SHORT THESIS 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

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

By Nardos Abebe Werissa, MSc

Supervisor: Róza Ádány, MD, PhD, DSc

Doctoral School of Health Sciences, University of Debrecen

Head of the Defense Committee: Endre Nagy, MD, PhD, DSc

Reviewers: László Márk, MD, PhD

Lajos Széles, PhD

Members of the Defense Committee: István Kiss, MD, PhD, DSc

Péter Fülöp, MD, PhD

The PhD Defense is held online and starts at 9:00 on 18th June, 2021

Publicity is provided online. If you wish to participate, please indicate it in a message sent to the e-mail address egdi@unideb.hu by 2 p.m. on 17 February 2021latest. After the deadline, it is no longer possible to connect to the PhD defense due to technical reasons.

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.

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 type 2 diabetes in adolescents and young adults is dramatically increasing…. raising the possibility of a future public health catastrophe”. Early onset of type 2 diabetes mellitus (EOT2DM) is defined as diagnosis at below 45 years of age and countries – among them China and the United Kingdom- documented an increase in the incidence and consequently the prevalence of EOT2DM. Moreover, Japan and Taiwan, 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. T2DM and its complications contribute massively to the mortality and disability globally. 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. Diabetes, next to HIV/AIDS, has the second biggest negative effect on reducing health adjusted life expectancy across the globe. 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. T2DM and its complications significantly affect the quality of life and exert a major burden on individuals, economy, and the healthcare system. 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. For instance, macrovascular and microvascular complications, cardiovascular disease and microalbuminuria are more common among diabetes patients with EOT2DM compared with the usual onset of type 2 diabetes mellitus (UOT2DM) (defined as diagnosis at ≥45 years of age). In general, individuals with EOT2DM have significantly poorer metabolic profiles than individuals with UOT2DM.

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

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. Environmental toxins, noise, increased exposure to residential traffic, and fine airborne particulate matter may also contribute to the development of T2DM.

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. 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. In addition, few studies also uncovered the existence of genetic factors on the age of onset of T2DM. 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.

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 but also, have uncovered several SNPs that influence the age of onset for T2DM. In 1998, the first candidate gene, PPARG which encodes the nuclear receptor PPAR-γ was identified to reproducibly associated with T2DM in Finnish population. 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. 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 and also observed on ethnically diverse populations. The majority of these SNPs exert their effect on the disease risk through deficient insulin secretion and few of them through insulin resistance.

Almost all these discovered SNPs separately have modest effects on the risk of T2DM (odds ratios $\leq$1.4); 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. 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, 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.

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. Approximately six million of Roma live in the European Union. Roma arrived in the Balkans from North India in the Xth century and then migrated to Europe in three migration waves. Currently, this minority group is clustered in the Central and Eastern European countries, largely in Bulgaria, North Macedonia, Hungary, Slovakia and Romania. 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, unhealthy living conditions, low life expectancy, severely limited access to health services, and discrimination, which are closely linked to a low level of education, a high rate of unemployment, and their low socio-economic status in general.

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 - and T2DM was shown in a previous study which compared PreDM and T2DM between Hungarian Roma and majority population (27.09% vs. 15.56%; $p<0.001$). 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%) and in Slovakia (30% vs. 10%; $p=0.0001$) was also reported. A 25% of higher prevalence of T2DM in Roma population compared to the Czech general population was also reported by the government of the Czech Republic.

However, the latest review by Nunes et al on publications related to the prevalence of diabetes mellitus in the Roma population 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.

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. 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, as well as increased CVD risk and significantly higher mortality among Roma. In addition, the higher prevalence of T2DM and genetically modified disturbances in other cardiometabolic traits were also detected in the Roma populations in Europe.

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.

Materials and methods

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

Study design

The current study consists of data assembled from previous three surveys involving 1168 individuals representative of Hungarian T2DM population (case population), 1783 individuals representative of Hungarian general population and 1260 individuals representative of Roma living in segregated colonies in North-East Hungary, where they mainly concentrated.

Samples

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. 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. 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. 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

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. 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.

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

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. 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.

Additional samples were obtained in the framework (Survey 3) of the Public Health Focused Model Program for Organizing Primary Care Services in 2013. 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. Within this framework further 535 samples from the Roma population dwelling in North-East of Hungary were collected, totally making the Roma sample 1260.

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).

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.

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.

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 protocol.

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)

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.

Statistical analysis

A $\chi 2$ test was used to assess whether the agreement of frequencies of genotypes for SNPs with Hardy-Weinberg equilibrium (HWE) expectations (by Plink software). Linkage disequilibrium (LD) between polymorphisms was tested by Haploview software (version 4.2). The r2 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. 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. 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 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 ($p<0.0042$).

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. 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 ith 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)

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

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

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.

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 r2 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.

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.

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.

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

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.

Results

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. 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.

Results of power calculations for the Hungarian general and Roma populations

The statistical power for individual SNPs was between 5.03% and 12.79%.

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. To avoid multicollinearity, only one SNP per LD block was used in the GRS calculation.

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. 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.

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.

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.

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 ($\beta=0.053$, $p=0.001$; OR=1.070, $p=0.027$) and in the Roma ($\beta=0.044$, $p=0.037$; OR=1.083, $p=0.010$) populations.

In the wGRS model the association was significant for both FG level and T2DM status in the Hungarian general population ($\beta=0.489$, $p<0.001$; 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 ($\beta=0.300$, $p=0.100$).

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. 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. 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, $p<0.001$).

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.

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; $\beta=-0.866$, $p=0.073$) to the weakest (rs2191349; $\beta=-0.004$, $p=0.990$). 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.

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: $\beta=2.352$, $p<0.001$).

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.

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.

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 ($\beta=-0.999$, $p=0.003$) 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.

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.

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, 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. 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.

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. 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, there is a decreasing trend in the share of genetic risk factors.

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, which shows a striking resemblance to the findings of the study carried out by Zhou et al. on a sample of the Scottish population.

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.

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 examined the allelic differences between the Czech general 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 general 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.

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, overweight or obesity and unhealthy diet 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 general population, irrespective of the country where they live. 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. 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. Several countries also reported an increasing prevalence of T2DM in younger adults and even in adolescents. 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 . 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.

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. 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. 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.

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). 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. 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, 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.

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. 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. 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. 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. 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.

New findings

Significant differences in allele frequencies were found for eight SNPs

Significant differences in allele frequencies were found for eight SNPs; five susceptible SNPs were more prevalent in Hungarian general population and 3 susceptible SNPs were more frequent in the Roma population.

Hungarian general carried greater risk allele load for the development of T2DM

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. 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.

Both GRSs and wGRSs were significantly associated with FG and T2DM status in the General population, but this association was modest in the case of the Roma population.

The GRS was significantly associated with both outcomes in the adjusted model both in the Hungarian general ($\beta=0.053$, $p=0.001$; OR=1.070, $p=0.027$) and in the Roma ($\beta=0.044$, $p=0.037$; OR=1.083, $p=0.010$) populations. In the case of wGRS model the association was significant for both FG level and T2DM status in the Hungarian general population ($\beta=0.489$, $p<0.001$; 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 ($\beta=0.300$, $p=0.100$).

In addition to GRS, Roma ethnicity has been identified as a risk factor for elevated fasting glucose and T2DM

When the two populations were analyzed together (ethnicity was integrated into the model as a covariate beside age, sex, BMI, HDL-C, TG, and GRSs), the ethnicity (FG levels: β ethnicity=0.918, $p<0.001$; T2DM status: OR ethnicity=2.484, $p<0.001$).and the GRSs had significant effects on both outcomes. The effect of Roma ethnicity was relatively strong on both outcomes compared with unweighted models (FG: β GRS=0.050, $p<0.001$; T2DM status: OR GRS=1.075, $p=0.001$), as well as weighted models (FG: β wGRS=0.425, $p<0.001$; T2DM status: OR wGRS=2.128, $p<0.001$).

The GRS showed a significant association with the age of onset for T2DM in the T2DM case population.

The GRS showed significant association with the age of onset for T2DM in the full case population: β GRS =-0.454, p<0.001as well as in male β GRS =-0.434, P=0.003 and female β GRS =-0.405, p=0.008. Individuals who developed T2DM at early age carried greater risk allele load compared with individuals who developed it at later age.

The GRS obtained in the Case population was replicated on the T2DM subpopulation of Hungarian general population

A significant correlation between patients' age and GRS (β=−0.999, p=0.003) was detected only in the subpopulation with T2DM. To estimate the age of onset for T2DM in Hungarian general population, AORS was calculated and when it was evaluated among the three subpopulations, the risk score was significantly higher in the T2DM subpopulation. 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 observed. 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 earlier the age of onset for T2DM is, the higher the GRS and wGRS are.

In conclusion, 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.

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.

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)

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)

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

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár

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