

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 Diamicon MR Controlled Evaluation study

Pavel Hamet^{a,b,*†}, Mounsi Haloui^b, François Harvey^b, François-Christophe Marois-Blanchet^b, Marie-Pierre Sylvestre^c, Muhammad-Ramzan Tahir^b, Paul H.G. Simon^b, Beatriz Sonja Kanzki^d, John Raelson^b, Carole Long^b, John Chalmers^e, Mark Woodward^e, Michel Marre^f, Stephen Harrap^g, and Johanne Tremblay^{a,b,*†}

Background: The prevalence of diabetic nephropathy varies according to ethnicity. Environmental as well as genetic factors contribute to the heterogeneity in the presentation of diabetic nephropathy. Our objective was to evaluate this heterogeneity within the Caucasian population.

Methods: The geo-ethnic origin of the 3409 genotyped Caucasian type 2 diabetes (T2D) patients of Action in Diabetes and Vascular Disease: Preterax and Diamicon MR Controlled Evaluation was determined using principal component analysis. Genome-wide association studies analyses of age of onset of T2D were performed for geo-ethnic groups separately and combined.

Results: The first principal component separated the Caucasian study participants into Slavic and Celtic ethnic origins. Age of onset of diabetes was significantly lower in Slavic patients ($P = 7.3 \times 10^{-20}$), whereas the prevalence of hypertension ($P = 4.9 \times 10^{-31}$) and albuminuria (5.1×10^{-9}) were significantly higher. Age of onset of T2D and albuminuria appear to have an important genetic component as the values of these traits were also different between Slavic and Celtic individuals living in the same countries. Common and geo-ethnic-specific loci were found to be associated to age of onset of diabetes. Among the latter, the *PROX1/PROX1-AS1* genes (rs340841) had the highest impact. Single-nucleotide polymorphism rs340841 CC genotype was associated with a 4.4 year earlier onset of T2D in Slavic patients living or not in countries with predominant Slavic populations.

Conclusion: These results reveal the presence of distinct genetic architectures between Caucasian ethnic groups that likely have clinical relevance, among them *PROX1* gene is a strong candidate of early onset of diabetes with variations depending on ethnicity.

Keywords: albuminuria, diabetic kidney disease, environment, ethnic groups, genetics

Abbreviations: ADVANCE, Action in Diabetes and Vascular Disease Preterax and Diamicon MR Controlled Evaluation; CKD, chronic kidney disease; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; eGFR, estimated glomerular filtration rate; GWAS, genome-wide association studies; MAF, minor allele frequency; PCA, principal component analysis; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes; UACR, urinary albumin-creatinine ratio

INTRODUCTION

The incidence of type 2 diabetes (T2D) is increasing even in younger study participants in both industrialized and economic transition countries totaling 415 million study participants worldwide in 2015 [1]. The major increased risk of mortality associated with both type 1

Journal of Hypertension 2017, 35 (Suppl 1):S24–S32

^aDepartment of Medicine, Gene Medicine Services, ^bDepartment of Medicine, ^cDepartment of Social and Preventive Medicine, Université de Montréal, CRCHUM, ^dDepartment of Software Engineering and the Information Technology, École de Technologie Supérieure, Montreal, Quebec, Canada, ^eThe George Institute for Global Health, University of Sydney, Sydney, New South Wales, Australia, ^fHôpital Bichat, Claude Bernard, Université Paris 7, Paris, France and ^gRoyal Melbourne Hospital, University of Melbourne, Melbourne, Victoria, Australia

Correspondence to Pavel Hamet, OQ, MD, PhD, FRCPC, Centre hospitalier de l'Université de Montréal, 900 Saint-Denis, room R14.404, Montréal, Québec, Canada, H2X 0A9. Tel: +1 514 890 8246; e-mail: pavel.hamet@umontreal.ca

*Pavel Hamet and Johanne Tremblay are members of College International de Recherche Servier and Johanne Tremblay received honoraria for presentations of ADVANCE at scientific meetings.

†Pavel Hamet and Johanne Tremblay contributed equally to this work.

Received 22 August 2016 **Revised** 10 November 2016 **Accepted** 5 December 2016

J Hypertens 35 (Suppl 1):S24–S32 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/HJH.0000000000001241

diabetes and T2D arises from diabetic nephropathy [2–4], which is estimated to affect about one-third of individuals with diabetes.

Different genetic architectures, such as variations in allele frequencies and linkage disequilibrium structure have long been noted between populations of different racial origins and the ability of this hidden population structure to confound genome-wide association studies (GWAS) findings has been well documented [5–8]. Moreover, it is known that GWAS using populations of differing racial backgrounds may help identify different sets of associated genes for complex diseases and drug responses [9–11]. Differences in genetic risks have been shown among Caucasians, Africans, and Asians for T2D [12–16] and for chronic kidney disease (CKD) [17–19]. Evidence exists for population substructure within Caucasian samples as well [20,21].

It is also well established that both environmental and genetic factors contribute to the occurrence of hypertension, diabetes, and CKD [22–25]. To distinguish the effects of environmental and lifestyle factors from genetic effects in explaining phenotypic differences in the development of renal complications of T2D, we studied T2D study participants of Caucasian origin and European descent from the Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation trial (ADVANCE) [26]. Study participants were recruited from a range of European countries as well as from countries of European settlements such as Canada, Australia, and New Zealand. Using principal component analysis (PCA), we identified two main ethnic genetic profiles (Celtic and Slavic) within the ADVANCE Caucasian study participants. To assess the relative effects of genetic and environmental factors, we compared study participants with a Slavic genetic profile living in countries with predominantly Celtic populations with individuals with a Slavic genetic profile living in predominantly Slavic countries of Europe. Significant differences between Slavs living in Slavic and Celtic countries would suggest an environmental/lifestyle effect, whereas no differences between Slavs living in Celtic or Slavic countries would support an impact of genetic influence.

As age of onset of T2D appears to be more dependent on genetic than environmental factors, we performed GWAS for ‘age of onset of T2D’ within the two ethnic groups separately and for the combined sample.

METHODS

Sample

In total, 11 140 participants recruited from 215 centers in 20 countries who were 55 years or older and had T2D since the age of 30 years or older were enrolled in ADVANCE, a factorial randomized controlled clinical trial of blood pressure (BP) lowering and intensive glucose control. All participants were ascertained for high outcome risk according to one of the following criteria: a history of major macrovascular or microvascular disease or diagnosis of T2D 10 years prior to entry in study or presence of another major risk factor for vascular disease, including smoking, dyslipidemia or microalbuminuria, or being 65 years or older. Detailed study methods have been published elsewhere [26].

Approval to conduct the trial was obtained from the ethics committee of each study center, and all participants provided written informed consent for the study conduct and a specific, separate consent for genetic substudy. Genotyping was performed only in patients who consented to the genetic substudy.

Complication phenotypes

Several phenotypes associated with diabetes and its complications were determined at baseline in each study participants by the ADVANCE study team. These included age at baseline, age at diagnosis of T2D, duration of diabetes at baseline, BMI, blood glucose, glycated hemoglobin, treatment for hypertension, heart rate, and SBP and DBP. Renal phenotypes included estimated glomerular filtration rate (eGFR), in ml/min per 1.73 m², estimated from serum creatinine levels using the CKD-Epidemiology Collaboration formula [27] and albuminuria, expressed as a ratio of urinary albumin and creatinine in µg/mg [urinary albumin–creatinine ratio (UACR)].

Genotyping

In this study, we have genotyped 3629 Caucasian study participants using the Affymetrix Genome-Wide Human SNP Arrays 5.0 or 6.0 (Affymetrix, Santa Clara, California, USA) following standard protocols recommended by the manufacturer. A quality control filtering step was applied to the genotype calls. The microarray data was analyzed using the Affymetrix power tools and individuals with a quality control call rate lower than 86% were filtered out. Additional quality control steps included coarse-grain stratification to ensure a Caucasian population ratio more than 0.8 (STRUCTURE software [28]), a genetic relatedness check to ensure independent samples (PLINK) and a sex check to ensure genetic accuracy and database integrity [29]. Quality control was also performed on the final genotypes to remove any single-nucleotide polymorphism (SNPs) with more than 4% of missing values across the entire cohort and any sample with more than 2% of missing SNP genotypes. A more stringent threshold was used for any SNPs with between 1 and 5% minor allele frequencies (MAF). Any of these low MAF SNPs with more than 1% of missing values was removed prior to the imputation; nonetheless, only SNPs with MAF higher than 5% were retained after imputation for use in the GWAS. After completion of the quality control process, a total of 3409 genotyped individuals remained available for analysis.

Principal component analysis

A subset of 139 186 independent SNPs was selected from the set of common genotyped SNPs from 5.0 and 6.0 arrays using the linkage disequilibrium pruning application sub-routine from PLINK. This set of SNPs was used to perform a PCA for the ADVANCE study participants of Caucasian origin using the EIGENSOFT 3.0 package [7]. The first principal component (PC1) was used to characterize the ethnic profiles of individuals from this Caucasian population that was sampled in European countries ranging east to west from Russia to Ireland and also from countries with populations of European descent, including Canada, Australia, and New Zealand.

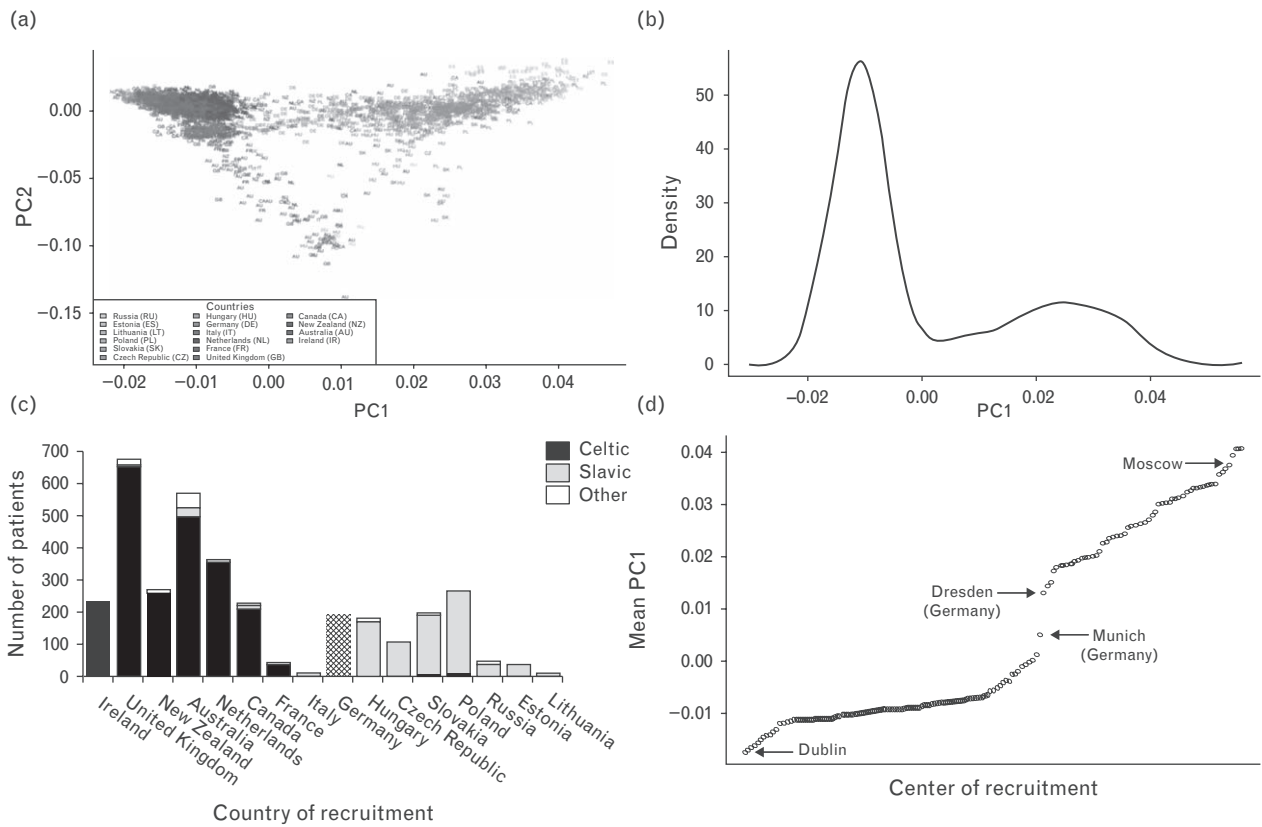


FIGURE 1 Distribution of genotyped ADVANCE study participants ($n = 3409$) according to principal components of genotype structure using EIGENSOFT 3.0 package. (a) ADVANCE individuals are plotted against the first two principal components PC1 (west–east gradient) and PC2 (north–south gradient), (b) frequency distribution of study participants by value of PC1. (c) Distribution of principal component values by countries of recruitment of patients in ADVANCE. (d) ADVANCE recruitment centers ordered by mean value of PC1.

TABLE 1. Demographic and clinical characteristics at baseline of Caucasian ADVANCE genotyped study participants stratified by ethnic origin

Trait	All ($n = 3409$) Mean (SD) or %	Celtic ($n = 2307$) Mean (SD) or %	Slavic ($n = 1102$) Mean (SD) or %	P Value
Age (years)	67.3 (6.6)	68.0 (6.6)	65.9 (6.6)	1.9×10^{-16}
Men (sex)	64.7	69.8	54.0	5.8×10^{-19}
Age at diagnosis of diabetes (years)	60.1 (8.5)	61.0 (6.1)	58.2 (6.1)	7.3×10^{-20}
Diabetes duration (years)	6.7 (6.1)	6.4 (6.1)	7.4 (6.1)	2.9×10^{-6}
BMI	30.1 (5.1)	30.1 (5.0)	30.0 (5.0)	7.6×10^{-1}
Blood glucose assessment				
HbA1c (%)	13.4 (2.7)	13.4 (2.8)	13.4 (2.8)	7.5×10^{-1}
Glucose (mmol/l)	18.8 (4.6)	18.8 (4.6)	18.8 (4.7)	7.2×10^{-1}
Blood pressure assessment				
SBP (mmHg)	185.5 (30.3)	182.0 (30.0)	192.8 (30.2)	4.5×10^{-22}
DBP (mmHg)	103.6 (16.7)	101.4 (16.5)	108.3 (16.7)	5.3×10^{-29}
Heart rate (beats/min)	94 (16)	92 (16)	98 (16)	3.0×10^{-21}
Currently treated hypertension ^c	60.0	53.0	74.6	4.9×10^{-31}
Renal function assessment				
eGFR _{CKD-EPI} (ml/min per 1.73 m ²)	69.6 (17.9)	70.9 (15.8)	66.8 (15.9)	1.2×10^{-11}
UACR (μg/mg)	78.1 (155)	63.5 (152)	96.7 (153)	5.1×10^{-9}
Microalbuminuria ^a	25.4	23.7	28.9	1.7×10^{-3}
Macroalbuminuria ^b	5.4	4.0	7.6	3.6×10^{-5}

Age and age at diagnosis of diabetes are adjusted for sex; diabetes duration, BMI, currently treated hypertension and micro and macroalbuminuria are adjusted for age and sex; all others traits are adjusted for age, sex, and respective treatments.

eGFR_{CKD-EPI}, estimated glomerular filtration rate calculated using Chronic Kidney Disease Epidemiology Collaboration equation; HbA1c, serum glycosylated hemoglobin; UACR, urinary albumin–creatinine ratio.

^aUrinary albumin–creatinine ratio between 30 and 300 μg/mg.

^bUrinary albumin–creatinine ratio >300 μg/mg.

^cBlood pressure >140/90 mmHg or receiving antihypertensive treatment

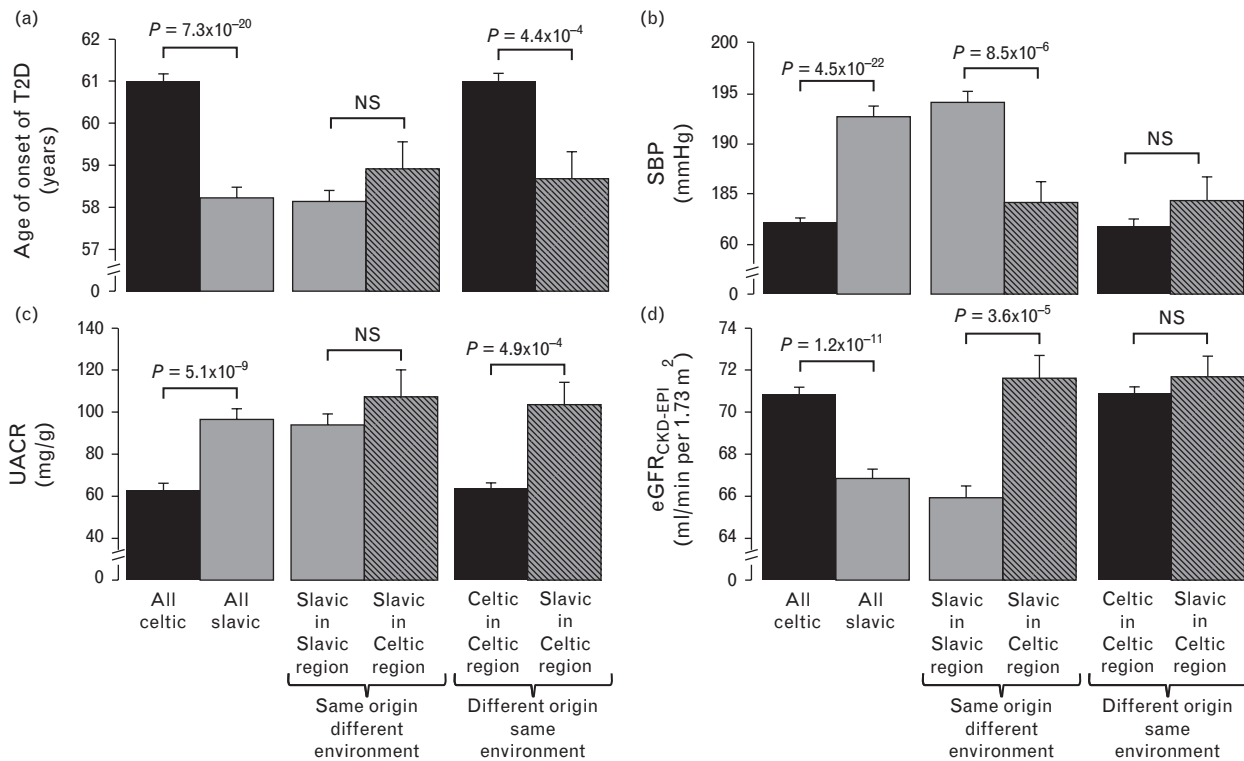


FIGURE 2 Effect of ethnic origin (Celtic vs. Slavic) and environment (Celtic region vs. Slavic region) on age of onset of T2D, SBP, UACR, and eGFR_{CKD-EPI}. Age of onset of T2D (a) and UACR (c) are more linked to genetics whereas SBP (b) and eGFR_{CKD-EPI} (d) are more impacted by environmental factors. Age of onset of T2D is adjusted for sex whereas all other traits are adjusted for sex, age, and treatment (nonparametric adjustment [34]). eGFR_{CKD-EPI}, estimated glomerular filtration rate calculated using Chronic Kidney Disease Epidemiology Collaboration equation; PC, principal component; UACR, urinary albumin–creatinine ratio; T2D, type 2 diabetes.

Imputation

Two sets of imputation were performed separately for the individuals genotyped on Affymetrix arrays 5.0 and 6.0 using SHAPEIT [30] and IMPUTE2 software [31] and the 1000 genome project [32] phased 3 data set as reference. Only those SNPs with a MAF greater than or equal to 5% and with an imputation quality score greater than or equal to 0.80 were kept as has been proposed in previous studies [33].

Statistical analysis

Analyses of the differences in phenotype values (mean values for quantitative traits and numbers of individuals affected for qualitative traits) between groups (between individuals with Celtic and Slavic genetic profiles; between individuals with Slavic profiles living in predominantly Germano–Celtic European or European descent countries (Celtic region) and individuals with Slavic profiles living in predominantly Slavic European countries (Slavic region); and between individuals with Slavic genetic profiles living in predominantly Germano–Celtic European countries and individuals with Celtic profiles living in those countries) were performed using general linear models included in the R statistical package [34].

Differences in age of onset of diabetes and duration of diabetes were tested using sex as a covariate. All other phenotype differences were tested using age and sex as covariates so that the significance of these differences are age and sex adjusted. Mean phenotype values were also adjusted for sex and age where appropriate using the

epicalc library [35] of the R statistical package [34]. When appropriate, adjustment for treatment for such traits as SBP, UACR, and eGFR were done using nonparametric adjustment as described [36].

Genome-wide association studies

GWAS were performed for age of onset of T2D separately for individuals with a Celtic or Slavic genetic profile as determined by their value for PC1 and for the combined Celtic and Slavic sample using linear regression with an additive genetic model and sex as well as the two respective first principal components of population stratification as covariates.

Association analyses were performed separately on the two imputed datasets for individuals that were genotyped on the different arrays (5 986 672 SNPs for 1015 individuals genotyped on chip 5.0 and 6 442 695 SNPs for 2394 individuals genotyped on chip 6.0) and results were merged using a fixed effects meta-analysis routine in the PLINK software [29] to avoid the possibility of any bias that might have arisen from uneven phenotype distributions across different genotyping chip technologies. The combined meta-analysis data set contained a total of 5 045 527 SNPs that passed all previous quality control steps in both data sets and also passed a combined test for Hardy–Weinberg equilibrium using a critical P value of 1×10^{-3} ($P < 10^{-3}$).

Effect size

The relative effect sizes (γ) for each SNP, weighted by the size of the β coefficient of regression (β), the standard error

TABLE 2. Summary statistics for association of age of onset of T2D for 25 independent single-nucleotide polymorphisms in Caucasians study participants from ADVANCE. Genome-wide association studies were done for all study participants (n = 3409 combined group) and separately for Celtic (n = 2307) and Slavic (n = 1102) individuals stratified by ethnic origin, which is determined by principal component analysis. Selected single-nucleotide polymorphisms for all groups are associated to age at diagnosis of diabetes with P value less than 1×10^{-5}

SNP ID	Chr	Position (bp, hg19)	RA	Locus ^a	Celtic			Slavic			Combined ^c		
					P value	RAF%	Effect size ^b	P value	RAF%	Effect size ^b	P value	RAF%	Effect size ^b
SNPs that are significantly associated in both Celtic and Slavic genetic profiles and that increase in significance in combined group ^c													
rs34428389	3	52 894 142	A	TMEM110 , MIR8064, SRMB1T1	3.3×10^{-4}	19.1	-2.00	8.9×10^{-3}	18.4	-1.43	8.1×10^{-6}	18.9	-2.47
rs17447640	4	42 555 811	G	ATP8A1 , SHISA3, GRXCR1	1.1×10^{-5}	13.9	-2.15	5.6×10^{-2}	15.7	-0.98	1.5×10^{-6}	14.5	-2.39
rs11298745	7	18 548 252	Del	HDAC9 , MIR1302-6, TWIST1	2.3×10^{-3}	90.5	-1.26	4.6×10^{-5}	90.0	-1.73	1.6×10^{-6}	90.3	-2.01
rs34620785	8	103 452 308	A	UBR5, ODF1	1.6×10^{-4}	88.4	-1.71	1.7×10^{-3}	84.4	-1.61	1.6×10^{-6}	87.1	-2.27
rs76703216	10	9 512 026	G	LOC101928272	3.8×10^{-5}	94.1	-1.37	3.5×10^{-2}	96.2	-0.57	3.6×10^{-6}	94.7	-1.46
rs35372009	14	38 782 341	Del	CLEC14A, LINC00639	1.1×10^{-2}	42.5	-1.78	9.0×10^{-6}	39.5	-3.07	3.3×10^{-6}	41.6	-3.24
rs148077446	16	77 310 608	Del	.SYCE1L, ADAMTS18	1.6×10^{-4}	28.7	-2.42	5.0×10^{-3}	31.9	-1.85	5.7×10^{-6}	29.7	-2.93
SNPs that are only significantly associated in Celtic profile													
rs12743974	1	67 708 357	A	IL23R , C10RF141, IL12RB2	5.3×10^{-6}	41.9	-3.18	3.8×10^{-1}	41.8	0.62	1.2×10^{-3}	41.8	-2.25
rs12736701	1	77 815 425	C	AK5 , PIGK, ZZZ3	7.6×10^{-6}	93.4	-1.57	8.4×10^{-1}	93.9	0.07	3.2×10^{-4}	93.6	-1.24
rs7412314	1	171 287 252	T	FMO4 , FMO1, TOP1P1	4.3×10^{-6}	73.3	-2.88	6.5×10^{-1}	69.8	-0.30	6.3×10^{-5}	72.1	-2.53
rs11099942	4	155 132 707	T	SFRP2, DCH52	4.3×10^{-6}	6.4	-1.59	6.7×10^{-1}	5.1	0.13	1.7×10^{-4}	5.9	-1.26
rs9502478	6	6 593 695	T	LY86-A51 , F13A1	4.3×10^{-6}	17.3	-2.46	6.3×10^{-1}	16.0	-0.25	3.2×10^{-5}	16.9	-2.20
rs2107167	7	109 590 911	A	EIF3P1	8.0×10^{-6}	13.7	-2.17	9.6×10^{-1}	9.6	-0.02	4.1×10^{-5}	12.4	-1.91
rs10973627	9	37 971 389	C	SHB , SLC25A51, ALDH1B1	4.1×10^{-6}	87.7	-2.14	6.4×10^{-1}	89.2	-0.21	9.9×10^{-6}	88.2	-2.02
rs113932007	11	115 491 436	Del	CADM1, LINC00900	5.9×10^{-7}	65.1	-3.37	5.1×10^{-1}	64.5	-0.45	8.8×10^{-6}	64.9	-3.00
rs10512488	17	40 963 904	G	BECN1 , CNTD1, MIR6781	3.1×10^{-6}	74.0	-2.90	2.9×10^{-1}	72.2	-0.67	9.5×10^{-6}	73.4	-2.77
SNPs that are only significantly associated in Slavic profile													
rs340841	1	214 124 470	C	PROX1-A51 , LINC00538, PROX1	7.9×10^{-1}	45.9	0.19	8.6×10^{-6}	52.8	-3.14	2.5×10^{-2}	48.1	-1.59
rs12714314	2	1 943 617	C	MYT1L , PXDN, MYT1L-AS1	9.3×10^{-1}	75.8	0.05	3.4×10^{-6}	72.8	-2.92	1.9×10^{-2}	74.9	-1.44
rs58383906	3	29 246 181	C	LINC00693, RBMS3-AS3	3.3×10^{-1}	51.5	-0.68	2.8×10^{-6}	50.3	-3.32	6.9×10^{-4}	51.1	-2.40
rs13166103	5	57 742 202	C	LOC101928569, PLK2	3.5×10^{-1}	81.3	0.51	4.1×10^{-6}	79.8	-2.61	9.4×10^{-2}	80.8	-0.93
rs12188216	5	166 425 736	G	CTB-7E3.1, TENM2	2.9×10^{-1}	10.7	-0.46	9.7×10^{-6}	9.7	-1.85	1.7×10^{-3}	10.3	-1.35
rs9384193	6	154 554 249	C	OPRM1 , CNKSR3	8.7×10^{-1}	36.3	0.11	8.8×10^{-6}	36.3	-3.02	2.7×10^{-2}	36.3	-1.50
rs11060464	12	130 097 850	G	TMEM132D , LOC101927735, LOC100190940	5.4×10^{-1}	70.9	0.39	3.7×10^{-6}	73.1	-2.90	6.0×10^{-2}	71.6	-1.20
rs1754680	14	38 848 003	A	CLEC14A, LINC00639	2.3×10^{-2}	59.4	-1.54	8.3×10^{-6}	57.4	-3.17	4.8×10^{-5}	58.7	-2.85
rs13043901	20	52 170 783	A	LOC101927770 , TSHZ2, ZNF217	3.3×10^{-1}	75.3	-0.60	5.0×10^{-6}	76.5	-2.74	4.2×10^{-4}	75.7	-2.14

RA, risk allele; RAF, risk allele frequency; SNP, single-nucleotide polymorphism.

^aIf gene is in bold, SNP lies within the gene.

^bThe effect size (ES) is determined by the β , MAF, and standard error using the equation described in literature [37].

^cAssociation is considered replicated in two independent samples when P values are nominally significant for each of Celtic and Slavic samples and are more significant for combined sample.

of β_j ($Se\beta_j$), and the MAF of the SNP (MAF_j) were estimated by the following equation [37]:

$$\gamma_j = \text{SQRT}[2MAF_j - (1-MAF_j)](\beta_j/Se\beta_j)$$

RESULTS

PCA of 3409 genotyped ADVANCE study participants using EIGENSOFT 3.0 package identified two major principal components. The first PC1 divided the individuals of Europe along an east–west region, whereas principal component 2 separated individuals of Europe into a north–south gradient (Fig. 1a). PC1 clearly separated countries with populations of predominantly Germano–Celtic ethnic background ('Celtic', $PC1 < 0$) from those with a Balto-Slavic ethnicity ('Slavic', $PC1 \geq 0$) with Germany aligned in the center of the distribution (Fig. 1b and c). When recruitment centers of the ADVANCE trial were ordered by values of PC1, we noted a pivot point between PC1 threshold values of 0.0 and 0.01 that separated Germany into Celtic (Munich) and Slavic (Dresden) origins (Fig. 1d).

Table 1 shows the main demographic and clinical characteristics of the two geo-ethnic groups at the entry of ADVANCE trial. The most striking difference between the two ethnic groups was the mean age of onset of diabetes. Individuals with Slavic profiles had T2D at a younger age ($P = 7.3 \times 10^{-20}$), had higher SBP and DBP ($P = 4.5 \times 10^{-22}$ and $P = 5.3 \times 10^{-29}$, respectively) despite the fact that a larger number of them were treated for hypertension and that they had a higher UACR at baseline ($P = 5.1 \times 10^{-9}$) even after adjusting for age, sex, and medication.

We then determined the effect of ethnic origin (Celtic vs. Slavic) and environment (Celtic region vs. Slavic region) on the most divergent phenotypes between Slavic and Celtic patients namely age of onset of T2D, BP, and renal function.

To assess the relative effects of genetic and environmental factors on the differences between individuals with Celtic and Slavic profiles, we compared study participants with a Slavic genetic profile living in countries with predominantly Germano–Celtic populations with individuals with a Slavic genetic profile living in predominantly Slavic countries of Europe. Significant differences between Slavs living in Slavic and Celtic countries would suggest an environmental/lifestyle effect, whereas no differences between Slavs living in Celtic or Slavic countries would support an impact of genetic influence. As shown in Fig. 2a, the highly significant earlier age of onset of T2D observed in individuals of Slavic origin was also present among the 175 Slavic study participants living in countries of predominantly Celtic populations, suggesting a genetic drive for this trait. Similarly, UACR was higher in Slavic individuals living in either Slavic or Celtic regions (Fig. 2c). This contrasted with eGFR that was higher in Celtic than Slavic individuals after adjustment of age, sex, and medication but was not different between Celtic and Slavic individuals living in the same environment (Fig. 2d). SBP is another good example of environmental effect as SBP was higher in Slavic than Celtic individuals but not different between Slavic and Celtic individuals living in Celtic countries (Fig. 2b).

As age of onset of T2D showed strong genetic differences between Slavic and Celtic geo-ethnic groups, we performed GWAS of this phenotype in Slavic, Celtic, and the two combined populations. All SNPs with association of nominal significance of P values below 10^{-5} from GWAS analysis are presented in Table 2. Associations that are nominally significant in each of the two independent Celtic and Slavic GWAS and that increase in significance in the combined Celtic and Slavic GWAS are considered to be replicated in two independent subcohorts, that is, Celtic and Slavic genetic profile populations. These SNPs are indicated in boldtype for combined sample in Table 2. Seven independent SNPs (not in linkage disequilibrium with each other) were found to be associated with age of onset of diabetes at $P < 10^{-5}$ having the most significant P value for the combined Celtic and Slavic cohorts, and thus considered replicated by the above criteria. Other SNPs were associated specifically to one or the other ethnic group. Nine independent SNPs were significant only for the Celtic group and a different set of nine independent SNPs were significant only for the Slavic group. Two SNPs within the same locus were associated with age of onset of diabetes in the two different groups. SNP, rs35372009, near the *CLEC14A* gene was the most significantly associated for the combined Celtic and Slavic cohort ($P = 3.3 \times 10^{-6}$; Fig. 3a) and SNP, rs1754680, was the most significantly associated for the Slavic only group ($P = 8.3 \times 10^{-6}$). The SNPs are 65 662 bp apart on chromosome 14q21.1 and are in high linkage disequilibrium. These two SNPs, which lie within a region that is 5' of the *CLEC14A* gene, are representing the same association (Fig. 3a and b).

The most interesting association is within the *PROX1/PROX1-AS1* gene locus (rs340841) that is characterized by one of the highest effect sizes for age of onset of T2D. The homozygous CC genotype for rs340841 is associated with 4.4 years earlier onset of T2D in Slavic patients living either in Slavic countries or in Celtic countries (Fig. 4). Furthermore, the C allele is the major allele in Slavic individuals (Table 2). This locus is also associated with eGFR decline in Slavs, with macroalbuminuria and hypertension in all ADVANCE study participants of Caucasian origin and with IL-6 levels at baseline (data not shown). A literature search indicated that the *PROX1* gene has been associated with abnormalities of glucose metabolism and risk of diabetes with ethnically specific individual polymorphisms [38–41].

DISCUSSION

The global burden of cardio-metabolic risk factors adjusted for age and sex has been shown to be greater in Eastern than in Western European countries [42]. We have recently reviewed the importance of lifestyle behavior in gene–environment interactions analysis [24]. The current study is adding the notion that analysis of migration of a population within a distinct population even before its admixture may help dissect environmental from genetic contributions.

There is some debate as to the original homeland of the Balto-Slavs. One hypothesis holds that modern Baltic and

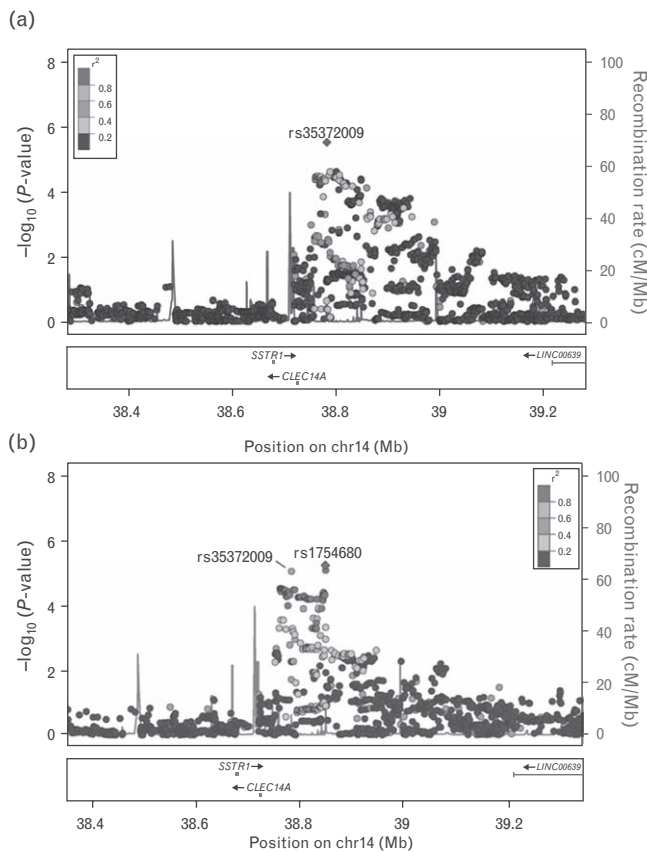


FIGURE 3 Regional association plots (1Mb window) of the *CLEC14A/LINC00639* locus (chromosome 14) identified by GWAS of age of onset of T2D in the combined (a) and Slavic only cohorts (b), respectively. $-\log_{10}(P)$ values are plotted against genomic position (build 37, hg19). The lead SNPs (rs35372009 and rs1754680) are indicated in purple diamonds. The SNPs surrounding the lead SNPs are color coded based upon their linkage disequilibrium with the lead SNPs (taken from pairwise r^2 values from the 1000 Genome EUR Database): red (r^2 with lead SNP 0.8–1.0), orange (0.6–0.8), green (0.4–0.6), light blue (0.2–0.4), and dark blue (<0.2). The recombination rates (cM/Mb) are plotted in blue to reflect local linkage disequilibrium structure. Genes, exons, and direction of transcription from UCSC genome browser (genome.ucsc.edu) are noted. Plots are generated using LocusZoom (<http://csg.sph.umich.edu/locuszoom>). GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes.

Slavic populations descend from a proto-Slavonic parental group most likely located in a homeland roughly corresponding to the modern western Ukraine and then expanded by the sixth and seventh centuries A.D. as the Prague–Penkov–Kolochin complex of cultures to an area defined by the Baltic Sea in the north, approximately the Volga river in the east, the area defined by the modern day Czech Republic and the Elbe River in modern Germany in the west and the Danube basin in the south [43].

The area of north eastern Germany between the Oder and Elbe rivers was occupied by the Polabian Slavic ethnic group by the sixth century. By the ninth century conflicts between Christian Germanic people and these western pagan Slavs began. The conflicts eventually resulted in the incorporation of the area into the Holy Roman Empire by the thirteenth century and the linguistic germanification of the populations [44]. Therefore, it is reasonable to hypothesize the surviving presence of genetic evidence for an ethnic divide between a Germano–Celtic group (here referred to as simply ‘Celtic’) and a Balto–Slavic group (here referred to simply as ‘Slavic’) centered roughly along

the Elbe river and the border of Czech republic in northern Europe. We have demonstrated that evidence of this ancestral Celtic–Slavic genetic divide still exists in the modern European population and that it is reflected in differences in genetic–phenotype correlations.

Noticeably, the Caucasian populations of the non-European countries involved in ADVANCE have founding populations that are principally of Germano–Celtic origin as a result of British Empire expansion. Slavic migration is more recent in these countries and therefore represents a minor ethnic component.

The principal difference between the Celtic and Slavic ethnic groups is age of onset of T2D, which is also correlated with other phenotypes such as albuminuria. We have previously introduced the concept of accelerated aging as being a primary cause of many complex genetic diseases [45]. It is a strong possibility that individuals with a Slavic genetic profile, despite their environment, are genetically more susceptible to accelerated aging resulting in earlier onset of T2D and associated albuminuria. In addition, our results from ADVANCE demonstrated that in contrast to macrovascular complications of diabetes that are strongly age dependent with an added risk conferred by duration of diabetes, the adverse effects of duration of diabetes on microvascular events were observed in the youngest age group [46], which is also compatible with observations of the Treatment Options for type 2 Diabetes in Adolescents and Youth trial [47]. Although the ADVANCE trial amply demonstrated, the decrease of renal events and total mortality by intensification of BP as well as of blood glucose control [48], a finding that is confirmed for glycemic control in the Veteran’s Affairs Diabetes Trial and Action to Control Cardiovascular Risk in Diabetes trials [49], the current study suggests that further specific functional benefits on eGFR and UACR should be analyzed with respect to geo-ethnicity.

PROX1 encodes the prospero homeobox 1 protein, a human homologue of the *Drosophila prospero* gene. This protein is a homeobox transcription factor involved in developmental processes such as cell fate determination, gene transcriptional regulation, and progenitor cell regulation in a number of organs. It plays a critical role in embryonic development. *PROX1* has been shown to be associated with diabetes and its complications in a number of studies [38–41,50–52]. Here, we present evidence that the genetic influence of *PROX1* on age of onset of diabetes is different within Caucasian ethnic groups. It is of interest that several polymorphisms at this locus are associated with insulin levels and its control in adolescence, selected for a lesser impact of environmental determinants at this age by Lecompte *et al.* [40]. As we have mentioned, the ADVANCE trial demonstrated that earlier onset of diabetes has more impact on micro than macrovascular complications [46]. We propose that earlier onset of T2D in context of genetic \times environmental influences and ethnicity deserves further attention as a potential new target for early detection and intervention in T2D.

In conclusion, genetic analyses have to consider geo-ethnic characteristics even within Caucasians, demonstrated here for cardinal features of T2D. Our data suggest that understanding of distinct genomic architectures is important to ascertain clinical utility.

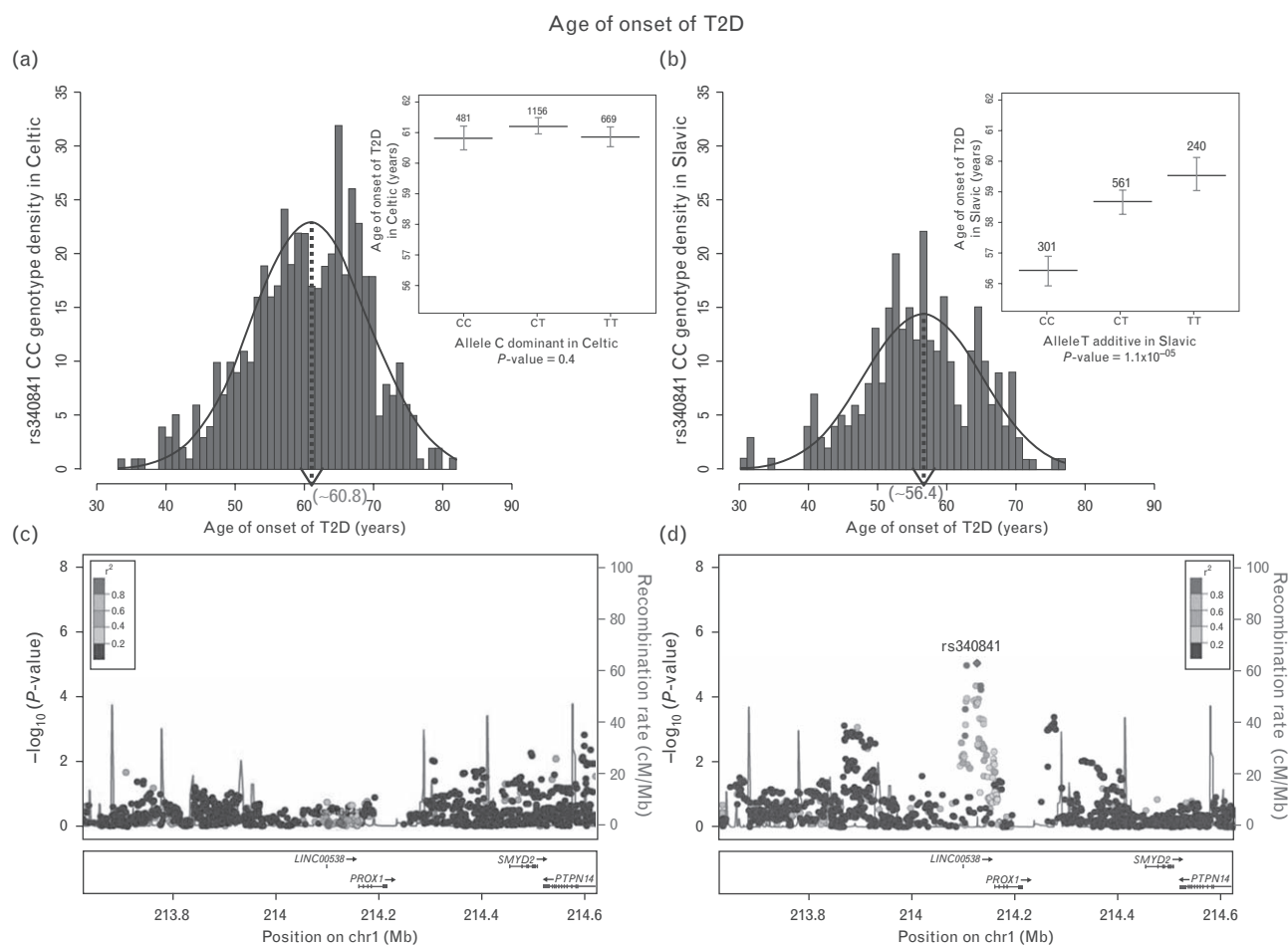


FIGURE 4 Distribution of the CC genotype frequencies for rs340841 located at the *PROX1* locus in Celtic (a) and Slavic (b) study participants separately. Histograms of means of age of onset of T2D vs. genotype of rs340841 (inserts). Regional association plots (1Mb window) of the *PROX1* locus (chromosome 1) identified by GWAS of age of onset of T2D in Celtic (c) and Slavic (d) study participants, respectively. $-\log_{10}(P\text{-value})$ are plotted against genomic position (build 37, hg19). The lead SNP (rs340841) is indicated in purple diamond. The SNPs surrounding rs340841 are color coded based on their linkage disequilibrium with the lead SNP (taken from pairwise r^2 values from the 1000 Genome EUR Database): red (r^2 with lead SNP 0.8–1.0), orange (0.6–0.8), green (0.4–0.6), light blue (0.2–0.4), and dark blue (<0.2). The recombination rates (cM/Mb) are plotted in blue to reflect local linkage disequilibrium structure. Genes, exons, and direction of transcription from UCSC genome browser (genome.ucsc.edu) are noted. Plots are generated using LocusZoom (<http://csg.sph.umich.edu/locuszoom>). GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes.

ACKNOWLEDGEMENTS

The authors acknowledge the support of grant OPTI-THERA, a private–public initiative of Government of Quebec in ‘Optimization of Therapeutic Approaches in Personalized Primary Care.’

The project is financed by the ‘Fond de Partenariat pour un Québec Innovant et en Santé’ from the Ministry of Economic Development, Innovation and Export Trade with additional funding from Canadian Institutes of Health Research, Genome Quebec, Servier Canada and Les Laboratoires Servier, Paris, France.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- International Diabetes Federation Diabetes Atlas. Seventh Edition. 2015
- Groop PH, Thomas MC, Moran JL, Waden J, Thorn LM, Makinen VP, *et al.*, FinnDiane Study Group. The presence and severity of chronic kidney disease predicts all-cause mortality in type 1 diabetes. *Diabetes* 2009; 58:1651–1658.
- Afkarian M, Sachs MC, Kestenbaum B, Hirsch IB, Tuttle KR, Himmel-farb J, de Boer IH. Kidney disease and increased mortality risk in type 2 diabetes. *J Am Soc Nephrol* 2013; 24:302–308.
- Chan GC, Tang SC. Diabetic nephropathy: landmark clinical trials and tribulations. *Nephrol Dial Transplant* 2016; 31:359–368.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; 361:598–604.
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, *et al.*, International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449:851–861.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38:904–909.
- Wang H, Haiman CA, Kolonel LN, Henderson BE, Wilkens LR, Le Marchand L, Stram DO. Self-reported ethnicity, genetic structure and the impact of population stratification in a multiethnic study. *Hum Genet* 2010; 128:165–177.
- Beckie TM, Beckstead JW, Groer MW. The association between variants on chromosome 9p21 and inflammatory biomarkers in ethnically diverse women with coronary heart disease: a pilot study. *Biol Res Nurs* 2011; 13:306–319.
- Dvornik V, Liu XH, Shen H, Lei SF, Zhao LJ, Huang QR, *et al.* Differentiation of Caucasians and Chinese at bone mass candidate genes: implication for ethnic difference of bone mass. *Ann Hum Genet* 2003; 67:216–227.

11. Ortega VE, Meyers DA. Pharmacogenetics: implications of race and ethnicity on defining genetic profiles for personalized medicine. *J Allergy Clin Immunol* 2014; 133:16–26.
12. Abdullah N, Attia J, Oldmeadow C, Scott RJ, Holliday EG. The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. *Int J Endocrinol* 2014; 2014:593982.
13. Hara K, Shojima N, Hosoe J, Kadowaki T. Genetic architecture of type 2 diabetes. *Biochem Biophys Res Commun* 2014; 452:213–220.
14. Keaton JM, Cooke Bailey JN, Palmer ND, Freedman BI, Langefeld CD, Ng MC, Bowden DW. A comparison of type 2 diabetes risk allele load between African Americans and European Americans. *Hum Genet* 2014; 133:1487–1495.
15. Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014; 46:234–244.
16. Ng MC. Genetics of type 2 diabetes in African Americans. *Curr Diab Rep* 2015; 15:74.
17. Liu CT, Garnaas MK, Tin A, Kottgen A, Franceschini N, Peralta CA, *et al.* Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. *PLoS Genet* 2011; 7:e1002264.
18. Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, *et al.* Kidney Gen Consortium, CKDGen Consortium, GUGC consortium. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* 2012; 44:904–909.
19. Yamada Y, Nishida T, Ichihara S, Kato K, Fujimaki T, Oguri M, *et al.* Identification of chromosome 3q28 and ALPK1 as susceptibility loci for chronic kidney disease in Japanese individuals by a genome-wide association study. *J Med Genet* 2013; 50:410–418.
20. McEvoy BP, Montgomery GW, McRae AF, Ripatti S, Perola M, Spector TD, *et al.* Geographical structure and differential natural selection among North European populations. *Genome Res* 2009; 19:804–814.
21. Zhu X, Luke A, Cooper RS, Quettermous T, Hanis C, Mosley T, *et al.* Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 2005; 37:177–181.
22. Arbeev KG, Ukraintseva SV, Akushevich I, Kulminski AM, Arbeeva LS, Akushevich L, *et al.* Age trajectories of physiological indices in relation to healthy life course. *Mech Ageing Dev* 2011; 132:93–102.
23. Keil U. The worldwide WHO MONICA project: results and perspectives. *Gesundheitswesen* 2005; 67 (Suppl 1):S38–S45.
24. Simon PH, Sylvestre MP, Tremblay J, Hamet P. Key considerations and methods in the study of gene-environment interactions. *Am J Hypertens* 2016; 29:891–899.
25. Sytkowski PA, Kannel WB, D'Agostino RB. Changes in risk factors and the decline in mortality from cardiovascular disease. The Framingham Heart Study. *N Engl J Med* 1990; 322:1635–1641.
26. Rationale and design of the ADVANCE study: a randomised trial of blood pressure lowering and intensive glucose control in high-risk individuals with type 2 diabetes mellitus. Action in Diabetes and Vascular Disease: PreterAx and DiamicroN Modified-Release Controlled Evaluation. *J Hypertens Suppl* 2001; 19:S21–S28.
27. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, *et al.*, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; 150:604–612.
28. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 2014; 197:573–589.
29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81:559–575.
30. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012; 9:179–181.
31. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010; 11:499–511.
32. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491:56–65.
33. Southam L, Panoutsopoulou K, Rayner NW, Chapman K, Durrant C, Ferreira T, *et al.* The effect of genome-wide association scan quality control on imputation outcome for common variants. *Eur J Hum Genet* 2011; 19:610–614.
34. R Development Core Team: A language and Environment for statistical computing. 2010
35. Chongsuvivatwong V. *Analysis of epidemiological data using R and EpiCalc*. Thailand: Epidemiology Unit, Prince of Songkla University; 2008.
36. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005; 24:2911–2935.
37. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; 340:1467–1471.
38. Kim HJ, Yoo YJ, Ju YS, Lee S, Cho SI, Sung J, *et al.* Combined linkage and association analyses identify a novel locus for obesity near PROX1 in Asians. *Obesity (Silver Spring)* 2013; 21:2405–2412.
39. Kretowski A, Adamska E, Maliszewska K, Wawrusiewicz-Kurylonek N, Citko A, Goscik J, *et al.* The rs340874 PROX1 type 2 diabetes mellitus risk variant is associated with visceral fat accumulation and alterations in postprandial glucose and lipid metabolism. *Genes Nutr* 2015; 10:454.
40. Lecompte S, Pasquetti G, Hermant X, Grenier-Boley B, Gonzalez-Gross M, De Henauf S, *et al.* Genetic and molecular insights into the role of PROX1 in glucose metabolism. *Diabetes* 2013; 62:1738–1745.
41. Ohshige T, Iwata M, Omori S, Tanaka Y, Hirose H, Kaku K, *et al.* Association of new loci identified in European genome-wide association studies with susceptibility to type 2 diabetes in the Japanese. *PLoS One* 2011; 6:e26911.
42. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G, Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration (BMI Mediated Effects). Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. *Lancet* 2014; 383:970–983.
43. Mallory JP, Adams DQ. *Encyclopedia of Indo-European culture*. London and Chicago: Fitzroy Dearborn Publishers; 1997.
44. Labuda G. Formation of the ethnic and cultural unity of Polabian Slavic tribes and its transformations in the historical development. In: Strzelczyk J, editor. *Polabian Slavdom between Germany and Poland*. Seria: Wyd. Nauk. UAM w Poznaniu; 1981.
45. El-Bikai R, Tahir MR, Tremblay J, Joffres M, Seda O, Sedova L, *et al.* Association of age-dependent height and bone mineral density decline with increased arterial stiffness and rate of fractures in hypertensive individuals. *J Hypertens* 2015; 33:727–735.
46. Zoungas S, Woodward M, Li Q, Cooper ME, Hamet P, Harrap S, *et al.*, ADVANCE Collaborative group. Impact of age, age at diagnosis and duration of diabetes on the risk of macrovascular and microvascular complications and death in type 2 diabetes. *Diabetologia* 2014; 57:2465–2474.
47. Lynch J, El L, Fisher L, Gidding SS, Laffel L, Libman I, *et al.*, TODAY Study Group. Rapid rise in hypertension and nephropathy in youth with type 2 diabetes: the TODAY clinical trial. *Diabetes Care* 2013; 36:1735–1741.
48. Zoungas S, de Galan BE, Ninomiya T, Grobbee D, Hamet P, Heller S, *et al.*, ADVANCE Collaborative Group. Combined effects of routine blood pressure lowering and intensive glucose control on macrovascular and microvascular outcomes in patients with type 2 diabetes: new results from the ADVANCE trial. *Diabetes Care* 2009; 32:2068–2074.
49. Hamet P. What matters in ADVANCE and ADVANCE-ON. *Diabetes Obes Metab* 2012; 14 (Suppl 1):20–29.
50. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010; 42:105–116.
51. Florez JC, Jablonski KA, McAteer JB, Franks PW, Mason CC, Mather K, *et al.*, Diabetes Prevention Program Research Group. Effects of genetic variants previously associated with fasting glucose and insulin in the Diabetes Prevention Program. *PLoS One* 2012; 7:e44424.
52. Gong Y, McDonough CW, Beitelshes AL, Karnes JH, O'Connell JR, Turner ST, *et al.* PROX1 gene variant is associated with fasting glucose change after antihypertensive treatment. *Pharmacotherapy* 2014; 34:123–130.