

# Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes

J. C. Florez · A. L. Price · D. Campbell · L. Riba · M. V. Parra · F. Yu · C. Duque · R. Saxena · N. Gallego · M. Tello-Ruiz · L. Franco · M. Rodríguez-Torres · A. Villegas · G. Bedoya · C. A. Aguilar-Salinas · M. T. Tusié-Luna · A. Ruiz-Linares · D. Reich

Received: 23 March 2009 / Accepted: 7 May 2009 / Published online: 13 June 2009  
© Springer-Verlag 2009

## Abstract

**Aims/hypothesis** Type 2 diabetes is more prevalent in US American minority populations of African or Native American descent than it is in European Americans. However, the proportion of this epidemiological difference that can be

ascribed to genetic or environmental factors is unknown. To determine whether genetic ancestry is correlated with diabetes risk in Latinos, we estimated the proportion of European ancestry in case–control samples from Mexico and Colombia in whom socioeconomic status had been carefully ascertained.

J. C. Florez and A. L. Price contributed equally to this work.

M. T. Tusié-Luna, A. Ruiz-Linares and D. Reich co-directed the study.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-009-1412-x) contains supplementary material, which is available to authorised users.

J. C. Florez · F. Yu  
Department of Medicine, Harvard Medical School,  
Boston, MA, USA

J. C. Florez · F. Yu · R. Saxena · D. Reich  
Program in Medical and Population Genetics,  
Broad Institute of Harvard and MIT,  
Cambridge, MA, USA

J. C. Florez (✉)  
Diabetes Research Center (Diabetes Unit),  
Massachusetts General Hospital,  
Simches Research Building, CPZN 5.250, 185 Cambridge Street,  
Boston, MA 02114, USA  
e-mail: jcflorez@partners.org

J. C. Florez · R. Saxena  
Center for Human Genetic Research,  
Massachusetts General Hospital,  
Boston, MA, USA

A. L. Price  
Department of Epidemiology, Harvard School of Public Health,  
Boston, MA, USA

A. L. Price  
Department of Biostatistics, Harvard School of Public Health,  
Boston, MA, USA

D. Campbell · A. Ruiz-Linares (✉)  
The Galton Laboratory,  
Department of Genetics, Evolution and Environment,  
University College London,  
Wolfson House, 4 Stephenson Way,  
London NW1 2HE, UK  
e-mail: a.ruizlin@ucl.ac.uk

L. Riba · M. Rodríguez-Torres · M. T. Tusié-Luna (✉)  
Unidad de Biología Molecular y Medicina Genómica,  
Instituto de Investigaciones Biomédicas,  
Universidad Nacional Autónoma de México/Instituto  
Nacional de Ciencias Médicas y Nutrición Salvador Zubirán,  
Vasco de Quiroga #15, Col. Sección 16,  
Tlalpan 14000 DF, Mexico  
e-mail: tusie@biomedicas.unam.mx

M. V. Parra · C. Duque · N. Gallego · L. Franco · A. Villegas ·  
G. Bedoya · A. Ruiz-Linares  
Laboratorio de Genética Molecular,  
Universidad de Antioquía,  
Medellín, Colombia

F. Yu  
Human Genome Sequencing Center,  
Department of Molecular and Human Genetics,  
Baylor College of Medicine,  
Houston, TX, USA

**Methods** We genotyped 67 ancestry-informative markers in 499 participants with type 2 diabetes and 197 controls from Medellín (Colombia), as well as in 163 participants with type 2 diabetes and 72 controls from central Mexico. Each participant was assigned a socioeconomic status scale via various measures.

**Results** Although European ancestry was associated with lower diabetes risk in Mexicans (OR [95% CI] 0.06 [0.02–0.21],  $p=2.0\times 10^{-5}$ ) and Colombians (OR 0.26 [0.08–0.78],  $p=0.02$ ), adjustment for socioeconomic status eliminated the association in the Colombian sample (OR 0.64 [0.19–2.12],  $p=0.46$ ) and significantly attenuated it in the Mexican sample (OR 0.17 [0.04–0.71],  $p=0.02$ ). Adjustment for BMI did not change the results.

**Conclusions/interpretation** The proportion of non-European ancestry is associated with both type 2 diabetes and lower socioeconomic status in admixed Latino populations from North and South America. We conclude that ancestry-directed search for genetic markers associated with type 2 diabetes in Latinos may benefit from information involving social factors, as these factors have a quantitatively important effect on type 2 diabetes risk relative to ancestry effects.

**Keywords** Genetic admixture · Genetic association · Latinos · Socioeconomic status · Type 2 diabetes

### Abbreviations

AIM Ancestry-informative marker

SNP Single nucleotide polymorphism

### Introduction

In comparison with the European American population, minority groups in the USA experience a disproportionate burden of type 2 diabetes [1]. This is particularly evident in some Native American tribes, with the Pima Indians presenting one of the highest population prevalences of type 2 diabetes ever reported for any ethnic group [2–4]. A similar trend can be observed among the admixed populations of Mexico, the Caribbean, and Central and South

America (including recent immigrants to the United States), who are commonly characterised as Latinos. Indeed, the risk of type 2 diabetes is two- to fivefold higher in Latinos from Puerto Rico, Texas, New Mexico or Colorado [5–9] than in whites, an epidemiological difference that persists after adjusting for other traits such as abdominal obesity [10].

These differences support the notion that diabetes risk factors occur at higher frequency in populations of Native American descent. A genetic component for this bias was suggested in the form of the ‘thrifty gene’ hypothesis and its interaction with the environment first posed by J. Neel over 40 years ago [11]. A similar hypothesis was proposed for Mexican-Americans, based on crude skin pigmentation measures [12] and subsequently supported by analysis of genetic markers [13]. Initial evidence of a genetic contribution to the higher risk of type 2 diabetes in Native Americans was provided by the observation that Pima Indians with type 2 diabetes have significantly more Native American ancestry than their normoglycaemic counterparts [14] and that the fraction of European ancestry was associated with protective metabolic phenotypes in this population [15]; the potential contribution of socioeconomic status to these estimates was not fully assessed. In addition, the recent identification of the *ABCA1* R230C variant in Mexican admixed individuals (mestizos) illustrates that exclusive gene variants derived from Native American ancestry can indeed influence type 2 diabetes risk [16, 17].

The advent of comprehensive databases cataloguing genetic variation [18], the development of high-throughput genotyping technologies and the availability of DNA samples from multiple populations make it possible to select a set of single nucleotide polymorphisms (SNPs) that are highly informative for geographic ancestry, commonly termed ancestry-informative markers (AIMs). Thus, SNPs that have widely divergent allele frequencies in ancestral populations can be compiled to make such determinations of ancestral origin. When evenly spaced throughout the genome, approximately 2,000 AIMs can be employed to infer ancestry at each genomic location for admixture mapping; but a much smaller set of randomly distributed AIMs (<100) can also be genotyped in admixed persons to derive a fairly accurate estimate of genetic ancestry, expressed as a proportion of each individual’s genome. Such compilations specific to Latino populations have recently been done by three different groups, including our own [19–21].

In anticipation of genome-wide admixture mapping, AIMs have been applied to Latino samples with the goal of estimating the genetic contribution to the increased diabetes prevalence in this population. An initial study performed in American Latinos and stratified by neighbourhood detected a strong association between Native American ancestry and socioeconomic status; however, the

R. Saxena · D. Reich  
Department of Genetics, Harvard Medical School,  
Boston, MA, USA

M. Tello-Ruiz  
Cold Spring Harbor Laboratory,  
Cold Spring Harbor, NY, USA

C. A. Aguilar-Salinas  
Departamento de Endocrinología y Metabolismo, Instituto  
Nacional de Ciencias Médicas y Nutrición Salvador Zubirán,  
Mexico City, Mexico

authors concluded that despite the presence of such confounding, a genetic component to the increased disease prevalence was likely [13]. The association between type 2 diabetes and Native American ancestry was further substantiated by Parra et al. in Mexican-Americans from the San Luis Valley, Colorado; but once again, controlling for income and education abolished the statistical significance of the finding [22]. More recently, a similar study was conducted in a sample of 286 unrelated diabetic patients and 275 controls assembled from users of the Social Security hospital in Mexico City, which is thought to capture a large middle segment of the population devoid of upper- and lower-income outliers. This report found a non-significant increase in Native American ancestry among participants with diabetes, but a much stronger association between higher educational level and both European ancestry and non-diabetic status [23].

The success of admixture mapping in Latino populations is predicated on the ability to dissect these extra-genetic confounders from the genetic association. The correlation of socioeconomic status with ancestry in samples from two distinct US American locations and in an additional sample from Mexico City suggests that this may be a general phenomenon. To replicate and expand these observations, and as a way of investigating its likely impact on ongoing admixture mapping studies, we evaluated the contribution of socioeconomic status to the ancestry–diabetes relationship in two separate, non-US Latino populations from North and South America.

## Methods

### Populations

We studied 499 patients with type 2 diabetes and 197 controls from Medellín (Colombia), as well as 163 patients with type 2 diabetes and 72 controls from central Mexico. Demographic characteristics are presented in Table 1. In Colombia, diabetic patients were recruited from diabetes clinics in and around Medellín, with diagnostic criteria including fasting plasma

glucose  $\geq 7$  mmol/l or 2 h glucose  $> 11.1$  mmol/l after a 75 g OGTT. Controls were unaffected participants over 40 years of age who had no history of diabetes among first-degree relatives (and at times also including grandparents). Medellín is divided into main sectors ('comunas') and the mean socioeconomic status is available for each one. We balanced the collection of patients and controls in Medellín by selecting the diabetes clinics (for the patients) and the centres for the care of the elderly (for the controls) from the same sectors of the city, with the express intention of reflecting a range of socioeconomic status strata. In Mexico, type 2 diabetic participants included individuals treated at the Diabetes Clinic of the Instituto Nacional de Ciencias Médicas y Nutrición in Mexico City, in whom diabetes was confirmed with a blood glucose sample obtained after a 9 to 12 h fast. Exclusion criteria included secondary causes of diabetes (e.g. diseases of the exocrine pancreas, endocrinopathies, both drug or chemical-induced), genetic syndromes associated with diabetes and insulin therapy during the first 2 years after diagnosis. Controls were selected among spouses of diabetic participants or unrelated patients who were seeking medical attention at the same Instituto Nacional de Ciencias Médicas y Nutrición for reasons other than diabetes (e.g. primary dyslipidaemia) and were older than 40 years of age, had no history of diabetes among first-degree relatives and had fasting plasma glucose less than 6.1 mmol/l. Individuals identified themselves as Mexican mestizos with both parents and grandparents born in Mexico. By this ascertainment strategy we tried to ensure that our cases and controls were selected from the same geographical area, and had similar socioeconomic status and comparable access to the public health system.

To estimate allele frequencies in the ancestral populations and project ancestry proportions, we also studied several unmixed populations: European Americans from Baltimore and Chicago ( $n=77$ ) and from the HapMap Centre d'Etude du Polymorphisme Humain (Utah residents with northern and western European ancestry) collection ( $n=60$ ); Spaniards from Valencia ( $n=31$ ); West Africans from Ghana ( $n=52$ ) and from the HapMap Yoruba in Ibadan, Nigeria collection ( $n=60$ ); and Native Americans from the Mazahua ( $n=22$ ), Zapotec ( $n=60$ ), Mixtec ( $n=23$ ) and Mixe ( $n=29$ ) populations. All participants gave informed consent and studies were carried out in accordance with the principles of the Declaration of Helsinki as revised in 2000.

### Estimation of socioeconomic status

In Colombia, we had access to a government-assigned 'property band' based on property valuation of each individual home for the purposes of billing for public utilities and ranging from 1 (lowest) to 6 (highest). An assignment to one of these strata was made for each participant based on the

**Table 1** Demographic characteristics of genotyped samples

	Mexico		Colombia	
	Cases	Controls	Cases	Controls
<i>n</i>	163	72	499	197
Men/women	72/91	32/40	183/316	53/144
Age (years)	48.5 $\pm$ 12.5	56.0 $\pm$ 9.9	64.1 $\pm$ 10.3	60.7 $\pm$ 9.9
BMI (kg/m <sup>2</sup> )	28.1 $\pm$ 5.8	25.6 $\pm$ 3.7	26.7 $\pm$ 4.6	25.4 $\pm$ 4.1

Quantitative traits are means  $\pm$  SD

telephone number they provided at the time of their interview. We contrasted this information with other data we collected (such as home or car ownership) and confirmed a close correlation between the governmental water usage rank and these other markers of socioeconomic status.

In Mexico, socioeconomic status was determined by social workers at the National Institute of Medical Sciences and Nutrition using a standardised and validated tool currently applied in all Mexican National Institutes of Health studies [24]. Questionnaires include information on six categories: family monthly income, occupation of the head of the household, percentage of family income spent on food, type and characteristics of residence (owner-occupied, rented or shared with extended family), place of residence and the presence of chronic illnesses in other family members. Points are given for each category and the sum is used to assign participants to one of six socioeconomic status bands (lowest to highest). Supporting documents are required to validate the information. When information was not complete or questionable, socioeconomic status assignments were further explored and confirmed via unscheduled home visits.

### Genotyping

Samples from all populations were genotyped using the Sequenom MassARRAY technology [25] at 67 AIMs (Electronic supplementary material [ESM] Table S1). These SNPs are on average 49% different in frequency between Native Americans and European Americans, and are spaced by at least 10 cM (or >10 Mb) on chromosomes 1 to 22 (see Supplementary Table B in Smith et al. [26], from which these markers were selected). The large number of markers and the high degree of informativeness per marker in this set yield precise estimates of the proportion of European ancestry for each individual and also a measure of the precision of each estimate as a standard error. The average standard error for the percentage European ancestry estimate was  $\pm 7.2\%$  (in Mexicans) and  $\pm 8.0\%$  (in Colombians) respectively. We had complete data for 89% of genotypes; this reduction below 100% reflects the fact that slightly different panels of markers were genotyped on some samples. The missing data are not expected to affect our estimates of ancestry proportion.

### Statistical methods

To search for genetic differences associated with ancestry between participants with type 2 diabetes and controls (specifically, under-representation of European ancestry in diabetic participants), we used a mixture-of-binomials model to estimate the ancestry proportions of each Latino individual as previously described [20], and compared the

distributions of ancestries in cases vs controls (Table 2). We note that Latino populations represent a three-way admixture of European, Native American and African ancestry. Thus, we had a choice of whether to base our analyses on per cent European ancestry (distinguishing European vs [Native American + African]) or on per cent Native American ancestry (distinguishing Native American vs [European + African]). Because the 67 AIMs used to infer ancestry were relatively less informative for Native American vs African ancestry, the standard errors for inferred per cent European ancestry were smaller than the standard errors for inferred per cent Native American ancestry: therefore, we based our analyses on per cent European ancestry.

Associations of European ancestry proportion, socioeconomic status and/or BMI with type 2 diabetes status were assessed via logistic regression with (1) European ancestry proportion, (2) socioeconomic status, (3) socioeconomic status and European ancestry proportion or (4) socioeconomic status, European ancestry proportion and BMI as covariates. We included a constant term in each regression analysis. For example, letting  $\pi$  denote disease outcome, the logistic regression model for (4) was

$$\pi = \frac{e^{c_0 + c_{\text{SES}}^{\text{SES}} + c_{\text{anc}}^{\text{anc}} + c_{\text{BMI}}^{\text{BMI}}}}{1 + e^{c_0 + c_{\text{SES}}^{\text{SES}} + c_{\text{anc}}^{\text{anc}} + c_{\text{BMI}}^{\text{BMI}}}}$$

where  $c_{\text{BMI}}$  is the logistic regression coefficient multiplying BMI,  $c_{\text{anc}}$  is genome-wide ancestry,  $c_{\text{anc}}$  is the logistic regression coefficient multiplying anc, SES is socioeconomic status,  $c_{\text{SES}}$  is the logistic regression coefficient multiplying socioeconomic status, and  $c_0$  is the logistic regression coefficient multiplying the constant term.

For each regression analysis, we computed effect sizes,  $p$  values and pseudo- $r^2$  values. We defined pseudo- $r^2$  as the reduction in magnitude of the mean square value of disease outcome minus the predicted probability of disease outcome, comparing each of the four logistic regression models to a constant-term only model.

### Results

We analysed the relationship between genetic ancestry and type 2 diabetes status in two Latino populations from Mexico and Colombia. Both of these populations inherit European, Native American and African ancestry, but we focused our analysis on the proportion of European ancestry as inferred from using 67 AIMs. In the Mexicans, we observed a statistically significant difference (OR [95% CI] 0.06 [0.02–0.21],  $p = 2 \times 10^{-5}$ ) in genetic ancestry between patients with type 2 diabetes and age- and sex-matched controls from central Mexico (Table 2). A similar phenomenon, although of lesser magnitude, was observed



**Table 2** Association of non-European ancestry with type 2 diabetes in Latinos

Country of origin	<i>N</i> (cases)	<i>N</i> (controls)	European ancestry, cases (%)	European ancestry, controls (%)	Logreg coefficient $c_{anc}$	Pseudo- $r^2$ for $c_{anc}$	<i>p</i> value for $c_{anc}$
Mexico	163	72	33±20	46±24	-2.84	0.08	$2 \times 10^{-5}$
Colombia	499	197	56±15	59±14	-1.36	0.01	0.02

Unless otherwise indicated, values are means ± SD

We report average ancestry in cases and controls, as well as results of a logistic regression using ancestry ( $c_{anc}$ ) as a predictor of disease outcome

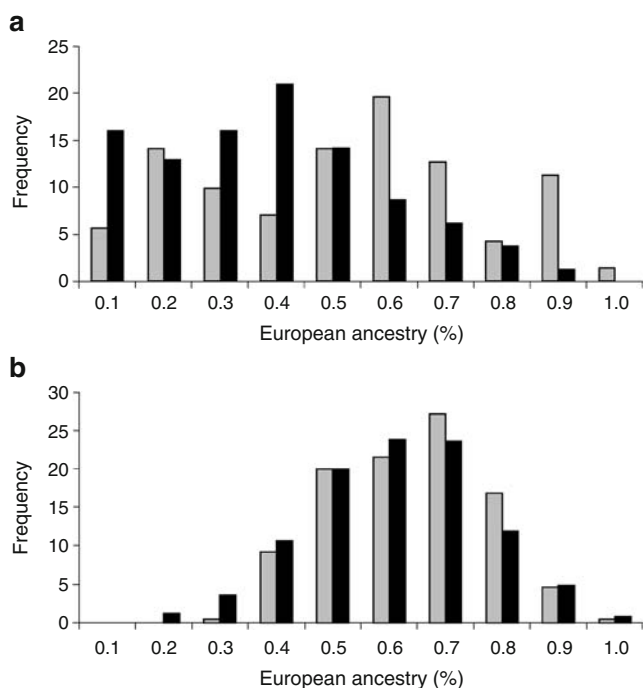
for the Colombians, with a nominally significant difference (OR 0.26 [0.08–0.78],  $p=0.02$ ) in genetic ancestry between diabetic participants and non-diabetic controls (Table 2). This ancestry difference was due to participants with type 2 diabetes having less European ancestry in both groups (Fig. 1). Overall, the proportion of European ancestry is estimated to be 33% in diabetic participants vs 46% in controls in Mexicans, and 56% vs 59% respectively in Colombians.

We investigated whether the observed association between genetic ancestry and type 2 diabetes was confounded by socioeconomic status. We determined that socioeconomic status was strongly correlated with European ancestry proportion in Mexicans (33% correlation,  $p=4 \times 10^{-7}$ ) and Colombians (34% correlation,  $p=2 \times 10^{-19}$ ), and therefore could potentially confound associations between genetic ancestry and type 2 diabetes. In the Mexicans, socioeconomic status was strongly predictive of case–control status

( $p=2 \times 10^{-10}$ ) and when socioeconomic status and ancestry were analysed together, the association between non-European ancestry and type 2 diabetes was significantly attenuated (OR 0.17 [0.04–0.71],  $p=0.02$ ; Table 3). In the Colombians, the association between socioeconomic status and type 2 diabetes was weaker, but still strongly significant due to the larger sample size ( $p=8 \times 10^{-7}$ ). Inclusion of both socioeconomic status and ancestry in the model abolished the significance of the ancestry–phenotype relationship (OR 0.64 [0.19–2.12],  $p=0.46$ ; Table 3). For both populations, inclusion of socioeconomic status in the model reduced the effect size as well as the statistical significance of the association between ancestry and type 2 diabetes (Table 3). However, in each case, the effect size and statistical significance of the association between socioeconomic status and type 2 diabetes was little changed by accounting for ancestry (Table 3).

As an alternative way to investigate the association between non-European ancestry and type 2 diabetes while accounting for socioeconomic status, we conducted a stratified analysis, in which we analysed each socioeconomic status stratum separately. We excluded the eight Colombian samples in socioeconomic status stratum 6 from this analysis, so that five strata were analysed for each population. In Mexicans and Colombians, the association between ancestry and type 2 diabetes did not reach statistical significance (nominal  $p=0.01$ , corrected  $p=0.05$  after accounting for five statistical tests) for any stratum (Fig. 2a,b). These results are consistent with the greatly reduced association between ancestry and type 2 diabetes when accounting for socioeconomic status (Table 3).

We also conducted a stratified analysis in which we considered each of five ancestry strata (0–20%, 20–40%, 40–60%, 60–80% or 80–100% European ancestry) and analysed associations between socioeconomic status and type 2 diabetes within each stratum (Fig. 2c,d). For Mexicans, we obtained nominal  $p$  values of 0.003, 0.12, 0.00001, 0.11 and 0.01, each with a negative coefficient for socioeconomic status. For Colombians, we excluded ancestry stratum 0–20% (which contained only six samples, each with type 2 diabetes) and obtained nominal  $p$  values of 0.004, 0.0006, 0.006 and 0.49, with a negative coefficient for socioeconomic status for each of the three significant



**Fig. 1** Histograms of European ancestry proportions in type 2 diabetes participants and controls from Mexico (a) and Colombia (b), using 67 AIMs. Black bars, type 2 diabetes; grey bars, controls

**Table 3** Association of non-European ancestry with type 2 diabetes in Latinos is confounded by socioeconomic status

Country of origin	$c_{SES}$			Logreg coefficients		Pseudo- $r^2$	$p$ values	
	Logreg coefficient	Pseudo- $r^2$	$p$ value	$c_{SES}$	$c_{anc}$		$c_{SES} + c_{anc}$	$c_{SES}^a$
Mexico	-0.95	0.18	$2 \times 10^{-10}$	-0.85	-1.79	0.20	$9 \times 10^{-8}$	0.02
Colombia	-0.41	0.03	$8 \times 10^{-7}$	-0.39	-0.45	0.03	$1 \times 10^{-5}$	0.46

We report results of logistic regressions (Logreg) using socioeconomic status ( $c_{SES}$ ) or socioeconomic status + ancestry ( $c_{SES} + c_{anc}$ ) as predictors of disease outcome.  $c$  refers to the logistic regression coefficients for each variable (shown in subscript)

<sup>a</sup> With ancestry as covariate

<sup>b</sup> With socioeconomic status as covariate

$p$  values. These results are consistent with the association between socioeconomic status and type 2 diabetes remaining strong after accounting for ancestry (Table 3).

We finally considered whether BMI confounds the observed associations between either socioeconomic status or non-European ancestry and type 2 diabetes. We determined that BMI is negatively correlated with socioeconomic status in Mexicans ( $-20\%$  correlation,  $p=0.002$ ) and Colombians ( $-7\%$  correlation,  $p=0.05$ ), but that correlations between BMI and European ancestry proportion are not statistically significant in Mexicans ( $-12\%$  correlation,  $p=0.06$ ) or in Colombians ( $-1\%$  correlation,  $p=0.72$ ). The significant negative correlation between BMI and socioeconomic status implies that BMI could potentially confound our analyses. However, inclusion of BMI as a covariate in our analyses did not change the effect size or statistical significance of the associations of socioeconomic status and non-European ancestry with type 2 diabetes either in Mexicans or in Colombians, even though BMI

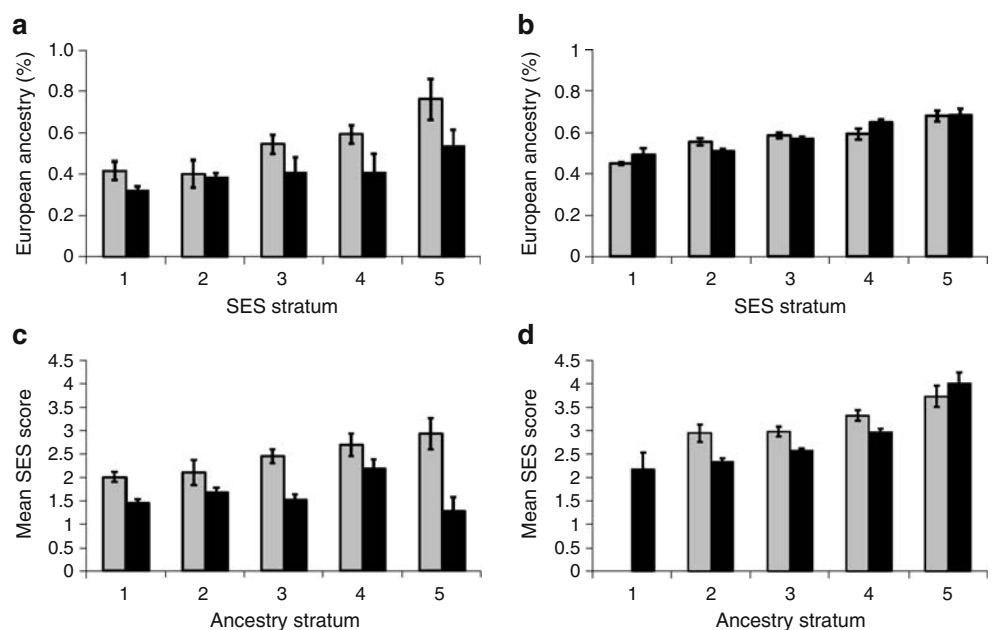
itself was associated with type 2 diabetes in both populations (Table 4).

## Discussion

These data demonstrate a genetic association between type 2 diabetes and individual non-European ancestry proportions in Latinos, while also showing that this evidence of association is highly confounded by socioeconomic status. Combining our study with previous results, this pattern has now been observed in North American, Central American and South American populations [13, 22, 23], with similar trends also noted in African-Americans [27].

Our study design has several limitations. First, our strategies for estimating socioeconomic status are necessarily imprecise and differ between both locations. Second, because type 2 diabetic participants and controls were ascertained separately, some bias may have been introduced

**Fig. 2** Distribution of ancestry proportions across strata of socioeconomic status (SES) in (a) Mexicans and (b) Colombians. c Distribution of mean socioeconomic status scores across strata of increasing proportion of European ancestry in Mexicans and (d) in Colombians. Error bars denote SEM. Black bars, type 2 diabetes; grey bars, controls



**Table 4** Association of socioeconomic status and non-European ancestry with type 2 diabetes is not confounded by BMI

Country of origin	Logreg coefficients			Pseudo- $r^2$	$p$ values		
	$c_{SES}$	$c_{anc}$	$c_{BMI}$		$c_{SES}^a$	$c_{anc}^b$	$c_{BMI}^c$
Mexico	-0.81	-1.72	0.08	0.22	$6 \times 10^{-7}$	0.02	0.03
Colombia	-0.37	-0.46	0.06	0.05	$3 \times 10^{-5}$	0.46	0.002

We report results of logistic regressions (Logreg) using socioeconomic status ( $c_{SES}$ ) + ancestry ( $c_{anc}$ ) + BMI ( $c_{BMI}$ ) as predictors of disease outcome.  $c$  refers to the logistic regression coefficients for each variable (shown in subscript)

<sup>a</sup> With ancestry and BMI as covariates

<sup>b</sup> With socioeconomic status and BMI as covariates

<sup>c</sup> With socioeconomic status and ancestry as covariates

despite best efforts to match recruitment procedures. Third, our sample size may have been too small to assess the relative effects of highly correlated, potentially confounding variables. And fourth, given the diversity inherent in Latinos, our findings may not be generalisable to all Latinos or to other populations with admixed Native American ancestry. Nevertheless, we believe we have taken analytical measures to address potential confounders and that the similar results obtained in two different populations strengthen our conclusions.

Our results show that, due to the correlation between socioeconomic status and Native American ancestry, it is difficult to disentangle the relationship between genetics and social factors in the contribution to disease risk. Low socioeconomic status can increase diabetes risk via a variety of mechanisms such as poor access to care, neglect of preventive strategies, a lower ability to exercise or an unhealthy diet. The question of whether the increased susceptibility to type 2 diabetes in Native American populations is caused by genetic or social factors (or a combination of both) is difficult to resolve accurately as long as low socioeconomic status is more prevalent among persons with greater Native American ancestry; answering it appropriately would require that admixed case-control cohorts be carefully matched for socioeconomic status, a much larger sample be used or twin studies discordant for socioeconomic status be undertaken. In some cases (e.g. the *ABCA1* R230C polymorphism), the strong association between a genetic variant and type 2 diabetes risk may be largely unaffected when adjusted for different confounders, including educational level as a surrogate of socioeconomic status [16, 17]. On the other hand, due to the strong association between Native American ancestry and socioeconomic status, use of the latter as a covariate in admixture studies might mask a true association signal in populations where low socioeconomic status is highly correlated to ancestral background.

The association between type 2 diabetes and Native American origin remained nominally significant after

adjustment for socioeconomic status in the Mexicans, but not in the Colombians. This may be due to a weaker initial signal in the Colombians, differences in socioeconomic status ascertainment between the two populations or a combination of both factors. Because the Mexican sample was enriched for early-onset cases ( $n=81$ ), it is possible that genetic susceptibility to type 2 diabetes may have been stronger among the Mexicans. Another potential reason for the stronger  $p$  value in Mexicans is the very different distribution of Native American heritage in Mexicans vs Colombians (Fig. 1a, b). Mexicans have a wide range of ancestry proportions, so enrichment of the type 2 diabetes group by individuals with more Native American ancestry could cause a wider separation of ancestry proportions that is easier to measure. Conversely, Colombians have a narrower range, which may be due to more generations of mixing and homogenisation [28]. Just as in African-Americans [26], this phenomenon limits the separation in ancestry proportion between cases and controls, even if a strong effect of ancestry causes oversampling of people who have inherited more of one ancestry. This may explain some of the discrepancies observed in the literature [22]. Thus it is possible that Pima Indians, like the samples in this study from central Mexico, have a wide range of ancestry proportions, making the oversampling of individuals with more extreme values of one ancestry easily detectable, whereas the San Luis Valley samples, like our samples from Colombia, have a narrow spread of ancestries making the effect more subtle. A higher degree of admixture and homogenisation in a society may also lead to decreased health disparities of a social nature. Alternatively, the lack of precision inherent in estimations of socioeconomic status (which often rely on information provided by participants who may have secondary motives for reporting a different stratum) may also have contributed to the observed differences. Our findings also illustrate the difficulties in making generalisations about Latino populations, which are often characterised by very diverse ancestral origins and environmental contexts that

may affect disease, socioeconomic status and their interactions.

These results also have implications for the prospects of gene mapping to find risk factors for type 2 diabetes in Latinos. Based on the epidemiological observation that type 2 diabetes is more prevalent in Latinos than in populations of European descent, it has been speculated that genetic risk factors for type 2 diabetes must be more common in Native Americans than in Europeans. Such variants could be located by admixture mapping, a technique that scans through the genome of affected individuals of mixed Native American, European and African ancestry, searching for regions with unusually high proportions of one ancestry compared with the genome-wide average. The observation that increased type 2 diabetes prevalence in Latinos is at least partially explained by environmental factors decreases the likelihood that significant genetic risk factors can be easily found through this approach, although it does not rule out the possibility that the method can work. These results do not imply that genome-wide association scans cannot detect variants associated with either increased or decreased Native American ancestry. Indeed, when a variant is much more common in one population than another, it will be easier to achieve genome-wide statistical significance in a genome-wide scan performed in the former than it will be in the latter [29–31].

It is important to recognise that socioeconomic status will not be a confounder in type 2 diabetes admixture mapping in the same way as it is for the present analysis. In admixture disease mapping, each locus in the genome is separately tested for association, using the rest of the genome as a control to assess whether the locus stands out. If an association is observed at any locus, it must signify a real genetic connection of that particular locus to disease (socioeconomic status is not expected to be locus-specific). However, because of the strong effect of socioeconomic status on type 2 diabetes risk in Latinos, rich information on socioeconomic status may be an important covariate that will increase statistical power in scans for genetic risk factors for type 2 diabetes. Tests for interactions between genes and environment in Latinos with type 2 diabetes may offer more power to detect risk factors for the disease than would be afforded if the modulatory effect of socioeconomic status on genetic risk were not taken into account.

**Acknowledgements** We thank all the participants for their contribution of phenotype and DNA samples, and S. Jiménez-Ramírez for technical assistance. J. C. Florez is supported by a Massachusetts General Hospital Physician Scientist Development Award and a Doris Duke Charitable Foundation Clinical Scientist Development Award. A. L. Price was supported by a Ruth Kirschstein National Research Service Award from the NIH. R. Saxena is supported by a Ruth Kirschstein National Research Service Award from the NIH. D. Reich is supported by a Burroughs Wellcome Career Development Award in the Biome-

dical Sciences. Support for this project was provided by discretionary funding from Harvard Medical School (to D. Reich), NIH grants NS043538 (to A. Ruiz-Linares) and DK073818 (to D. Reich), Colciencias grants 1115-04-16451 and 1115-04-012986 (to G. Bedoya and A. Villegas), and a Universidad de Antioquia grant (CODI/sostenibilidad 9889-E01321) to G. Bedoya and A. Villegas.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

## References

1. Carter JS, Pugh JA, Monterrosa A (1996) Non-insulin-dependent diabetes mellitus in minorities in the United States. *Ann Intern Med* 125:221–232
2. Bennett PH, Burch TA, Miller M (1971) Diabetes mellitus in American (Pima) Indians. *Lancet* 2:125–128
3. Hamman RF, Bennett PH, Miller M (1978) Incidence of diabetes among the Pima Indians. *Adv Metab Disord* 9:49–63
4. Knowler WC, Pettitt DJ, Saad MF, Bennett PH (1990) Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. *Diabetes Metab Rev* 6:1–27
5. Flegal KM, Ezzati TM, Harris MI et al (1991) Prevalence of diabetes in Mexican Americans, Cubans, and Puerto Ricans from the Hispanic Health and Nutrition Examination Survey, 1982–1984. *Diabetes Care* 14:628–638
6. Stern MP, Gaskill SP, Allen CR Jr, Garza V, Gonzales JL, Waldrop RH (1981) Cardiovascular risk factors in Mexican Americans in Laredo, Texas. II. Prevalence and control of hypertension. *Am J Epidemiol* 113:556–562
7. Hanis CL, Ferrell RE, Barton SA et al (1983) Diabetes among Mexican Americans in Starr County, Texas. *Am J Epidemiol* 118:659–672
8. Samet JM, Coultas DB, Howard CA, Skipper BJ, Hanis CL (1988) Diabetes, gallbladder disease, obesity, and hypertension among Hispanics in New Mexico. *Am J Epidemiol* 128:1302–1311
9. Hamman RF, Marshall JA, Baxter J et al (1989) Methods and prevalence of non-insulin-dependent diabetes mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am J Epidemiol* 129:295–311
10. Haffner SM, Stern MP, Hazuda HP, Pugh JA, Patterson JK (1986) Hyperinsulinemia in a population at high risk for non-insulin-dependent diabetes mellitus. *N Engl J Med* 315:220–224
11. Neel JV (1962) Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet* 14:353–362
12. Gardner LI Jr, Stern MP, Haffner SM et al (1984) Prevalence of diabetes in Mexican Americans. Relationship to percent of gene pool derived from Native American sources. *Diabetes* 33:86–92
13. Chakraborty R, Ferrell RE, Stern MP, Haffner SM, Hazuda HP, Rosenthal M (1986) Relationship of prevalence of non-insulin-dependent diabetes mellitus to Amerindian admixture in the Mexican Americans of San Antonio, Texas. *Genet Epidemiol* 3:435–454
14. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) Gm3;5, 13, 14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 43:520–526
15. Williams RC, Long JC, Hanson RL, Sievers ML, Knowler WC (2000) Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *Am J Hum Genet* 66:527–538
16. Villarreal-Molina MT, Aguilar-Salinas CA, Rodriguez-Cruz M et al (2007) The ATP-binding cassette transporter A1 R230C variant



- affects HDL cholesterol levels and BMI in the Mexican population: association with obesity and obesity-related comorbidities. *Diabetes* 56:1881–1887
17. Villarreal-Molina MT, Flores-Dorantes MT, Arellano-Campos O et al (2008) Association of the ATP-binding cassette transporter A1 R230C variant with early-onset type 2 diabetes in a Mexican population. *Diabetes* 57:509–513
  18. Reich DE, Gabriel SB, Altshuler D (2003) Quality and completeness of SNP databases. *Nat Genet* 33:457–458
  19. Tian C, Hinds DA, Shigeta R et al (2007) A genome-wide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet* 80:1014–1023
  20. Price AL, Patterson N, Yu F et al (2007) A genome-wide admixture map for Latino populations. *Am J Hum Genet* 80:1024–1036
  21. Mao X, Bigham AW, Mei R et al (2007) A genome-wide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet* 80:1171–1178
  22. Parra EJ, Hoggart CJ, Bonilla C et al (2004) Relation of type 2 diabetes to individual admixture and candidate gene polymorphisms in the Hispanic American population of San Luis Valley, Colorado. *J Med Genet* 41:e116
  23. Martinez-Marignac VL, Valladares A, Cameron E et al (2007) Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 120:807–819
  24. Silva Arciniega MR, Brain Calderon ML (2006) Validez y confiabilidad del estudio socioeconómico. Universidad Nacional Autónoma de México y Escuela Nacional de Trabajo Social, Mexico City
  25. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H (1999) Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A* 96:10016–10020
  26. Smith MW, Patterson N, Lautenberger JA et al (2004) A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 74:1001–1013
  27. Reiner AP, Ziv E, Lind DL et al (2005) Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet* 76:463–477
  28. Bedoya G, Montoya P, García J et al (2006) Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A* 103:7234–7239
  29. Yasuda K, Miyake K, Horikawa Y et al (2008) Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40:1092–1097
  30. Unoki H, Takahashi A, Kawaguchi T et al (2008) SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 40:1098–1102
  31. McCarthy MI (2008) Casting a wider net for diabetes susceptibility genes. *Nat Genet* 40:1039–1040