion, modulation of beta	A	1.10	[49]

		cell neogenesis and pancreatic beta cell differentiation and proliferation [125] and it is thought that mutation in the rs11671664 may results in T2DM.			
rs780094	GCKR	Glucokinase regulatory protein regulates the glycolytic enzyme (glucokinase) and polymorphism in this variant leads to T2DM [108]	C	1.08	[49, 104, 108, 126]
rs1387153	MTNR1B	Melatonin receptor 1B encodes the melatonin receptor MT2, a G protein-coupled receptor, which is expressed in pancreatic islets [127]. The MTNR1B polymorphism is associated with higher fasting glucose levels and lower dynamic beta cell response [128] and increased the risk of isolated impaired fasting glycaemia but not isolated impaired glucose tolerance [129].	T	1.07	[130, 131]
rs10830963			G	1.06	[49, 104, 126, 130]
rs340874	PROX1	Prospero homeobox protein 1 is a transcription factor that plays a key regulatory role in neurogenesis and embryonic development of the pancreas. Polymorphism of the gene affect the beta cell development and leads to the development of T2DM [132, 133].	C	1.07	[104]
rs2191349	DGKB–TMEM195	Diacylglycerol kinase beta DGKB encodes the β isotype of the catalytic domain of diacylglycerol kinase, which regulates the intracellular concentration of the second messenger diacylglycerol. In pancreatic islets diacylglycerol activates protein kinase C (PKC) and thus potentiates insulin secretion. mutation of this variant leads to the development of T2DM [134]	T	1.06	[104]
rs174550	FADS1	Fatty acid desaturase 1 encodes rate limiting enzyme known as delta-5 desaturase (D5D). D5D is responsible for the double bond formation in the n-3 poly unsaturated fatty acid (PUFA) pathway and is linked with fatty acid composition in plasma, adipose tissue and membrane fluidity [135]. Mutation rs174550 mediate the development of T2DM by impairing insulin sensitivity [136, 137].	T	1.04	[104]
rs10885122	ADRA2A	Adrenoceptor alpha 2A encodes the alpha2A-adrenergic receptor (alpha(2A)AR), a Gi-coupled receptor expressed in pancreatic beta cells and whose activation leads to an outward potassium current independent of the islet ATP-sensitive potassium channel. By this way they modify the release of insulin. The rs10885122 polymorphism mediates adrenergic suppression of insulin secretion, and in turn increase the development of T2DM [138].	G	1.04	[104, 111]
rs11071657	C2CD4B	C2 calcium dependent domain containing 4B expressed in the pancreatic beta cells and regulates insulin release or beta cell function. polymorphism in the gene variant exerts an impact on the development of T2DM [134]	A	1.03	[104]
rs7034200	GLIS3	GLIS family zinc finger 3 plays a key role in controlling insulin gene transcription, insulin secretion and pancreatic beta cell survival. The	A	1.03	[104]

		rs7034200 is associated with fasting glucose and impaired β cell function [139, 140] and associated with reduced glucose-stimulated β cell function [141].			
rs560887	G6PC2	Glucose-6-phosphatase catalytic subunit 2 encodes the enzyme islet-specific glucose-6-phosphatase catalytic subunit related protein (IGRP) that takes part in the counter player to glucokinase by dephosphorylating glucose-6-phosphate and ends up with glucose stimulated insulin secretion, thus mutation in the rs560887 leads to the development of T2DM [142].	T	1.03	[104]
Additional SNPs identified that were associated with FG level					
rs11920090	SLC2A2	SLC2A2 encodes GLUT2, a glucose transporter and a member of the facilitative glucose transporter family, is highly expressed in pancreatic beta cells and liver. GLUT2 is involved in the regulation of both glucose uptake and output. SLC2A2 polymorphism may probably influence basal insulin secretion and mediates the progression of T2DM [142].	T	-	-
rs11558471			G	-	-
rs7944584	MADD	The biological function of MADD is linked with pancreatic beta cell proliferation and development [143]. It encodes mitogen-activated protein kinase (MAPK) activating death domain, an adaptor protein that interacts with the tumor necrosis factor alpha receptor to activate MAPK. MAPK is believed to be involved in the proliferation of pancreatic beta cells and insinuating that MADD polymorphism plays crucial progression of T2DM through beta cell dysfunctions [134].	A	-	-
rs10838687			T	-	-
rs3736594	MRPL33	Mitochondrial ribosomal protein L33 gene encodes a large mitoribosomal subunit protein, which may be involved in mitochondrial translation. The rs3736594 associated with fasting glucose and insulin levels [144].	C	-	-
rs7173964	C2CD4B	C2 calcium dependent domain containing 4B expressed in the pancreatic beta cells and regulates insulin release or beta cell function. Polymorphism in the gene variant exerts an impact on the development of T2DM [134]	G	-	-
rs10906115	CDC123/CAMK1D	Cell division-cycle 123 (CDC123/CAMK1D) encodes a protein involved in cell cycle regulation and nutritional control of gene transcription [145]; however its role in the development of T2DM is still unclear.	G	-	-

3.6. Power calculation for SNPs for the Roma and Hungarian general populations

The statistical power calculations were based on the average effect sizes obtained from meta-analyses, assuming an alpha-level of 0.05 and a given sample size. In the estimation, we applied the allele frequencies for Utah Residents (CEPH) with Northern and Western Ancestry (CEU) and for GIH (Gujarati Indian from Houston, Texas) populations from the 1000 genome project, phase 3 considering that the Roma population of Europe had arrived to the Western Balkans from North India and then migrated to Europe. Power calculation was carried out by Peter Piko

3.7. Statistical analysis

A χ^2 test was used to assess whether the agreement of frequencies of genotypes for SNPs with Hardy-Weinberg equilibrium (HWE) expectations (by Plink software [146]). Linkage disequilibrium (LD) between polymorphisms was tested by Haploview software (version 4.2). The r^2 values were defined and visualized using standard D'/LOD color scheme. In the presence of LD blocks, one SNP (the second SNP from each LD block) was selected to avoid multicollinearity. Power calculations were performed by the software package Quanto 1.2.4 [147]. The normality of data for quantitative variables was tested using the Shapiro-Wilk test; and when it was necessary, non-normal variables were transformed using Templeton's two-step approach [148]. Two-tailed Student's t-tests were used to assess the statistical difference of variables among the groups. Associations between GRSs and FG levels (as continuous variable) and Prediabetes or T2DM status (as binary variable, hereafter referred to as T2DM status) were investigated by multiple regression models (adjusted by age, sex, BMI, TG, HDL-C and ethnicity as covariates) in separate and in combined study populations, as well. In addition, multiple linear regression analyses were used to estimate the individual and combined (GRS) effect of SNPs on the early onset of T2DM by adjusting for age, sex, BMI and TG/HDL-C ratio. Jonckheere-Terpstra trend test was [149]

used to analyze the statistically significant trend between the ordinal independent variable and continuous or ordinal dependent variables for the age of onset for T2DM.

IBM SPSS statistics for Windows (version 26, IBM Company, Armonk, NY, USA), STATA statistical software (version 12) and SNPStats online tool was used to carry out regression analyses. The Bonferroni correction was applied when several statistical tests were being performed simultaneously ($p<0.0042$). Statistical analysis was done by Nardos Abebe Werissa

3.8. Calculation and computation of GRS and wGRS values

To examine the cumulative effect of selected SNPs, unweighted (GRS) and weighted (wGRS) genetic risk scores were computed and compared in study populations. Individuals with any missing genotype or phenotype data were excluded from the calculation.

In the GRS, each person was assigned a score based on the number of risk alleles carried. Thus, risk allele homozygotes were coded as genotype "2", heterozygotes as genotype "1", and "0" indicated absence of the risk allele [150]. By using these codes, a simple count score (unweighted) was calculated as described by equation (1) in which Gi is the number of the risk alleles for the i^{th} SNP. This model sums up all risk alleles over all loci as a summary score assuming that all alleles have the same effect in size and direction:

$$GRS = \sum_{i=1}^{I} Gi \quad (1)$$

In the weighted approach, rather than giving equal weight to each SNP, SNPs with larger effects contributed more to the score. The calculation of the wGRS is described by equation (2). In this weighted score, average weights (wβ_i) were derived from the risk coefficient for each allele based on relative effect size determined previously in studies. These average weights (wβ_i) were

multiplied by 0, 1 or 2 according to the number of effect alleles carried by each person (Xi) [150, 151]

$$wGRS = \sum_{i=1}^{I} w\beta\_iXi \quad (2)$$

The average effect size estimate for wGRS calculation was computed by meta-analyses under the random-effects model using OpenMetaAnalyst software [152].

3.9. Determination of the best fitted genetic model for the age of onset for T2DM

For each SNP we have tested which of the genetic model of inheritance (codominant, dominant, and recessive) shows the strongest correlation with the outcome (age of onset for T2DM) in the case population. Adjusted (by age, sex, TG/HDL-C ratio) regression analyses were applied to test the association of SNPs individually with the age of onset for T2DM by SNPStats online tool (http://bioinfo.iconcologia.net/SNPstats). The Akaike information criterion (AIC), Bayesian information criterion (BIC), and p value were used to find the best fitting genetic model of inheritance under the selection process [153].

3.10. Calculation and optimization of the GRS model

Based on the result of best-fitting genetic model of inheritance, SNPs were coded according to the criteria of the model as follows:

In case of the codominant genetic model:

- homozygote genes with two risk alleles were counted as ‘’2’’, while heterozygote genotypes as ‘’1’’ and homozygote non-risk genes as ‘’0’’.

In case of the dominant genetic model:

- homo- and heterozygote genes with two or one risk alleles were counted as ‘’2’’, while homozygote non-risk genes as ‘’0’’.

In case of the recessive genetic model:

- homozygote genes with two risk alleles were counted as ‘’2’’, while heterozygote genotypes with one risk allele and homozygotes without risk allele as ‘’0’’.

Subsequently, the number of risk effects (2, 1, or 0) was summed using equation 1, where Gi is the number of risk effects in the respective locus.

During the optimization of the GRS model, SNPs which do not reinforce the association of the model with the outcome variable were excluded. To avoid the possibility of false-positive association, SNPs were tested in an ascending order of p value (from the strongest association to the weakest one). Starting with the SNP with the lowest p value, we inserted them one by one in GRS model, the association with the age of onset was tested after each inserted SNP. For each step, the number of risk alleles for the SNP inserted was added to the GRS. Regression analysis was applied to monitor the changes in the strength of the association. The SNPs were selected for the final GRS model only if they increased the r^2 value and decreased the p value in the model. Under the optimization process, all calculation was adjusted by BMI, TG/HDL-C ratio, sex, and duration of T2DM.

3.11. Estimation of the effect of genetic (GRS) and non-genetic (sex, BMI, and TG/HDL-C ratio) factors on the age of onset for T2DM on the case population

Linear regression was used to estimate the effect of GRS and non-genetic factors (sex, BMI, and TG/HDL ratio) on the age of onset for T2DM on the case population. The results of this calculation were used to determine the weighted GRS as well as to construct a risk estimation model for the age of onset for T2DM on the Hungarian general population.

3.12. wGRS calculation for the age of onset for T2DM

wGRS calculation was performed on the Hungarian general population by using the beta values determined on the case population for weighting (wβ_i). Then the GRS for each person (xi) was multiplied by the weight (wβ_i). Equation (3) describes the calculation of the wGRS.

$$wGRS = \sum_{i=1}^{I} w\beta\_iXi \quad (3)$$

3.13. Calculation of a score for an estimated age of onset for T2DM

The weight of genetic and non-genetic factors was determined on the case population. Using these weights, it is possible to calculate a score to estimate the age of onset for T2DM. To investigate the combined effect of non-genetic (sex, BMI, and TG/HDL-C ratio) and genetic (GRSs) factors with a reasonable impact on the development of T2DM, a score was calculated for each sample. The effect of non-genetic and genetic factors on the age of onset for T2DM was estimated on the case population and it was tested on the Hungarian general one.

4. Results

4.1. Characteristics of the study populations

Samples without full geno- and phenotype data were excluded from the analyses. In total, 881 individuals from the case, 1415 from the Hungarian general population and 1008 individuals from Roma population were included (Table 2). The population characteristics of the case population were similar to the prediabetic and T2DM subpopulations of the Hungarian general and significantly differed from the subpopulation with a normal FG level. A statistically significant increase was observed in the proportion of males, and in the average age, BMI, TG level, and TG/HDL-C ratio by subgroups ranging from normal FG level through prediabetes to T2DM cases in the Hungarian general population, while HDL-C level showed significant decrease on the same path by subpopulations. The age and sex differences between the case and the Hungarian subpopulations are partly due to the age category (20–69 years) applied in the sample collection of the Hungarian general population.

Table 2. Characteristics of the T2DM case, Hungarian general subpopulations, Hungarian general full population, and Roma full population.

	T2DM Case population (n =881)	Hungarian general subpopulations				Full Hungarian general population (n=1415)	Full Roma population (n=1008)	p value
		Normal FG level (FG<5.6mmol/L, n=1197)	Prediabetes (FG: 5.6–6.9mmol/L, n=108)	T2DM (FG≥7mmol/L and/or treated, n=110)	p for trend			
Male in % (95%CI)[a]	49.3 (46.0–52.6)	44.0 * (41.2–46.9)	64.8 * (55.5–73.3)	65.5 * (56.3–73.8)	<0.001	47.2 (44.6–49.9)	39.1 (36.2–42.3)	<0.001
Female in % (95%CI)[a]	50.7 (47.4–54.0)	56.0 * (53.1–58.8)	35.2 * (26.7–44.5)	34.5 * (26.2–43.7)		52.8 (50.1–55.4)	60.9 (57.7–63.8)	
Age in years (95%CI)[a]	66.14 (65.53–66.74)	42.68 ** (41.99–43.37)	50.57 ** (48.77–52.38)	54.33 ** (52.87–55.79)	<0.001	44.17 (32.07–56.27)	40.03 (27.61–52.45)	<0.001
BMI in kg/m2 (95%CI)	31.33 (30.97–31.69)	26.81 ** (26.53–27.10)	30.40 (29.17–31.64)	31.06 (29.95–32.17)	<0.001	27.43 (22.04–32.82)	26.76 (17.3–36.22)	0.024
HDL-C in (mmol/L (95%CI)	1.26 (1.24–1.29)	1.45 ** (1.42–1.47)	1.38 * (1.27–1.48)	1.19 (1.12–1.27)	<0.001	1.42 (1.40–1.45)	1.31 (1.28–1.34)	<0.001
TG in mmol/L (95%CI)	2.52 (2.36–2.69)	1.47 ** (1.41–1.53)	2.01 (1.64–2.39)	3.13 (2.36–3.89)	<0.001	1.64 (1.55–1.73)	1.63 (1.54–1.72)	0.864
TG/HDL-C ratio (95%CI)	2.34 (2.12–2.56)	1.24 ** (1.16–1.33)	2.12 (1.37–2.87)	3.54 (2.34–4.75)	<0.001	1.74 (1.45–2.04)	1.31 (1.28–1.34)	0.857
Elevated FG level (>5.6 mmol/L) and/or diabetes treatment (%)	-	-	-	-	-	15.16	22.87	<0.001
Age at diagnosis of T2DM in years (95%CI)	57.25 (56.6–57.9)	-	-	unknown	-	-	-	-

Notes: [a] The age and sex differences between the case population and the Hungarian subpopulations are due to the age category (20–69 years) applied in the sample collection of the Hungarian general population. *statistically significant ($p<0.05$) difference from T2DM case population. ** Significant ($p<0.0042$) difference from T2DM case population after Bonferroni correction.

4.2. Results of power calculations for the Hungarian general and Roma populations

The statistical power for individual SNPs was between 5.03% and 12.79% (Table 3).

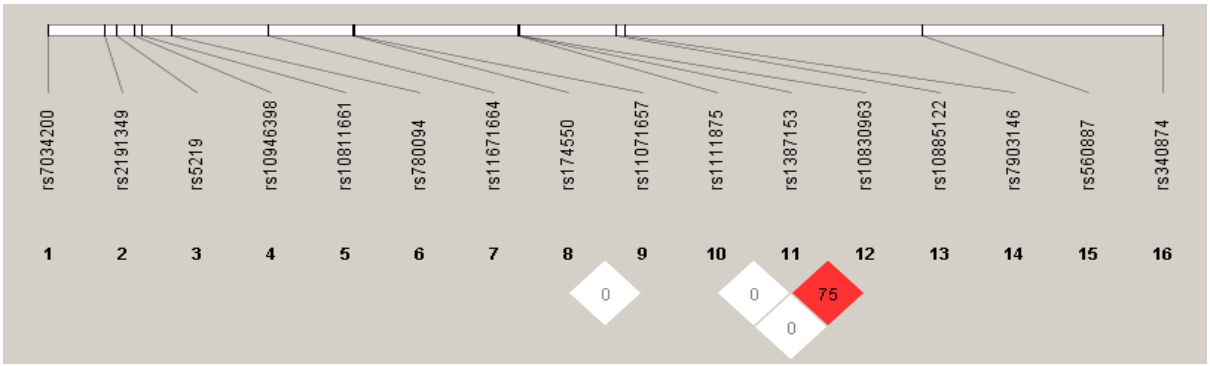
Table 3. Statistical power of the susceptible alleles considered separately for study populations

SNP	Gene	Effect allele	Power for Hungarian general population	Power for Roma population
rs7903146	TCF7L2	T	12.79%	9.92%
rs10811661	CDKN2A/B	T	5.71%	5.45%
rs10946398	CDKAL1	C	5.63%	5.40%
rs1111875	HHEX	C	5.27%	5.19%
rs5219	KCNJ11	T	7.92%	7.00%
rs11671664	GIPR	A	5.16%	5.09%
rs780094	GCKR	C	5.37%	5.23%
rs1387153	MTNR1B	T	6.54%	6.23%
rs340874	PROX1	C	5.37%	5.25%
rs10830963	MTNR1B	C	6.77%	6.37%
rs2191349	DGKB–TMEM195	T	5.26%	5.19%
rs174550	FADS1	T	5.10%	5.07%
rs10885122	ADRA2A	G	5.06%	5.04%
rs11071657	C2CD4B	A	5.06%	5.05%
rs7034200	GLIS3	A	5.07%	5.05%
rs560887	G6PC2	T	5.03%	5.03%

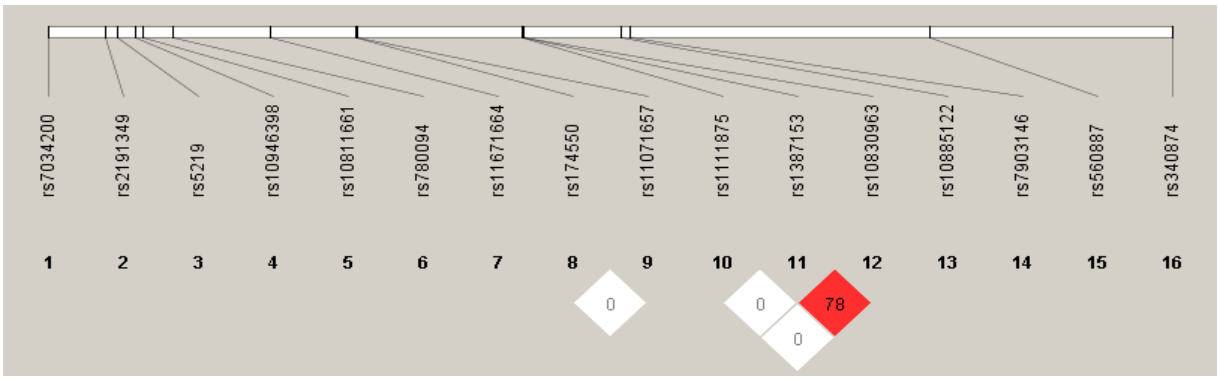
4.3. Results of the Hardy-Weinberg equilibrium and Linkage disequilibrium analyses in the case, the Hungarian general and Roma populations

In the case of the observed genotype distributions, no significant deviation from HWE was found in the populations. Two blocks were identified within linkage disequilibrium (LD) (Block 1: rs10838687 and rs7944584; Block 2: rs1387153 and rs10830963) in the case population and no LD block was identified in Hungarian general and Roma populations (Figure 2). To avoid multicollinearity, only one SNP per LD block was used in the GRS calculation.

A



B



C

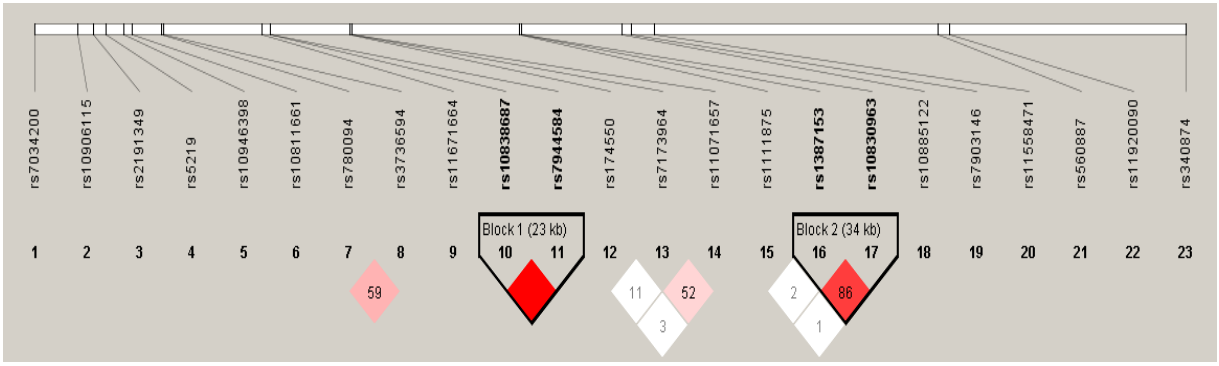


Figure 2. Linkage disequilibrium map of SNPs identified in the Hungarian general (A), Roma (B) and T2DM case (C) populations.

4.4. Comparison of allele frequencies in the Hungarian general and Roma populations

Allele frequencies calculated on the basis of genotype distributions obtained in the study populations, are shown in Table 4. Differences between the Roma and Hungarian general populations were significant for eight SNPs. Five susceptible alleles (rs7903146, rs1167664, rs340874, rs11071657, rs10946398) were more prevalent in the Hungarian general population and three (rs1387153, rs780094, rs10830963) among Roma.

Table 4. Comparison of susceptible allele frequencies between study populations

SNP	Gene	Effect allele	Allele frequency		
			Hungarian general population	Roma population	P value
rs7903146	TCF7L2	T	29.45%	24.21%	**0.004**
rs10811661	CDKN2A/B	T	82.78%	85.27%	0.096
rs10946398	CDKAL1	C	31.36%	25.45%	**0.002**
rs1111875	HHEX	C	58.82%	60.62%	0.377
rs5219	KCNJ11	T	35.58%	32.34%	0.099
rs11671664	GIPR	A	11.41%	8.53%	**0.022**
rs780094	GCKR	C	73.64%	79.09%	**0.002**
rs1387153	MTNR1B	T	29.38%	35.42%	**0.002**
rs340874	PROX1	C	47.60%	37.45%	**<0.001**
rs10830963	MTNR1B	G	29.05%	33.23%	**0.029**
rs2191349	DGKB–TMEM195	T	57.93%	58.88%	0.636
rs174550	FADS1	T	70.37%	71.63%	0.504
rs10885122	ADRA2A	G	88.67%	89.29%	0.633
rs11071657	C2CD4B	A	64.06%	59.29%	**0.030**
rs7034200	GLIS3	A	44.62%	47.87%	0.110
rs560887	G6PC2	T	86.93%	86.34%	0.652

Note: A p value in bold indicates at least a nominally significant difference in allele frequency between the study populations. An allele frequency shaded in gray is the higher allele frequency value.

4.5. Comparison of GRS and wGRS distribution

The GRS calculated for Roma subjects ranged from 6 to 24, and that for individuals of the Hungarian general population ranged from 7 to 24. The mean of the GRS was 14.8±2.68 in the Roma and 15.38±2.70 in the Hungarian general population sample. The distribution of the GRS in the two study groups was found to be significantly different ($p<0.001$), being right shifted in the Hungarian general population relative to the Roma (Figure 3.).

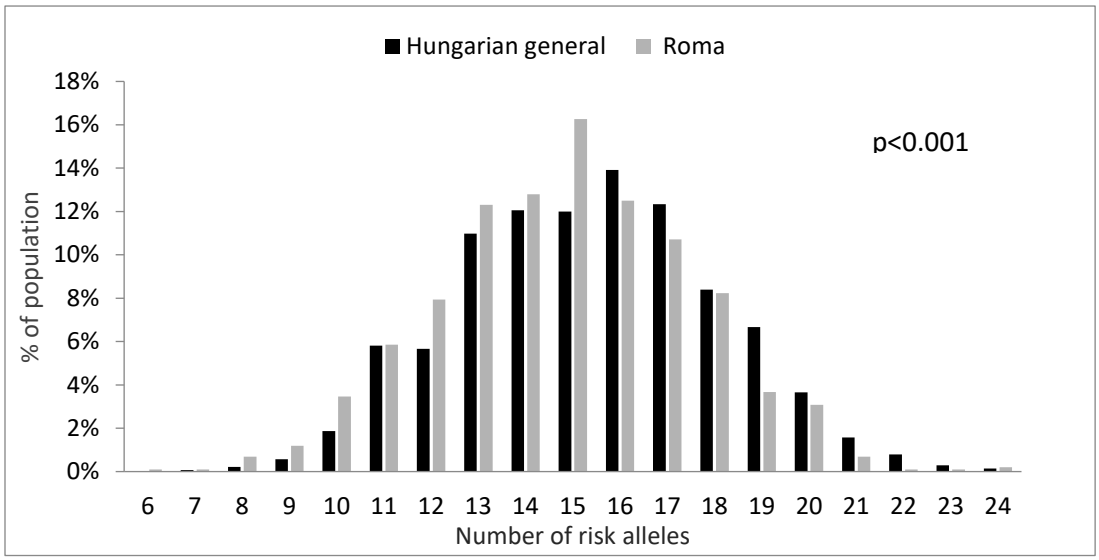


Figure 3. Distribution of GRSs based on 16 SNPs by study population samples.

The average wGRS in the Roma group was 1.36±0.31, while it was 1.41±0.32 for the Hungarian general population. The distribution of wGRS was significantly ($p<0.001$) different between the study populations. The distribution curves of the wGRS values for the study populations are shown in Figure 4.

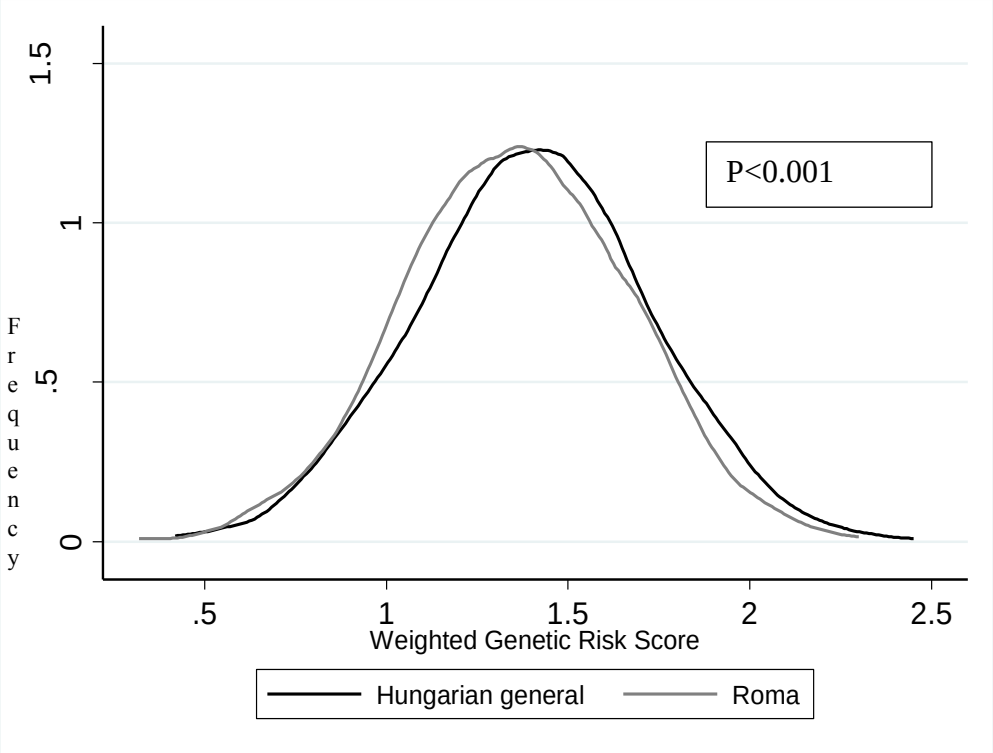


Figure 4. Distribution curves of wGRS by study population.

4.6. Association of GRS and wGRS with FG levels and T2DM status

Both the GRS and wGRS were analyzed for the association with FG level as a continuous variable and with T2DM status as a binary variable. The GRS was significantly associated with both outcomes in the adjusted (sex, age, BMI, HDL-C and TG levels were the covariates) model both in the Hungarian general (β=0.053, p=0.001; OR=1.070, p=0.027) and in the Roma (β=0.044, p=0.037; OR=1.083, p=0.010) populations (Table 5.).

Table 5. Association of GRS with FG level and T2DM status by study groups. The association was evaluated under adjusted (sex, age, BMI, HDL-C and TG level) regression models.

A	**FG level**					
	Hungarian general population			**Roma population**		
	β	**95% CI**	**p value**	**β**	**95% CI**	**p value**
GRS	0.053	0.023–0.082	0.001	0.044	0.003–0.085	0.037
Sex (male as reference)	-0.432	-0.602– -0.262	<0.001	-0.033	-0.264–0.200	0.775
Age	0.042	0.351–0.050	<0.001	0.022	0.013–0.031	<0.001
BMI	0.028	0.011–0.046	0.001	0.048	0.035–0.061	<0.001
HDL-C	-0.313	-0.532– -0.929	0.005	-0.601	-0.899– -0.303	<0.001
TG	0.021	-0.038–0.079	0.486	0.151	0.063–0.238	0.001
	T2DM status					
B	**Hungarian general population**			**Roma population**		
	OR	**95% CI**	**p value**	**OR**	**95% CI**	**p value**
GRS	1.07	1.008–1.137	0.027	1.083	1.020–1.151	0.010
Sex (male as reference)	0.385	0.271–0.547	<0.001	0.701	0.505–0.973	0.034
Age	1.069	1.052–1.086	<0.001	1.03	1.016–1.045	<0.001
BMI	1.09	1.052–1.130	<0.001	1.031	1.012–1.051	0.001
HDL-C	0.897	0.571–1.410	0.638	0.628	0.406–0.970	0.036
TG	1.242	1.095–1.411	0.001	1.54	1.347–1.761	<0.001

In the wGRS model the association was significant for both FG level and T2DM status in the Hungarian general population (β=0.489, p<0.001; OR=2.564, p<0.001); however, in the case of Roma population, a significant association was found only for T2DM status (OR=1.932, p=0.016) but not for FG level (β=0.300, p=0.100) (Table 6.).

Table 6. Association of wGRS with FG level (A) and T2DM status (B) by study groups. The association was evaluated under adjusted (sex, age, BMI, HDL-C and TG level) regression models.

A	**FG level**					
	Hungarian general population			**Roma population**		
	β	**95% CI**	**p value**	**β**	**95% CI**	**p value**
wGRS	0.489	0.240–0.738	<0.001	0.300	-0.062–0.663	0.104
Sex (male as reference)	-0.436	-0.605– -0.266	<0.001	-0.032	-0.263–0.198	0.783
Age	0.043	0.035–0.050	<0.001	0.022	0.013–0.031	<0.001
BMI	0.029	0.013–0.046	0.001	0.048	0.035–0.061	<0.001
HDL-C	-0.312	-0.531– -0.093	0.005	-0.596	-0.895– -0.297	<0.001
TG	0.020	-0.038–0.079	0.501	0.152	0.064–0.240	0.001
	T2DM status					
B	**Hungarian general population**			**Roma population**		
	OR	**95% CI**	**p value**	**OR**	**95% CI**	**p value**
wGRS	2.564	1.526–4.309	<0.001	1.932	1.133–3.292	0.016
Sex (male as reference)	0.384	0.270–0.547	<0.001	0.710	0.512–0.987	0.041
Age	1.07	1.053–1.087	<0.001	1.030	1.016–1.044	<0.001
BMI	1.091	1.052–1.131	<0.001	1.032	1.023–1.052	0.001
HDL-C	0.897	0.570–1.411	0.637	0.623	0.403–0.964	0.034
TG	1.249	1.099–1.418	0.001	1.539	1.346–1.760	<0.001

In further analysis the two study populations were combined and Roma ethnicity (Hungarian general population was used as reference) was integrated into the models (Model I and II) as a covariate (beside to age, sex, BMI, HDL, TG level and GRS) to eliminate the effect of all ethnicity-related (environmental and/or cultural) factors. In these models, the effect of GRS (Model I) and wGRS (Model II) could be examined independently from the ethnicity (Table 7.). The associations between the GRS (Model I) and FG level and T2DM status were significant (FG: β GRS=0.050, p<0.001; T2DM status: OR GRS=1.075, p=0.001), as in the case of the weighted models (Model II) (FG: β wGRS=0.425, p<0.001; T2DM status: OR wGRS=2.128, p<0.001).

In addition to genetic risk score and Roma ethnicity - in harmony with previously published findings - to be a male, to be older and having higher TG level have also identified as risk factors for elevated FG level and/or development of T2DM (Table 5 A and B, Table 6 A and B, Table 7). It is important to highlight that in these multivariate models, the effect of Roma ethnicity was relatively strong on both outcomes (FG levels: β ethnicity=0.918, p<0.001; T2DM status: OR ethnicity=2.484, p<0.001).

Table 7. Association of Roma ethnicity (Hungarian general population was used as reference) with FG level and T2DM status.

	FG level					
	Model I*			**Model II****		
	β	**95% CI**	**p value**	**β**	**95% CI**	**p value**
Ethnicity	0.918	0.779–1.058	<0.001	0.910	0.770–1.049	<0.001
GRS	0.050	0.026–0.075	<0.001	0.425	0.216–0.634	<0.001
Sex (male as reference)	-0.262	-0.400– -0.123	<0.001	-0.263	-0.402– -0.125	<0.001
Age	0.032	0.027–0.038	<0.001	0.032	0.027–0.038	<0.001
BMI	0.042	0.032–0.052	<0.001	0.042	0.033–0.052	<0.001
HDL-C	-0.387	-0.564– -0.211	<0.001	-0.383	-0.560– -0.206	<0.001
TG	0.061	0.116–0.110	0.015	0.061	0.012–0.110	0.015
	T2DM status					
	Model I*			**Model II****		
	OR	**95% CI**	**p value**	**OR**	**95% CI**	**P value**
Ethnicity	2.484	1.954–3.156	<0.001	2.472	1.945–3.141	<0.001
GRS	1.075	1.031–1.121	0.001	2.128	1.477–3.067	<0.001
Sex (male as reference)	0.552	0.436–0.698	<0.001	0.554	0.438–0.701	<0.001
Age	1.047	1.036–1.058	<0.001	1.047	1.037–1.058	<0.001
BMI	1.053	1.035–1.070	<0.001	1.053	1.036–1.071	<0.001
HDL-C	0.808	0.597–1.094	0.169	0.809	0.597–1.095	0.170
TG	1.378	1.261–1.505	<0.001	1.381	1.264–1.509	<0.001

Note: The association of Roma ethnicity with fasting glucose level and T2DM status was evaluated under regression models (Model I and II) in the combined population

*Model I was adjusted for ethnicity and GRS as well as sex, age, BMI, HDL-C and TG level

**Model II was adjusted for ethnicity and wGRS as well as sex, age, BMI, HDL-C and TG level

4.7. The best fitting genetic models for SNPs in the case population

Adjusted (by BMI, TG/HDL-C ratio, sex, and duration of T2DM) linear regression analyses were used to test the association of SNPs with the age of onset for T2DM in the case population. For each SNP, we have tested which of the three most commonly used genetic models of inheritance (codominant, recessive, and dominant) shows the strongest correlation with the age of onset for T2DM. The model with the lowest AIC, BIC, and p value was chosen for GRS calculation. In 15 cases in the recessive, in 5 cases in the dominant, and in 1 case in the codominant model SNPs showed the strongest correlation with the age of onset (Table 8.).

Table 8. Genetic models most fitted i.e., the strongest correlation with the patient's age of onset of T2DM could be detected in the adjusted regression model by SNPs.

No.	SNP	Gene	Effect allele	Genetic model	β (95%CI)	p value
1	rs174550	FADS1	C	Recessive	-0.866 (-1.812–0.079)	0.073
2	rs7903146	TCF7L2	T	Recessive	-0.782 (-1.719–0.155)	0.102
3	rs7944584	MADD	A	Recessive	-0.467 (-1.033–0.099)	0.106
4	rs10830963	MTNR1B	G	Dominant	-0.426 (-0.995–0.143)	0.142
5	rs7034200	GLIS3	A	Dominant	-0.413 (-1.067–0.240)	0.215
6	rs10885122	ADRA2A	T	Recessive	-1.326 (-3.586–0.934)	0.250
7	rs5219	KCNJ11	T	Recessive	-0.427 (-1.181–0.326)	0.266
8	rs3736594	MRPL33	C	Recessive	-0.575 (-1.695–0.545)	0.314
9	rs560887	G6PC2	T	Recessive	-0.479 (-1.472–0.512)	0.344
10	rs11671664	GIPR	G	Recessive	-0.272 (-0.932–0.388)	0.419
11	rs10946398	CDKAL1	C	Codominant	-0.322 (-1.168–0.524)	0.455
12	rs11920090	SLC2A2	T	Recessive	-0.242 (-0.901–0.416)	0.470
13	rs7173964	C2CD4B	G	Recessive	-0.200 (-0.795–0.394)	0.509
14	rs10811661	CDKN2A/B	C	Recessive	-0.573 (-2.437–1.291)	0.546
15	rs340874	PROX1	C	Dominant	-0.133 (-0.768–0.502)	0.680
16	rs10906115	CDC123/CAMK1D	G	Dominant	-0.119 (-0.703–0.466)	0.691
17	rs11071657	C2CD4B	G	Dominant	-0.084 (-0.664–0.496)	0.776
18	rs780094	GCKR	C	Recessive	-0.066 (-0.666–0.534)	0.830
19	rs1111875	HHEX	C	Recessive	-0.055 (-0.632–0.521)	0.851
20	rs11558471	SLC30A8	G	Recessive	-0.087 (-1.106–0.931)	0.866
21	rs2191349	DGKB	T	Recessive	-0.004 (-0.596–0.589)	0.990

4.8. Results of the optimization of the GRS model

In calculating the GRS, we have selected those SNPs that strengthened the association of the GRS with the outcome (age of onset for T2DM) in the linear regression model by moving from the SNP with the strongest correlation (rs174550; β=−0.866, p=0.073) to the weakest (rs2191349; β=−0.004, p=0.990). The SNPs were individually inserted and tested by adjusted (by BMI, TG/HDL-C, sex, and duration of T2DM) linear regression models. All SNPs that strengthened the association of GRS with the outcome variable (raised the value of r^2) were selected and inserted in the optimized GRS model, while those that weakened (reduced the value of r^2) were excluded. Finally, 12 SNPs were selected for the optimized GRS model (Table 9.).

Table 9. SNPs tested for the GRS calculation to optimize the model

Steps	Inserted SNP	GRS association			Included/excluded from the GRS
		β	p value	r^2	
1	rs174550	-0.866 (-1.812 – 0.079)	0.073	0.228 ↑	Included
2	rs7903146	-0.865 (-1.546– -0.184)	0.013	0.230 ↑	Included
3	rs7944584	-0.627 (-1.060– -0.193)	0.005	0.232 ↑	Included
4	rs10830963	-0.533 (-0.871– -0.195)	0.002	0.233 ↑	Included
5	rs7034200	-0.510 (-0.810– -0.210)	$9x10^{-4}$	0.235 ↑	Included
6	rs10885122	-0.527 (-0.825– -0.229)	$6x10^{-4}$	0.235 ↑	Included
7	rs5219	-0.510 (-0.785– -0.234)	$3x10^{-4}$	0.236 ↑	Included
8	rs3736594	-0.526 (-0.797– -0.255)	$1.5x10^{-4}$	0.238 ↑	Included
9	rs560887	-0.517 (-0.777– -0.258)	$1x10^{-4}$	0.238 ↑	Included
10	rs11671664	-0.478 (-0.713– -0.243)	$7x10^{-5}$	0.239 ↑	Included
11	rs10946398	-0.463 (-0.713– -0.243)	$9x10^{-5}$	0.238 ↓	Excluded
12	rs11920090	-0.456 (-0.679– -0.234)	$6x10^{-5}$	0.239 ↑	Included
13	rs7173964	-0.410 (-0.615– -0.205)	$9x10^{-5}$	0.238 ↓	Excluded
14	rs10811661	-0.454 (-0.674– -0.234)	$5.5x10^{-5}$	0.239 ↑	Included
15	rs340874	-0.420 (-0.628– -0.212)	$8x10^{-5}$	0.238 ↓	Excluded
16	rs10906115	-0.387 (-0.587– -0.187)	$1.5x10^{-4}$	0.238 ↓	Excluded
17	rs11071657	-0.415 (-0.623– -0.208)	$9x10^{-4}$	0.238 ↓	Excluded
18	rs780094	-0.392 (-0.595– -0.190)	$1.5x10^{-4}$	0.238 ↓	Excluded
19	rs1111875	-0.386 8-0.591– -0.181)	$2x10^{-4}$	0.237 ↓	Excluded
20	rs11558471	-0.434 (-0.648– -0.220)	$8x10^{-5}$	0.238 ↓	Excluded
21	rs2191349	-0.403 (-0.610– -0.195)	$1.5x10^{-4}$	0.238 ↓	Excluded

SNPs have an improving effect of correlation are highlighted by shadow

4.9. Effect of GRS on the age of onset for T2DM in the case population

The mean value of GRS was 7.72 (7.55–7.88) in the full case population; 7.75 (7.49–8.00) for males and 7.69 (7.46–7.91) for females. The GRS showed a significant association with the age of onset for T2DM in the full case population and also separately in both sexes. The TG/HDL-C ratio significantly associated with the age of onset for T2DM in the male population ($\beta=-0.556$, $p<0.001$), while it was not observed in the female one ($\beta=-0.136$, $p=0.251$). Females are more protected against the early manifestation of T2DM compared to males (males vs. females: $\beta=2.352$, $p<0.001$) (Table 10.).

Table 10. Association of GRS with the age of onset for T2DM in the full case population (A) and separately for sexes (B in males, C in females). The association was analysed under adjusted regression models (sex/full population/, BMI, TG/HDL-C ratio).

A- full case population	**β (95%CI)**	**p value**
Sex	2.352 (1.228–3.475)	<0.001**
BMI	-0.330 (-0.434– -0.227)	<0.001**
TG/HDL-C ratio	-0.354 (-0.511– -0.198)	<0.001**
Duration of T2DM	-0.607 (-0.69– -0.515)	<0.001**
GRS	-0.454 (-0.67– -0.234)	<0.001**
B- males	**β (95%CI)**	**p value**
BMI	-0.290 (-0.435– -0.145)	<0.001**
TG/HDL-C ratio	-0.556 (-0.767– -0.346)	<0.001**
Duration of T2DM	-0.646 (-0.774– -0.517)	<0.001**
GRS	-0.434 (-0.722– -0.145)	0.003**
C- females	**β (95%CI)**	**p value**
BMI	-0.376 (-0.523– -0.229)	<0.001**
TG/HDL-C ratio	-0.136 (-0.369–0.097)	0.251
Duration of T2DM	-0.579 (-0.708– -0.450)	<0.001**
GRS	-0.405 (-0.796– -0.120)	0.008*

Notes: * statistically significant p value ($p<0.05$). ** significant p value with Bonferroni correction

There is a significant association between GRS and age of onset for T2DM appearing as decreasing trend by age in the total T2DM case population, as well as in both sexes. The development of T2DM occurred at a younger age among individuals with higher GRS values (Table 11.).

Table 11. The average values of GRS in the age categories created by the onset of T2DM, and results of p for trend analyses

	≤49 years	**50-59 years**	**≥60 years**	**p for trend**
	Mean GRS (95%CI)	**Mean GRS (95%CI)**	**Mean GRS (95%CI)**	
Full population	8.36 (7.97–8.75, n=191)	7.79[a] (7.52–8.06, n=340)	7.30[b] (7.04–7.55, n=350)	<0.001**
Males	8.18 (7.63–8.73, n=111)	8.00 (7.61–8.39, n=176)	7.12[b] (6.71–7.52, n=147)	0.002**
Females	8.60 (8.05–9.15, n=80)	7.56[b] (7.20–7.92, n=164)	7.43[b] (7.10–7.76, n=203)	0.0038**

Notes: [a]statistically significant difference ($p<0.05$) compared with the ≤49 years old T2DM subpopulation. [b]significant difference with Bonferroni correction compared with the ≤49 years old T2DM subpopulation. ** significant p value with Bonferroni correction.

4.10. Association of GRS with T2DM in the Hungarian general population

Based on the results obtained in the adjusted logistic regression model the GRS did not show a significant association with existing T2DM in the Hungarian general population. All conventional risk factors (age, sex, BMI, and TG/HDL-C ratio) showed a significant correlation with the outcome in the model (Table 12.).

Table 12. Association of GRS with T2DM status in the Hungarian general population. The association was investigated under the adjusted (age, sex, BMI, TG/HDL-C ratio) regression model.

	OR (95%CI)	**p value**
Age	1.087 (1.063–1.112)	<0.001**
Sex (male as reference)	0.502 (0.320–0.788)	0.003**
BMI	1.066 (1.025–1.109)	0.001**
TG/HDL-C ratio	1.312 (1.186–1.451)	<0.001**
GRS	1.032 (0.945–1.126)	0.488

Notes: ** statistically significant p value with Bonferroni correction.

4.11. Association of GRS with the age in the subpopulations of the Hungarian general population

The association of GRS with age was tested by adjusted linear regression model on the subpopulations (based on FG level and/or treatment for diabetes) of the Hungarian general population sample. A significant correlation between patients' age and GRS (β=−0.999, p=0.003) was detected only in the subpopulation with T2DM (FG level of 7 mmol/L or higher and/or under antidiabetic treatment). Sex showed a significant association with age in the subpopulation with normal glucose level, while BMI was significantly associated with the age of patients in the subpopulation with normal FG level and prediabetes. TG/HDL-C ratio had no significant effect in any of the subpopulations (Table 13.).

Table 13. Association of GRS with the age of subpopulations with normal FG level (n=1197), prediabetes (n=108) and T2DM (n=110). The association was evaluated under adjusted (sex, BMI, and TG/HDL-C ratio) linear regression model.

Subpopulation with normal FG	**β (95%CI)**	**p value**
Sex	3.293 (1.991–4.596)	<0.001**
BMI	0.621 (0.510–0.732)	<0.001**
TG/HDL-C ratio	0.088 (-0.213–0.388)	0.567
GRS	0.104 (-0.158–0.365)	0.437
Subpopulation with prediabetes	**β (95CI)**	**p value**
Sex	1.559 (-2.358–5.476)	0.432
BMI	0.330 (0.024–0.635)	0.035*
TG/HDL-C ratio	-0.338 (-1.092–0.416)	0.376
GRS	-0.245 (-0.972–0.483)	0.507
Subpopulation with T2DM	**β (95CI)**	**p value**
Sex	0.128 (-3.296–3.553)	0.941
BMI	0.165 (-0.086–0.416)	0.195
TG/HDL-C ratio	-0.413 (-1.088–0.262)	0.228
GRS	-0.999 (-1.660– -0.337)	0.003**

Notes: * statistically significant p value (p<0.05). ** statistically significant p value with Bonferroni correction.

Five categories were formed based on the GRS values. Between the average age of people and GRS categories, a significant decreasing trend in age was found only in the subpopulation with T2DM (Table 14.).

Table 14. The average age of people by GRS categories in the Hungarian general subpopulations.

No. of risk alleles	**GRS<4, n=91**	**GRS=4, n=286**	**GRS=6, n=469**	**GRS=8, n=379**	**GRS>8, n=190**	**p for trend**
	Average age (95%CI)					
Subpopulation with normal FG	41.68 (38.84–44.53)	43.30 (41.75–44.85)	42.16 (40.96–43.35)	42.50 (41.14–43.85)	43.91 (42.08–45.74)	0.563
Subpopulation with prediabetes	49.71 (39.64–59.78)	50.88 (46.00–55.75)	52.20 (48.35–56.05)	49.03 (46.20–51.86)	50.64 (47.11–54.18)	0.222
Subpopulation with T2DM	58.38 (53.28–63.47)	57.32 (55.73–58.90)	54.81 (52.46–57.15)	52.19a (49.05–55.32)	51.80[a] (46.11–57.49)	0.0045*

*statistically significant p value ($p<0.05$)

[a]: average age at least nominal significantly differed compared with the GRS<4 subgroup

4.12. Estimation of the age of onset for T2DM by a score based on genetic and non-genetic factors in the Hungarian general population

An age of onset risk score (AORS) for T2DM was calculated based on the individuals' sex, BMI, TG/HDL-C ratio, and GRS by multiplying these components with their effects measured on the case population (see more details in Table 10. A), to estimate the age of onset for T2DM in the Hungarian general population.

The mean AORS values (normal FG: 13.26 vs. prediabetes: 15.27 and T2DM: 16.00) showed a significant difference between the samples with prediabetes or T2DM and subpopulation with normal FG. In terms of mean values, all non-genetic components (sex, BMI, and TG/HDL-C ratio) differed at a statistically significant level ($p<0.05$) between the subpopulations with normal FG and prediabetic or T2DM patients. The mean values of wGRS did not differ significantly between the study subpopulations (Figure 5.). This result is consistent with the fact that genetic determination of the age of onset for T2DM remains constant from birth, but environmental and lifestyle factors play a significant role in the development of T2DM. This finding is in good harmony with data obtained previously in different studies [24, 154, 155].

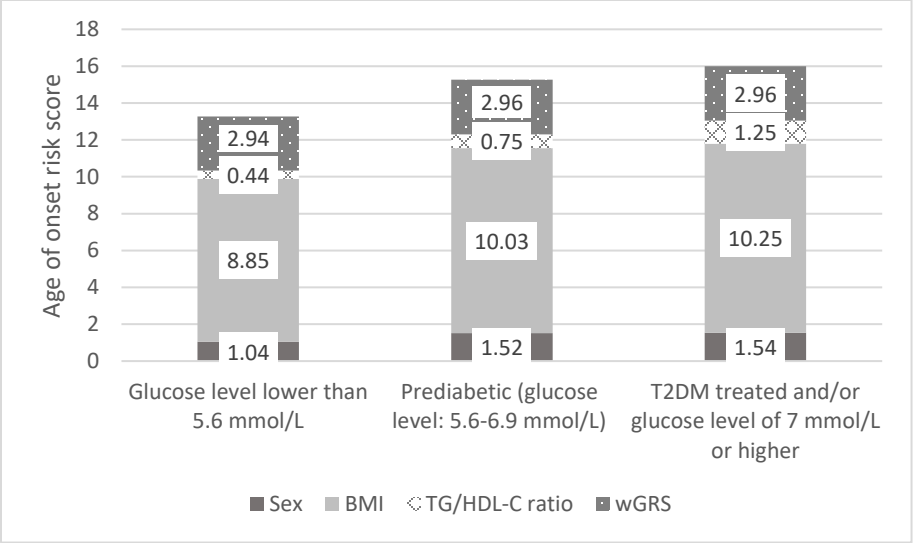


Figure 5. Representation of different components (sex, BMI, TG/HDL-C ratio, and wGRS) in the AORS values in the Hungarian general population by subpopulations.

The representation of AORS's components (%) in the subpopulations was also examined. There are statistically significant trend ($p<0.05$) in changing the representation of sex, BMI and TG/HDL-C ratio across the subpopulations, but the contribution of wGRS remains unchanged. Regarding sex, its effect on the AORS is higher in the prediabetic and T2DM groups compared to the normal one, which is in harmony with the observation on a higher proportion of men in the T2DM group. In case of the TG/HDL-C ratio, its contribution to the AORS is higher in the prediabetic and T2DM groups than in the normal one, the increasing trend can be explained by the fact that lipid and glucose metabolism are closely linked and the TG/HDL-C ratio is considered as a sensitive indicator of susceptibility to T2DM [156]. The "weight" of non-genetic factors is increasing with the progression of disturbances in carbohydrate metabolism; and although the weight of genetic component never changes (see Figure 5.), there is a decreasing trend in the share of genetic risk factors (Figure 6.).

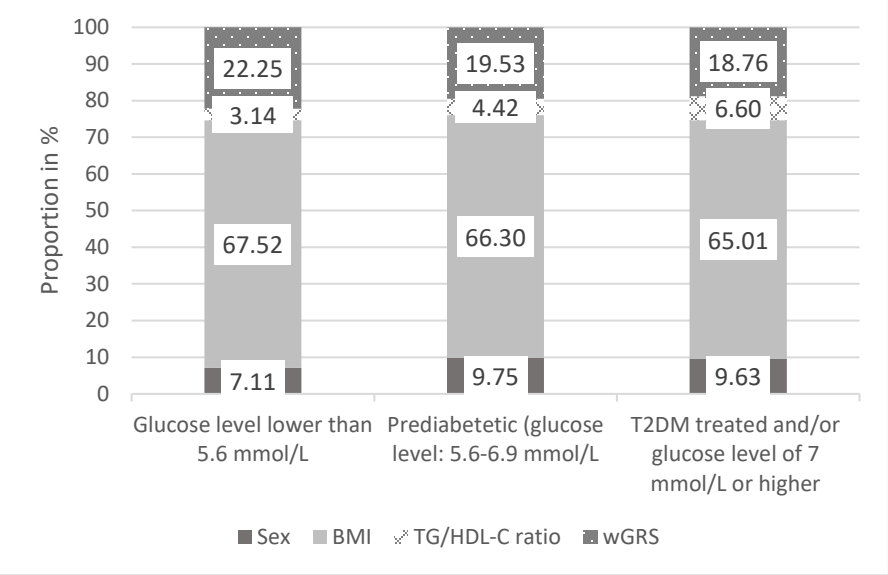


Figure 6. The representation of different components (sex, BMI, TG/HDL-C ratio, and wGRS) in the age of onset risk score in the Hungarian general population by subpopulations.

4.13. The effect of wGRS on the age of onset for T2DM in the Hungarian general population

Linear regression analyses were performed to examine the effect of AORS's components on the age of onset for T2DM on T2DM subpopulation in the Hungarian general one. Out of the four inserted components (sex, BMI, TG/HDL-C ratio, and wGRS), only wGRS showed a significant ($p=0.0036$) association with the age of onset for T2DM. A one-unit increase in wGRS results in developing T2DM two years earlier (see more details in Table 15), which shows a striking resemblance to the findings of the study carried out by Zhou et al. on a sample of the Scottish population [157].

Table 15. Results of the linear regression analysis on AORS's components related to the age of onset for T2DM in the Hungarian general population's T2DM subpopulation.

	β (95%CI)	p value
Sex	-0.346 (-1.667–0.975)	0.604
BMI	0.485 (-0.250–1.221)	0.194
TG/HDL-C ratio	-0.167 (-0.803–0.469)	0.605
wGRS	-2.011 (-3.347– -0.674)	0.0036 **

Notes: ** statistically significant p value with Bonferroni correction.

To describe the association between the wGRS and the age of onset for T2DM, we examined the representation of wGRS in AORS in three different age groups (≤49 yrs, 50–59 yrs, and ≥60 yrs) in the T2DM subpopulation. The representation of wGRS decreased significantly (p=0.023) from the under 50 years (20.95%) through the 50–59 (19.31%) to the over 60 years of age group (15.49%) among type 2 diabetic patients. The same trend was observed in case of the representation of sex (≤49 yrs: 9.79%, 50–59 yrs: 7.67%, ≥60 yrs: 14.00%; p=0.016) while in case of that of BMI (p=0.383) and TG/HDL-C ratio (p=0.365) no significant change in trend was observed across the age groups (see more details in Figure 7).

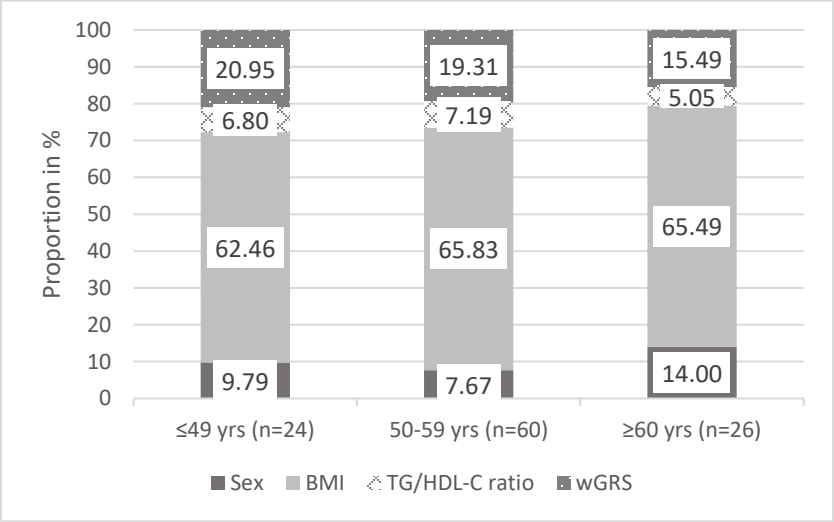


Figure 7. The representation of AORS' components (sex, BMI, TG/HDL-C ratio, and wGRS) among individuals in different age groups of the T2DM subpopulation of the Hungarian general population.

5. Discussion

Our study was carried out to determine whether genetic factors contribute to the higher prevalence of raised FG level and/or T2DM among Roma by comparing differences in frequencies and load of the risk alleles to T2DM between the Hungarian general and the Roma populations. Sixteen SNPs associated with T2DM were genotyped, and differences in eight SNPs were significant when the two groups were compared. Five susceptibility alleles were found more prevalent in the Hungarian general population, whilst three alleles were more frequent among Roma.

Recently in a similar study Hubáček et al. [58] examined the allelic differences between the Czech population and the Roma populations. From the examined eight SNPs, only two SNPs were identical to those analysed in our investigation. The allele frequencies between the Hungarian general and the Czech population population did not differ significantly (29.5% vs. 27.5% and 82.8% vs. 82.1% for rs7903146 and rs10811661, respectively). Hungarian Roma have higher prevalence (24.2% vs. 17.8%) of risk allele consisting of rs7903146, however in the case of rs10811661, they have lower frequency (85.3 vs. 90.1%) compared with Roma residing in Czech territory [58].

We also constructed GRS and wGRS based on sixteen SNPs and compared their distribution between the study populations. The results indicate that the Hungarian general population has greater genetic risk load for the development of T2DM compared with the Roma population. Our result is quite opposite to the recently reported findings of the above cited Czech authors, who reported higher genetic load within the Roma population. The divergence of our result from the finding of the above authors may be explained by selection of SNPs; the Czech researchers have chosen 8 SNPs, of which only two (rs7903146 and rs10811661) were identical with ours.

Our multivariate regression analysis has shown that both GRS and wGRS were significantly associated with FG and T2DM status in the Hungarian general population, while this association was modest in the case of the Roma population. The two populations were combined and analyzed together when ethnicity as a covariate was inserted into the model in addition to age, sex, BMI, HDL-C, TG and GRSs and it was showed that ethnicity and GRSs had significant impact on the outcomes. By this combined analysis, the effect of ethnicity-related factors (such as lifestyle, environmental or even unknown genetic factors) could be adjusted for. The combined effect of 16 SNPs incorporated in our GRS model significantly influenced the development of T2DM in the Hungarian general population, and this effect was significantly modulated by ethnicity-related factors among the Roma.

Our results reveal that the higher prevalence of elevated FG and/or T2DM among Roma is not connected directly to their increased genetic load. Based on our findings it is reasonable to suppose that lifestyle and/or environmental factors could explain the higher prevalence of disturbances in glucose metabolism. It is well known fact that environmental factors and unhealthy lifestyles such as physical inactivity [158], overweight or obesity [159] and unhealthy diet [160] strongly increase the risk of developing T2DM and are linked to poor health conditions in general. Roma is more likely to suffer from conditions than the majority population, irrespective of the country where they live [161-163]. Moreover, accumulated reports revealed that healthy diet (relatively low intake of fats, and high consumption of fruits and vegetables) and physical activities are less common in the Roma population [164, 165]. It seems likely that the burden of unhealthy lifestyles and cultural attributes contribute to the high prevalence of prediabetes or T2DM among Roma, but still the role of unknown genetic components in the development of T2DM cannot be excluded.

During the last ten years, the global burden of diabetes has been escalating at an alarming rate. Currently, half a billion people (9.3% of adults) have been living with diabetes worldwide. The number of people living with diabetes increased by 62%, from 285 million in 2009 to 463 million in 2019. The growing prevalence of fasting glucose level and T2DM has been reported in the younger adults as well [2, 166-168]. Several countries also reported an increasing prevalence of T2DM in younger adults and even in adolescents [154, 169, 170]. This is an alarming trend since the EOT2DM is expected to be associated with higher risk of cardiovascular (micro- and macrovascular) complications and increased frequency of comorbidities at later life. It is obvious that early identification of EOT2DM risk is essential for the development of effective preventive intervention strategies against T2DM in general. Developing sensitive and precise risk assessment tools are important to identify the inheritable and non-inheritable risk factors contributing to the disease manifestation and by using this information to stratify populations accordingly.

The other aim of our study is to quantify the combined effect of T2DM associated SNPs (using genetic risk score modelling) and known non-inheritable risk factors such as sex, BMI, and TG/HDL-C ratio on the age of onset for T2DM in the Hungarian population. To best of our knowledge, this is the first study to explore impact of genetic influences on the age of onset for T2DM on the Hungarian population.

Twenty-three SNPs that have a role on the development of T2DM were genotyped, and no SNP was identified to have significant individual association with the age of onset for T2DM. Indeed, two SNPs were excluded from the analysis due to failed to be within the linkage equilibrium. Very limited studies have assessed the association of these SNPs with the age of onset for T2DM. Our result is in contradictory with previously published findings of Silbernagel and his colleagues who reported association of rs7903146 with age at onset of T2DM [171]. Behind this fact it may exist

that we have adjusted the model with covariates (BMI, TG/HDL-C ratio, sex, and duration of T2DM); however, Silbernagel et al. adjusted it only for sex and BMI. Our finding also disagrees with more recently published results on impact of T2DM variants identified through GWAS in early-onset T2DM from South Indian population by Liju et al. The authors found a significant association between rs1111875 and early onset of T2DM[44].

In the optimization process of GRS calculation, twelve SNPs have been detected that improved the strength of the association between the GRS and the age of onset for T2DM in the T2DM case population. In our multivariate linear regression analysis (adjusted for sex, BMI, TG/HDL-C and duration of T2DM), the GRS showed a strong significant association with the age of onset for T2DM in the whole Hungarian population and in both sexes.

When the relationship between the GRS trend and the age groups created on the basis of age at onset of T2DM was studied in the case population, a significant trend was noted between the average GRS vales and onset age groups and it found significant separately for males and for females as well.

The optimized GRS model which was created on the case population was tested on the Hungarian general population. In the adjusted logistic regression model, we did not observe significant association between the GRS and the presence of T2DM, but significant association between age, sex, BMI and TG/HDL-C ratio and T2DM was detected.

A small number of studies investigated and reported the impact of GRS on the age of onset for T2DM. Our current findings agree with all these previously published findings. Iwata and his colleagues explored that the GRS, constructed by incorporating 14 SNPs, showed association with early onset of T2DM in the Japanese population [172]. Similar result was also reported recently

by Kong et al. in the Chinese population who contracted 24 SNPs into a single quantitatively measurable risk score (GRS) and evaluated its association with early onset of T2DM [173]. Our results also supported by the observations on pooled data of the Framingham Offspring study that convincingly show that considering also GRS in addition to clinical factors efficiently improved the predictive ability of risk assessment in younger adults (<50years of age) but not for individuals above 50 years of age [155].

The results of the EPIC InterAct case-cohort study which examined the association between genetic risk score (integrating 49 SNPs) and the age of onset for T2DM also support our findings. The authors observed higher relative genetic risk for persons who developed T2DM at the younger age (below 55 years of age) compared to individuals who developed T2DM at later age (55-65 years or ≥65 years of age) [174]. Similarly, a more recent study by Mars et al. reports that individuals with higher GRS developed the disease at an earlier age than people with lower GRS [175]. The researchers also conclude that GRS has an influence on the age of onset for T2DM.

In our replication study on the Hungarian general population, significant association between GRSs (both unweighted and weighted GRS) and the age of onset was observed in group of persons with T2DM. We could show that the higher the GRS is, the lower the age of onset for T2DM is. When the AORS was evaluated among the three subpopulations (normal FG, prediabetes and T2DM) created from the Hungarian general population, the risk score was significantly higher in the T2DM subpopulation compared with the normal subpopulation. In this analysis, we could observe significant differences for all non-genetic factors amongst the three subpopulations. However, significant difference was not detected for GRS between these subpopulations. This corroborates that the genetic determination remains constant throughout once life. Individual's increased genetic risk for the early onset of T2DM is manifested only if the effects of non-genetic

risk factors are high enough. When the effect of AORS and its components was assessed on the T2DM subpopulation, significant association between GRS and age of onset for T2DM was witnessed. The earlier the age of onset for T2DM is, the higher the GRS and wGRS are.

Our study also showed a strong association between age of onset for T2DM and TG/HDL-C ratio for men only. Nguyen and his colleagues in the Bogalusa Heart Study also reported a significant association between TG/HDL-C ratio and age of onset for T2DM [156], however, the authors did not evaluate their association separately for males and females. In fact, compared with women, men are usually diagnosed with diabetes at earlier age [176].

The strength of this study is that the results obtained on the T2DM case population were validated on an independent sample population. It is obvious that our study has limitations: the first limitation is that although majority of the Roma population resides in the catchment area, this sample cannot be interpreted as a representative sample for the whole Hungarian Roma population. Samples were excluded based on various criteria (refuse to participate, missing pheno- and genotype data….), so the sample population used in the analyses can no longer be considered representative of the Hungarian general population. Since some Roma people are reluctant to self-define their identities as ethnically ‘’Roma’’, the representative Hungarian general sample included some people who are Roma. It is possible that their inclusion could have resulted in a slight underestimation of the differences between the two populations. Due to unavailability of data on gene-gene interactions, gene-environmental interactions, epigenetic factors, and structural variants, we did not integrate them into the models. In fact, all these factors can modify the genetic risk. In our study we considered only the major confounding factors (age, sex, BMI, HDL-C and TG). Several behavioral factors (such as physical inactivity and diet) that can undeniably modify susceptibility to the studied trait were not investigated and consequently they can account for

differences in the prevalence between the studied populations to some extent. In our study we have considered sixteen and twelve SNPs which have an effect on the development of T2DM for comparison of risk allele load between the Roma and Hungarian general population and for age of onset for T2DM study, respectively in the GRS model. Integrating more SNPs into the GRS model could further increase the informative ability of the GRS model, although adding many more SNPs into the GRS model does not necessarily boost up the informative capability of the model [177, 178]. Since the current study was designed to define and compare the genetic risk for T2DM at population level among the Hungarian general population and Roma population, the difference between the effect of homozygous and heterozygous gene variants on FG level and/or T2DM cannot be estimated.

Regarding Roma a relatively high consanguinity was demonstrated about. High endogamy was proved by the gipsy origin of male partners in 90% of couples. The incidence of first cousin couples was sixteen times higher than that of the majority population at large [179, 180]. Based on this fact, it is reasonable to suppose that a number of private founder mutations could have an influence on trait among Roma. The founder mutations identified so far are related to diseases following Mendelian inheritance. Out of these, the intron 9 +1 G>T mutation in the SLC12A3 gene is associated with impaired glucose metabolism and significantly impaired insulin secretion in a study involving small number of samples [181]. Indeed, the effects of other still unknown founder mutations related to carbohydrate metabolic pathways - if they exist at all - cannot be excluded.

Understanding the SNPs-mediated development of T2DM could increase the clinical applicability of the present study. Our results need to be validated in other non-Hungarian populations. Certainly, our results may pave the way for the development of genetic tests that can be used to

predict the timing of T2DM development and delay or avert its manifestation through targeted interventions, which would also reduce the burden on health care systems.

Due to the advent of “big data”, and the evolution of analytical tools based on the results of genomic, epigenetic, metabolomic, proteomic and pharmacogenetic studies, personalized T2DM treatments are emerging, and a one-size-fits-all method is becoming outdated. Owing to polygenic nature of the disease and the influence of both environmental and genetic factors on its development, defining subgroups using molecular testing is difficult in type 2 diabetes mellitus patients. Hence, the best approach for the accurate and most convenient treatment of T2DM is to categorize patients based on their SNPs-based expected response to medicines. Studying how the SNPs influence drug efficacy may help us uncover new drug targets and personalized treatments.

The current study is not only the first to explore the possible genetic influence on the high prevalence of prediabetes and T2DM among Roma inhabiting in segregated colonies and to compare them with the general population but also it is the first to examine the impact of joint effect of T2DM associated SNPs using GRS modeling on the age of onset for T2DM in the Hungarian population. Compared with the Roma, the general population carries genetic load for the development of PreDM/T2DM. The combined impact of these genetic alterations on the development of PreDM/T2DM was stronger in the general population. However, the effect of genetic factors appears to be overwritten by ethnicity-related factors (such as environmental and lifestyle characteristics) in the Roma population. GRS modeling demonstrated that the combined effect of T2DM related SNPs was associated with the age of onset for T2DM. Compared with people who developed T2DM at later age, individuals who developed T2DM at earlier age carried greater risk alleles. Our study uncovered the considerable genetic susceptibility for the early onset of T2DM. Hence, GRS can be utilized as a tool for stratification and estimation of the risk of the

early onset of T2DM in the Hungarian population. We recommend that interventions targeting T2DM prevention in the Roma population ought to focus on harmful environmental exposures related to their unhealthy lifestyle. In fact, identifying individuals that are more susceptible to T2DM can more effectively improve the preventive interventions related to this disease in both populations.

6. Summary

Background: Compared with the Hungarian general population, type 2 diabetes mellitus (T2DM) and/or elevated fasting glucose level are more frequent in the Roma population. Genetic factors could be behind the difference in the prevalence between the two populations and may influence the age of onset of the disease.

Objective: The aims of our study were to assess whether the distribution of 16 single nucleotide polymorphisms (SNPs) with unequivocal effects on the development of T2DM contributes to the higher prevalence of T2DM among Roma and to evaluate the impact of genetic factors on the age of onset for T2DM in addition to conventional risk factors also in the Hungarian population.

Methods: A total of 1168 samples of T2DM individuals, 1783 samples from Hungarian general population and 1260 samples from segregated colonies of Roma were included in our study. Genetic risk scores, unweighted (GRS) and weighted (wGRS), were computed and compared between the study populations. Associations between GRSs and fasting glucose level and T2DM status were investigated in separate and combined study populations. For the impact of genetic factors on the age of onset for T2DM, twenty-one SNPs were tested on the case population. Twelve SNPs were chosen for the GRS analysis and the GRS was tested for validation on the Hungarian general population.

Results: The Hungarian general population carried a greater genetic risk for the development of T2DM ($GRS_{General}$ =15.38±2.70 vs. GRS_{Roma} =14.80±2.68, $p<0.001$; $wGRS_{General}$ = 1.41±0.32 vs. $wGRS_{Roma}$ = 1.36±0.31, $p<0.001$). In the combined population models, GRSs and wGRSs showed significant associations with elevated FG ($p<0.001$) and T2DM ($p<0.001$) after adjusting for ethnicity, age, sex, BMI, HDL-C, and TG. In these models, the effect of ethnicity was relatively strong on both outcomes (FG levels: $\beta_{ethnicity}$ =0.918, $p<0.001$; T2DM status:

$OR_{ethnicity}$ =2.484, p<0.001). For the impact of genetic factors on the age of onset for T2DM, the GRS showed a significant association with the age of onset for T2DM (β= -0.454, p=0.001) in the case population and among T2DM patients in the HG one (β= -0.999, p=0.003) during the replication. The higher the GRS, the earlier was the T2DM onset.

Conclusions: The higher prevalence of elevated FG level and/or T2DM among Roma does not appear to be directly linked to their increased genetic load but rather to their environmental/cultural attributes. Our results also suggest that there is a considerable genetic predisposition for early onset of T2DM among them. Interventions targeting T2DM prevention should focus on harmful environmental exposures related to their unhealthy lifestyle and GRS can be used as a tool for stratifying and estimating the risk of earlier onset of T2DM in addition to conventional risk factors.

7. References

1. World Health Organization. *Classification of diabetes mellitus*. 2019 [Accessed on Junuary 27, 2021]; Available from: https://apps.who.int/iris/handle/10665/325182.
2. Lascar, N., et al., *Type 2 diabetes in adolescents and young adults.* Lancet Diabetes Endocrinol, 2018. **6**(1): p. 69-80.
3. Wilmot, E. and I. Idris, *Early onset type 2 diabetes: risk factors, clinical impact and management.* Ther Adv Chronic Dis, 2014. **5**(6): p. 234-44.
4. Xu, Y., et al., *Prevalence and control of diabetes in Chinese adults.* JAMA, 2013. **310**(9): p. 948-59.
5. Song, S. and C. Hardisty, *Early onset type 2 diabetes mellitus: a harbinger for complications in later years—clinical observation from a secondary care cohort.* QJM: An International Journal of Medicine, 2009. **102**(11): p. 799-806.
6. Cockram, C.S., *The epidemiology of diabetes mellitus in the Asia-Pacific region.* Hong Kong Med J, 2000. **6**(1): p. 43-52.
7. Wei, J.N., et al., *National surveillance for type 2 diabetes mellitus in Taiwanese children.* JAMA, 2003. **290**(10): p. 1345-50.
8. World Health Organization. *Diabetes*. 2020 [Accessed on January 27, 2021]; Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes.
9. Roth, G.A., et al., *Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet, 2018. **392**(10159): p. 1736-1788.
10. Saeedi, P., et al., *Mortality attributable to diabetes in 20–79 years old adults, 2019 estimates: Results from the International Diabetes Federation Diabetes Atlas.* Diabetes research and clinical practice, 2020: p. 108086.
11. Chen, H., et al., *Contribution of specific diseases and injuries to changes in health adjusted life expectancy in 187 countries from 1990 to 2013: retrospective observational study.* bmj, 2019. **364**.
12. Wu, Y., et al., *Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention.* International journal of medical sciences, 2014. **11**(11): p. 1185.
13. Akshintala, D., et al., *Nonalcoholic Fatty Liver Disease: The Overlooked Complication of Type 2 Diabetes*, in *Endotext [Internet]*. 2019, MDText. com, Inc.
14. Zheng, Y., S.H. Ley, and F.B. Hu, *Global aetiology and epidemiology of type 2 diabetes mellitus and its complications.* Nature Reviews Endocrinology, 2018. **14**(2): p. 88.
15. Mata-Cases, M., et al., *Direct medical costs attributable to type 2 diabetes mellitus: a population-based study in Catalonia, Spain.* The European Journal of Health Economics, 2016. **17**(8): p. 1001-1010.
16. Deshpande, A.D., M. Harris-Hayes, and M. Schootman, *Epidemiology of diabetes and diabetes-related complications.* Physical therapy, 2008. **88**(11): p. 1254-1264.
17. Dart, A.B., et al., *Earlier onset of complications in youth with type 2 diabetes.* Diabetes Care, 2014. **37**(2): p. 436-43.
18. Hillier, T.A. and K.L. Pedula, *Characteristics of an adult population with newly diagnosed type 2 diabetes: the relation of obesity and age of onset.* Diabetes Care, 2001. **24**(9): p. 1522-7.

19. Dart, A.B., et al., *Earlier onset of complications in youth with type 2 diabetes.* Diabetes care, 2014. **37**(2): p. 436-443.
20. Hillier, T.A. and K.L. Pedula, *Complications in young adults with early-onset type 2 diabetes: losing the relative protection of youth.* Diabetes care, 2003. **26**(11): p. 2999-3005.
21. Lee, S.C., et al., *Factors predicting the age when type 2 diabetes is diagnosed in Hong Kong Chinese subjects.* Diabetes Care, 2001. **24**(4): p. 646-9.
22. Ripsin, C.M., H. Kang, and R.J. Urban, *Management of blood glucose in type 2 diabetes mellitus.* American family physician, 2009. **79**(1): p. 29-36.
23. Sterns, J.D., et al., *Epigenetics and type II diabetes mellitus: underlying mechanisms of prenatal predisposition.* Frontiers in cell and developmental biology, 2014. **2**: p. 15.
24. Kolb, H. and S. Martin, *Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes.* BMC medicine, 2017. **15**(1): p. 131.
25. Chatterjee, S., K. Khunti, and M.J. Davies, *Type 2 diabetes.* The Lancet, 2017. **389**(10085): p. 2239-2251.
26. Lang, I.A., et al., *Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults.* Jama, 2008. **300**(11): p. 1303-1310.
27. Scott RA, L.C., Sharp SJ, Franks PW, Rolandsson O, Drogan D, van der Schouw YT, Ekelund U, Kerrison ND, Ardanaz E, Arriola L, Balkau B, Barricarte A, Barroso I, Bendinelli B, Beulens JW, Boeing H, de Lauzon-Guillain B, Deloukas P, Fagherazzi G, Gonzalez C, Griffin SJ, Groop LC, Halkjaer J, Huerta JM, Kaaks R, Khaw KT, Krogh V, Nilsson PM, Norat T, Overvad K, Panico S, Rodriguez-Suarez L, Romaguera D, Romieu I, Sacerdote C, Sánchez MJ, Spijkerman AM, Teucher B, Tjonneland A, Tumino R, van der A DL, Wark PA, McCarthy MI, Riboli E, Wareham NJ, *The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study.* Diabetologia, 2013. **56**(1): p. 60-69.
28. Meigs, J.B., L.A. Cupples, and P. Wilson, *Parental transmission of type 2 diabetes: the Framingham Offspring Study.* Diabetes, 2000. **49**(12): p. 2201-2207.
29. Medici, F., et al., *Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis.* Diabetologia, 1999. **42**(2): p. 146-150.
30. Candler, T., et al., *Continuing rise of type 2 diabetes incidence in children and young people in the UK.* Diabetic Medicine, 2018. **35**(6): p. 737-744.
31. Zouali, H., et al., *A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene.* Human molecular genetics, 1997. **6**(9): p. 1401-1408.
32. Almgren, P., et al., *Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study.* Diabetologia, 2011. **54**(11): p. 2811.
33. Prasad, R.B. and L. Groop, *Genetics of type 2 diabetes—pitfalls and possibilities.* Genes, 2015. **6**(1): p. 87-123.
34. Stoffers, D.A., V. Stanojevic, and J.F. Habener, *Insulin promoter factor-1 gene mutation linked to early-onset type 2 diabetes mellitus directs expression of a dominant negative isoprotein.* J Clin Invest, 1998. **102**(1): p. 232-41.
35. Hegele, R.A., et al., *The hepatic nuclear factor-1α G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree.* The Journal of Clinical Endocrinology & Metabolism, 1999. **84**(3): p. 1077-1082.

36. Aguilar-Salinas, C.A., et al., *Early-onset type 2 diabetes: metabolic and genetic characterization in the mexican population.* J Clin Endocrinol Metab, 2001. **86**(1): p. 220-6.
37. Li, J., et al., *A Missense Mutation in IRS1 is Associated with the Development of Early-Onset Type 2 Diabetes.* Int J Endocrinol, 2020. **2020**: p. 9569126.
38. Chang, T.J., et al., *Genetic variation of SORBS1 gene is associated with glucose homeostasis and age at onset of diabetes: A SAPPHIRe Cohort Study.* Sci Rep, 2018. **8**(1): p. 10574.
39. Yamada, Y., et al., *Identification of four genes as novel susceptibility loci for early-onset type 2 diabetes mellitus, metabolic syndrome, or hyperuricemia.* Biomedical reports, 2018. **9**(1): p. 21-36.
40. Hamet, P., et al., *PROX1 gene CC genotype as a major determinant of early onset of type 2 diabetes in slavic study participants from Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation study.* J Hypertens, 2017. **35 Suppl 1**(Suppl 1): p. S24-S32.
41. Liu, L., et al., *Mutations in KCNJ11 are associated with the development of autosomal dominant, early-onset type 2 diabetes.* Diabetologia, 2013. **56**(12): p. 2609-18.
42. Ma, L., et al., *PCLO variants are nominally associated with early-onset type 2 diabetes and insulin resistance in Pima Indians.* Diabetes, 2008. **57**(11): p. 3156-60.
43. Lim, D.M., N. Huh, and K.Y. Park, *Hepatocyte nuclear factor 1-alpha mutation in normal glucose-tolerant subjects and early-onset type 2 diabetic patients.* Korean J Intern Med, 2008. **23**(4): p. 165-9.
44. Liju, S., et al., *Impact of type 2 diabetes variants identified through genome-wide association studies in early-onset type 2 diabetes from South Indian population.* Genomics & informatics, 2020. **18**(3).
45. Deeb, S.S., et al., *A Pro12Ala substitution in PPARγ2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity.* Nature genetics, 1998. **20**(3): p. 284-287.
46. Horikawa, Y., et al., *Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus.* Nature genetics, 2000. **26**(2): p. 163-175.
47. Mahajan, A., et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.* Nature genetics, 2014. **46**(3): p. 234.
48. Waters, K.M., et al., *Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups.* PLoS genetics, 2010. **6**(8): p. e1001078.
49. Saxena, R., et al., *Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci.* The American Journal of Human Genetics, 2012. **90**(3): p. 410-425.
50. Scott, L.J., et al., *Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample.* diabetes, 2006. **55**(9): p. 2649-2653.
51. Groves, C.J., et al., *Association analysis of 6,736 UK subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk.* Diabetes, 2006. **55**(9): p. 2640-2644.
52. Dimas, A.S., et al., *Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity.* Diabetes, 2014. **63**(6): p. 2158-2171.
53. Das, S.K. and S.C. Elbein, *The genetic basis of type 2 diabetes.* Cellscience, 2006. **2**(4): p. 100.

54. Burgess, S. and S.G. Thompson, *Use of allele scores as instrumental variables for Mendelian randomization.* International journal of epidemiology, 2013. **42**(4): p. 1134-1144.
55. Keaton, J.M., et al., *A comparison of type 2 diabetes risk allele load between African Americans and European Americans.* Human genetics, 2014. **133**(12): p. 1487-1495.
56. Klimentidis, Y.C., et al., *Multiple Metabolic Genetic Risk Scores and Type 2 Diabetes Risk in Three Racial/Ethnic Groups.* Journal of Clinical Endocrinology & Metabolism, 2014. **99**(9): p. E1814-E1818.
57. Abdullah, N., et al., *Characterizing the genetic risk for Type 2 diabetes in a Malaysian multi-ethnic cohort.* Diabetic Medicine, 2015. **32**(10): p. 1377-1384.
58. Hubáček, J.A., et al., *Different prevalence of T2DM risk alleles in Roma population in comparison with the majority Czech population.* Molecular Genetics & Genomic Medicine, 2020. **8**(9): p. e1361.
59. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, T.C., THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS,. *An EU Framework for National Roma Integration Strategies up to 2020*. Accessed 28 April 2021]; Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0173&from=HR.
60. Werissa, N.A., et al., *SNP-Based Genetic Risk Score Modeling Suggests No Increased Genetic Susceptibility of the Roma Population to Type 2 Diabetes Mellitus.* Genes, 2019. **10**(11): p. 942.
61. Kalaydjieva, L., D. Gresham, and F. Calafell, *Genetic studies of the Roma (Gypsies): a review.* BMC medical genetics, 2001. **2**(1): p. 5.
62. *Communication from the commission to the European Parliament, the Council, the European Economic and Social committee and the commttee of the regions.* An EU Framework for National Roma Integration Strategies up to 2020 2011.
63. Ádány, R., *Roma health is global ill health*. 2014, Oxford University Press.
64. Kósa, K., L. Daragó, and R. Ádány, *Environmental survey of segregated habitats of Roma in Hungary: a way to be empowering and reliable in minority research.* The European Journal of Public Health, 2009. **21**(4): p. 463-468.
65. Sepkowitz, K.A., *Health of the world's Roma population.* The Lancet, 2006. **367**(9524): p. 1707-1708.
66. Arora, V.S., C. Kühlbrandt, and M. McKee, *An examination of unmet health needs as perceived by Roma in Central and Eastern Europe.* The European Journal of Public Health, 2016. **26**(5): p. 737-742.
67. Kühlbrandt, C., et al., *An examination of Roma health insurance status in Central and Eastern Europe.* The European Journal of Public Health, 2014. **24**(5): p. 707-712.
68. McFadden, A., et al., *Gypsy, Roma and Traveller access to and engagement with health services: a systematic review.* The European Journal of Public Health, 2018. **28**(1): p. 74-81.
69. Cook, B., et al., *Revisiting the evidence on health and health care disparities among the Roma: a systematic review 2003–2012.* International journal of public health, 2013. **58**(6): p. 885-911.
70. Tabák, A.G., et al., *Prediabetes: a high-risk state for developing diabetes.* Lancet, 2012. **379**(9833): p. 2279.

71. Kosa, Z., et al., *Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary.* European Journal of Public Health, 2015. **25**(2): p. 299-304.
72. Beljić Živković, T., et al., *Screening for diabetes among Roma people living in Serbia.* Croatian medical journal, 2010. **51**(2): p. 144-150.
73. de Courten, B.V., et al., *Higher prevalence of type 2 diabetes, metabolic syndrome and cardiovascular diseases in gypsies than in non-gypsies in Slovakia.* Diabetes research and clinical practice, 2003. **62**(2): p. 95-103.
74. Nunes, M., et al., *Prevalence of Diabetes Mellitus among Roma Populations—A Systematic Review.* International journal of environmental research and public health, 2018. **15**(11): p. 2607.
75. Chan, J.C., et al., *Diabetes in Asia: epidemiology, risk factors, and pathophysiology.* Jama, 2009. **301**(20): p. 2129-2140.
76. Simko, V. and E. Ginter, *Short life expectancy and metabolic syndrome in Romanies (gypsies) in Slovakia.* Central European journal of public health, 2010. **18**(1): p. 16.
77. Macejova, Z., et al., *The Roma Population Living in Segregated Settlements in Eastern Slovakia Has a Higher Prevalence of Metabolic Syndrome, Kidney Disease, Viral Hepatitis B and E, and Some Parasitic Diseases Compared to the Majority Population.* International journal of environmental research and public health, 2020. **17**(9): p. 3112.
78. Zeljko, H., et al., *Traditional CVD risk factors and socio-economic deprivation in Roma minority population of Croatia.* Collegium antropologicum, 2008. **32**(3): p. 667-676.
79. Dobranici, M., A. Buzea, and R. Popescu, *The cardiovascular risk factors of the Roma (Gypsies) people in Central-Eastern Europe: a review of the published literature.* Journal of medicine and life, 2012. **5**(4): p. 382.
80. Babinska, I., et al., *Is the cardiovascular risk profile of people living in Roma settlements worse in comparison with the majority population in Slovakia?* International journal of public health, 2013. **58**(3): p. 417-425.
81. Piko, P., et al., *Comparative risk assessment for the development of cardiovascular diseases in the Hungarian general and Roma population.* Scientific reports, 2021. **11**(1): p. 3085-3085.
82. Bogdanović, D., et al., *Mortality of Roma population in Serbia, 2002-2005.* Croatian medical journal, 2007. **48**(5): p. 720.
83. Kohler, I.V. and S.H. Preston, *Ethnic and religious differentials in Bulgarian mortality, 1993–98.* Population studies, 2011. **65**(1): p. 91-113.
84. Fiatal, S., et al., *Genetic profiling revealed an increased risk of venous thrombosis in the Hungarian Roma population.* Thrombosis research, 2019. **179**: p. 37-44.
85. Pikó, P., et al., *Genetic factors exist behind the high prevalence of reduced high-density lipoprotein cholesterol levels in the Roma population.* Atherosclerosis, 2017. **263**: p. 119-126.
86. Macejova, Z., et al., *The Roma Population Living in Segregated Settlements in Eastern Slovakia Has a Higher Prevalence of Metabolic Syndrome, Kidney Disease, Viral Hepatitis B and E, and Some Parasitic Diseases Compared to the Majority Population.* International Journal of Environmental Research and Public Health, 2020. **17**(9): p. 3112.
87. Széles, G., et al., *A preliminary evaluation of a health monitoring programme in Hungary.* The European Journal of Public Health, 2005. **15**(1): p. 26-32.

88. Nagy, A., R. Adany, and J. Sandor, *Effect of diagnosis-time and initial treatment on the onset of type 2 diabetes mellitus complications: a population-based representative cross-sectional study in Hungary.* Diabetes Research and Clinical Practice, 2011. **94**(3): p. e65-e67.
89. Szigethy, E., et al., *Epidemiology of the metabolic syndrome in Hungary.* Public Health, 2012. **126**(2): p. 143-149.
90. Kósa, Z., et al., *Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary.* The European Journal of Public Health, 2015. **25**(2): p. 299-304.
91. Ádány, R., et al., *General practitioners' cluster: a model to reorient primary health care to public health services.* The European Journal of Public Health, 2013. **23**(4): p. 529-530.
92. Expert Committee on the, D. and M. Classification of Diabetes, *Report of the expert committee on the diagnosis and classification of diabetes mellitus.* Diabetes Care, 2003. **26 Suppl 1**: p. S5-20.
93. Sandor, J., K. Kosa, and M. Papp, *Capitation-based financing hampers the provision of preventive services in primary health care. Front Public Health 4: 200*. 2016.
94. Agena Bioscience Inc. *Single Nucleotide Polymorphism Detection with the iPLEX® Assay and the MassARRAY® System*. 28 April 2021]; Available from: https://agenabio.com/wp-content/uploads/2015/07/51-20061R1.0-iPLEX-Application-Note_WEB.pdf.
95. Lyssenko, V., et al., *Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes.* The Journal of clinical investigation, 2007. **117**(8): p. 2155-2163.
96. Florez, J.C., et al., *TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program.* N Engl J Med, 2006. **355**(3): p. 241-50.
97. Palmer, N.D., et al., *Association of TCF7L2 gene polymorphisms with reduced acute insulin response in Hispanic Americans.* J Clin Endocrinol Metab, 2008. **93**(1): p. 304-9.
98. Lyssenko, V., et al., *Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes.* J Clin Invest, 2007. **117**(8): p. 2155-63.
99. Villareal, D.T., et al., *TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action.* diabetes, 2010. **59**(2): p. 479-485.
100. Huang, Z.-q., et al., *Possible role of TCF7L2 in the pathogenesis of type 2 diabetes mellitus.* Biotechnology & Biotechnological Equipment, 2018. **32**(4): p. 830-834.
101. Tong, Y., et al., *Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis.* BMC medical genetics, 2009. **10**(1): p. 15.
102. Peng, S., et al., *TCF7L2 gene polymorphisms and type 2 diabetes risk: a comprehensive and updated meta-analysis involving 121 174 subjects.* Mutagenesis, 2012. **28**(1): p. 25-37.
103. Guan, Y., et al., *Correlation of the TCF7L2 (rs7903146) polymorphism with an enhanced risk of type 2 diabetes mellitus: a meta-analysis.* Genet Mol Res, 2016. **15**(3).
104. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nature genetics, 2010. **42**(2): p. 105.
105. Ding, W., et al., *Meta-analysis of association between TCF7L2 polymorphism rs7903146 and type 2 diabetes mellitus.* BMC medical genetics, 2018. **19**(1): p. 38.

106. Cauchi, S., et al., *TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis.* Journal of molecular medicine, 2007. **85**(7): p. 777-782.
107. Peng, F., et al., *The relationship between five widely-evaluated variants in CDKN2A/B and CDKAL1 genes and the risk of type 2 diabetes: a meta-analysis.* Gene, 2013. **531**(2): p. 435-443.
108. Li, H., et al., *Association of glucokinase regulatory protein polymorphism with type 2 diabetes and fasting plasma glucose: a meta-analysis.* Molecular biology reports, 2013. **40**(6): p. 3935-3942.
109. Cugino, D., et al., *Type 2 diabetes and polymorphisms on chromosome 9p21: a meta-analysis.* Nutrition, Metabolism and Cardiovascular Diseases, 2012. **22**(8): p. 619-625.
110. Bao, X.Y., C. Xie, and M.S. Yang, *Association between type 2 diabetes and CDKN2A/B: a meta-analysis study.* Molecular biology reports, 2012. **39**(2): p. 1609-1616.
111. Chen, X., et al., *Association of the ADRA2A polymorphisms with the risk of type 2 diabetes: A meta-analysis.* Clinical biochemistry, 2013. **46**(9): p. 722-726.
112. Steinthorsdottir, V., et al., *A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.* Nature genetics, 2007. **39**(6): p. 770-775.
113. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-1341.
114. Liang, Y., et al., *Correlation between CDKAL1 rs10946398C> A single nucleotide polymorphism and type 2 diabetes mellitus susceptibility: A meta-analysis.* Open Life Sciences, 2017. **12**(1): p. 501-509.
115. Dehwah, M., M. Wang, and Q. Huang, *CDKAL1 and type 2 diabetes: a global meta-analysis.* Genet Mol Res, 2010. **9**(2): p. 1109-20.
116. Rudovich, N.N., H.J. Rochlitz, and A.F. Pfeiffer, *Reduced hepatic insulin extraction in response to gastric inhibitory polypeptide compensates for reduced insulin secretion in normal-weight and normal glucose tolerant first-degree relatives of type 2 diabetic patients.* Diabetes, 2004. **53**(9): p. 2359-2365.
117. Bort, R., et al., *Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas.* Development, 2004. **131**(4): p. 797-806.
118. Cai, Y., et al., *Meta-analysis of the effect of HHEX gene polymorphism on the risk of type 2 diabetes.* Mutagenesis, 2010. **26**(2): p. 309-314.
119. Li, X., et al., *Hematopoietically-expressed homeobox gene three widely-evaluated polymorphisms and risk for diabetes: a meta-analysis.* PLoS One, 2012. **7**(11): p. e49917.
120. Wang, D.-d., et al., *Association of Kir6. 2 gene rs5219 variation with type 2 diabetes: A meta-analysis of 21,464 individuals.* Primary care diabetes, 2018. **12**(4): p. 345-353.
121. Qin, L., Y. Lv, and Q. Huang, *Meta-analysis of association of common variants in the KCNJ11-ABCC8 region with type 2 diabetes.* Genet Mol Res, 2013. **12**(3): p. 2990-3002.
122. Gong, B., et al., *The effect of KCNJ11 polymorphism on the risk of type 2 diabetes: a global meta-analysis based on 49 case-control studies.* DNA and cell biology, 2012. **31**(5): p. 801-810.
123. Qiu, L., et al., *Quantitative assessment of the effect of KCNJ11 gene polymorphism on the risk of type 2 diabetes.* PloS one, 2014. **9**(4): p. e93961.

124. Sokolova, E.A., et al., *Replication of KCNJ11 (p. E23K) and ABCC8 (p. S1369A) association in Russian diabetes mellitus 2 type cohort and meta-analysis.* PLoS One, 2015. **10**(5): p. e0124662.
125. Wang, T., et al., *The effect of glucose-dependent insulinotropic polypeptide (GIP) variants on visceral fat accumulation in Han Chinese populations.* Nutrition & diabetes, 2017. **7**(5): p. e278-e278.
126. Wang, H., et al., *Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk.* PloS one, 2013. **8**(6): p. e67665.
127. Yin, X., *MTNR1B gene polymorphisms are associated with the therapeutic responses to repaglinide in Chinese patients with type 2 diabetes mellitus.* Frontiers in pharmacology, 2019. **10**: p. 1318.
128. Zheng, C., et al., *A common variant in the MTNR1b gene is associated with increased risk of impaired fasting glucose (IFG) in youth with obesity.* Obesity (Silver Spring), 2015. **23**(5): p. 1022-9.
129. Sparso, T., et al., *G-allele of intronic rs10830963 in MTNR1B confers increased risk of impaired fasting glycemia and type 2 diabetes through an impaired glucose-stimulated insulin release: studies involving 19,605 Europeans.* Diabetes, 2009. **58**(6): p. 1450-6.
130. Xia, Q., et al., *Association between the melatonin receptor 1B gene polymorphism on the risk of type 2 diabetes, impaired glucose regulation: a meta-analysis.* PloS one, 2012. **7**(11): p. e50107.
131. Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.* Nature genetics, 2010. **42**(7): p. 579.
132. Takeda, Y. and A.M. Jetten, *Prospero-related homeobox 1 (Prox1) functions as a novel modulator of retinoic acid-related orphan receptors α-and γ-mediated transactivation.* Nucleic acids research, 2013. **41**(14): p. 6992-7008.
133. Boesgaard, T., et al., *Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people.* Diabetologia, 2010. **53**(8): p. 1647-1655.
134. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nature genetics, 2010. **42**(2): p. 105-116.
135. Huang, M.-C., et al., *FADS gene polymorphisms, fatty acid desaturase activities, and HDL-C in type 2 diabetes.* International Journal of Environmental Research and Public Health, 2017. **14**(6): p. 572.
136. Warensjö, E., et al., *Associations between estimated fatty acid desaturase activities in serum lipids and adipose tissue in humans: links to obesity and insulin resistance.* Lipids in health and disease, 2009. **8**(1): p. 37.
137. Cormier, H., et al., *Polymorphisms in Fatty Acid Desaturase (FADS) gene cluster: Effects on glycemic controls following an omega-3 Polyunsaturated Fatty Acids (PUFA) supplementation.* Genes, 2013. **4**(3): p. 485-498.
138. Rosengren, A.H., et al., *Overexpression of alpha2A-adrenergic receptors contributes to type 2 diabetes.* Science, 2010. **327**(5962): p. 217-220.
139. Hong, K.W., M. Chung, and S.B. Cho, *Meta-analysis of genome-wide association study of homeostasis model assessment beta cell function and insulin resistance in an East Asian population and the European results.* Mol Genet Genomics, 2014. **289**(6): p. 1247-55.

140. Barrett, J.C., et al., *Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes.* Nat Genet, 2009. **41**(6): p. 703-7.
141. Boesgaard, T.W., et al., *Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people.* Diabetologia, 2010. **53**(8): p. 1647-55.
142. Heni, M., et al., *The impact of genetic variation in the G6PC2 gene on insulin secretion depends on glycemia.* The Journal of Clinical Endocrinology & Metabolism, 2010. **95**(12): p. E479-E484.
143. Li, L.-c., et al., *IG20/MADD plays a critical role in glucose-induced insulin secretion.* Diabetes, 2014. **63**(5): p. 1612-1623.
144. Liang, X., et al., *Integrating Genome-Wide Association and eQTLs Studies Identifies the Genes and Gene Sets Associated with Diabetes.* Biomed Res Int, 2017. **2017**: p. 1758636.
145. Bieganowski, P., et al., *CDC123 and checkpoint forkhead associated with RING proteins control the cell cycle by controlling eIF2γ abundance.* Journal of Biological Chemistry, 2004. **279**(43): p. 44656-44666.
146. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* The American journal of human genetics, 2007. **81**(3): p. 559-575.
147. Gauderman, A., *QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.* http://hydra. usc. edu/gxe, 2006.
148. Templeton, G.F., *A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research.* Communications of the Association for Information Systems, 2011. **28**(1): p. 4.
149. Jonckheere, A.R., *A Distribution-Free k-Sample Test Against Ordered Alternatives.* Biometrika, 1954. **41**(1/2): p. 133–145.
150. Talmud, P.J., et al., *Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.* Bmj, 2010. **340**.
151. Sebastiani, P., N. Solovieff, and J. Sun, *Naïve Bayesian classifier and genetic risk score for genetic risk prediction of a categorical trait: not so different after all!* Frontiers in genetics, 2012. **3**: p. 26.
152. Wallace, B.C., et al., *Closing the gap between methodologists and end-users: R as a computational back-end.* J Stat Softw, 2012. **49**(5): p. 1-15.
153. Salanti, G., et al., *Underlying genetic models of inheritance in established type 2 diabetes associations.* Am J Epidemiol, 2009. **170**(5): p. 537-45.
154. Wilmot, E. and I. Idris, *Early onset type 2 diabetes: risk factors, clinical impact and management.* Therapeutic advances in chronic disease, 2014. **5**(6): p. 234-244.
155. de Miguel-Yanes, J.M., et al., *Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms.* Diabetes care, 2011. **34**(1): p. 121-125.
156. Nguyen, Q.M., et al., *Correlates of age onset of type 2 diabetes among relatively young black and white adults in a community: the Bogalusa Heart Study.* Diabetes care, 2012. **35**(6): p. 1341-1346.
157. Zhou, K., et al., *Clinical and genetic determinants of progression of type 2 diabetes: a DIRECT study.* Diabetes Care, 2014. **37**(3): p. 718-724.
158. Smith, A.D., et al., *Physical activity and incident type 2 diabetes mellitus: a systematic review and dose–response meta-analysis of prospective cohort studies*. 2016, Springer.

159. Guh, D.P., et al., *The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis.* BMC public health, 2009. **9**(1): p. 88.
160. Sami, W., et al., *Effect of diet on type 2 diabetes mellitus: A review.* International journal of health sciences, 2017. **11**(2): p. 65.
161. Sándor, J., et al., *The decade of Roma Inclusion: did it make a difference to health and use of health care services?* International Journal of Public Health, 2017. **62**(7): p. 803-815.
162. Carrasco-Garrido, P., et al., *Health status of Roma women in Spain.* European Journal of Public Health, 2011. **21**(6): p. 793-798.
163. Dolinska, S., M. Kudlackova, and E. Ginter, *The prevalence of female obesity in the world and in the Slovak Gypsy women.* Bratislavske lekarske listy, 2007. **108**(4-5): p. 207-211.
164. Sedova, L., et al., *Qualification of Food Intake by the Roma Population in the Region of South Bohemia.* International journal of environmental research and public health, 2018. **15**(2): p. 386.
165. Hoxha, A., et al., *Assessment of nutritional status and dietary patterns of the adult Roma community in Albania.* AMJ AMJ, 2013. **3**: p. 32.
166. Saeedi, P., et al., *Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas.* Diabetes research and clinical practice, 2019. **157**: p. 107843.
167. Piko, P., et al., *Changes in the Prevalence of Metabolic Syndrome and Its Components as Well as in Relevant Preventive Medication between 2006 and 2018 in the Northeast Hungarian Population.* Journal of Personalized Medicine, 2021. **11**(1): p. 52.
168. Centers for Disease Control and Prevention. *National Diabetes Statistics Report*. 2020 [Accessed on February 10, 2021]; Available from: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf.
169. Ogawa, Y., et al., *Proportion of diabetes type in early-onset diabetes in Japan.* Diabetes Care, 2007. **30**(5): p. e30-e30.
170. Pan, J. and W. Jia, *Early-onset diabetes: an epidemic in China.* Frontiers of medicine, 2018. **12**(6): p. 624-633.
171. Silbernagel, G., et al., *Association of TCF7L2 SNPs with age at onset of type 2 diabetes and proinsulin/insulin ratio but not with glucagon-like peptide 1.* Diabetes/metabolism research and reviews, 2011. **27**(5): p. 499-505.
172. Iwata, M., et al., *Genetic risk score constructed using 14 susceptibility alleles for type 2 diabetes is associated with the early onset of diabetes and may predict the future requirement of insulin injections among Japanese individuals.* Diabetes Care, 2012. **35**(8): p. 1763-70.
173. Kong, X., et al., *Early-onset type 2 diabetes is associated with genetic variants of β-cell function in the Chinese Han population.* Diabetes/metabolism research and reviews, 2019: p. e3214.
174. Langenberg, C., et al., *Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study.* PLoS Med, 2014. **11**(5): p. e1001647.
175. Mars, N., et al., *Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers.* Nature Medicine, 2020. **26**(4): p. 549-557.

176. Kautzky-Willer, A., J. Harreiter, and G. Pacini, *Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus.* Endocr Rev, 2016. **37**(3): p. 278-316.
177. Walford, G.A., et al., *Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes.* Diabetes Care, 2014. **37**(9): p. 2508-14.
178. Gan, W., et al., *Evaluation of type 2 diabetes genetic risk variants in Chinese adults: findings from 93,000 individuals from the China Kadoorie Biobank.* Diabetologia, 2016. **59**(7): p. 1446-1457.
179. Hajioff, S. and M. McKee, *The health of the Roma people: a review of the published literature.* Journal of epidemiology & community health, 2000. **54**(11): p. 864-869.
180. Assal, S., E. Susanszky, and A. Czeizel, *High consanguinity rate in Hungarian gipsy communities.* Acta Paediatrica Hungarica, 1991. **31**(3): p. 299-304.
181. Yuan, T., et al., *Glucose tolerance and insulin responsiveness in Gitelman syndrome patients.* Endocrine connections, 2017. **6**(4): p. 243-252.

8. Key words

Type 2 diabetes mellitus, genetic risk score, Roma, targeted intervention, single nucleotide polymorphism, age of onset for type 2 diabetes, Hungarian

9. Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Dr. Róza Ádány for her tremendous guidance, encouragement, and help throughout this dissertation work.

I would like to extend my sincere gratitude to Péter Pikó for his benevolence, assistance, fruitful and valuable comments, suggestions, guidance, and for all his backing in realizing of this dissertation. My heartfelt gratefulness also goes to Dr. Szilvia Fiatal for her insightful comments and suggestions in this paper.

I am also indebted to the Head of the Department of Public Health and Epidemiology, Professor János Sándor, who gave me the opportunity to carry out my research. Special thanks to Koós Zoltánné and all staff members of Department of Public Health and Epidemiology from whom I received an endless support and assistance.

My earnest thanks to my mother, Tsega Bekele Desta, my brothers Berhanu Abebe and Ahadu Kassahun and my sister Amelework Abebe for their interminable encouragements throughout my PhD study.

Lastly but not least, I would like to dedicate this dissertation to my father, Abebe Werissa Molla sole who always encouraged and inspired me to carry on my education.

10. Funding

The project was co-financed by the European Union under the European Social Fund (TÁMOP 4.2.1. B-09/1/KONV-2010-0007 and TÁMOP 4.2.2. A-11/1/KONV-2012-0031) and European Regional Development Fund (GINOP-2.3.2-15-2016-00005), as well as by the Hungarian Academy of Sciences (MTA11010, TK2016-78)

11. Publications

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/89/2021.PL
Subject: PhD Publication List

Candidate: Nardos Abebe Werissa
Doctoral School: Doctoral School of Health Sciences
MTMT ID: 10060836

List of publications related to the dissertation

1. Pikó, P., **Werissa, N. A.**, Fiatal, S., Sándor, J., Ádány, R.: Impact of Genetic Factors on the Age of Onset for Type 2 Diabetes Mellitus in Addition to the Conventional Risk Factors.
   JPM. 11 (1), 6-, 2020.
   DOI: http://dx.doi.org/10.3390/jpm11010006
   IF: 4.433 (2019)
2. **Werissa, N. A.**, Pikó, P., Fiatal, S., Kósa, Z., Sándor, J., Ádány, R.: SNP-Based Genetic Risk Score Modeling Suggests No Increased Genetic Susceptibility of the Roma Population to Type 2 Diabetes Mellitus.
   Genes. 10 (11), 1-16, 2019.
   DOI: http://dx.doi.org/10.3390/genes10110942
   IF: 3.759

List of other publications

3. Pikó, P., Fiatal, S., **Werissa, N. A.**, Begashaw, B., Rácz, G., Kósa, Z., Sándor, J., Ádány, R.: The Effect of Haplotypes in the CETP and LIPC Genes on the Triglycerides to HDL-C Ratio and Its Components in the Roma and Hungarian General Populations.
   Genes. 11 (56), 1-13, 2020.
   DOI: http://dx.doi.org/10.3390/genes11010056
   IF: 3.759 (2019)

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár

Total IF of journals (all publications): 11,951
Total IF of journals (publications related to the dissertation): 8,192

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

09 March, 2021