

## Short Report

# Association of *TCF7L2* polymorphisms with type 2 diabetes in Mexico City

Parra EJ, Cameron E, Simmonds L, Valladares A, McKeigue P, Shriver M, Wachter N, Kumate J, Kittles R, Cruz M. Association of *TCF7L2* polymorphisms with type 2 diabetes in Mexico City. Clin Genet 2007: 71: 359–366. © Blackwell Munksgaard, 2007

Polymorphisms within the transcription factor 7-like 2 gene (*TCF7L2*) have been associated with type 2 diabetes (T2D) in several recent studies. We characterized three of these polymorphisms (rs12255372, rs7903146 and the microsatellite DG10S478) in an admixed sample of 286 patients with T2D and 275 controls from Mexico City. We also analyzed three samples representative of the relevant parental populations: Native Americans from the state of Guerrero (Mexico), Spanish from Valencia and Nigerians (Bini from the Edo region). In order to minimize potential confounding because of the presence of population stratification in the sample, we evaluated the association of the three *TCF7L2* polymorphisms with T2D by using the program ADMIXMAP to fit a logistic regression model incorporating individual ancestry, sex, age, body mass index and education. The markers rs12255372, rs7903146 and DG10S478 are in tight disequilibrium in the Mexican sample. We observed a significant association between the single-nucleotide polymorphism (SNP) rs12255372 and the microsatellite DG10S478 with T2D in the Mexican sample [rs12255372, odds ratio (OR) = 1.78,  $p = 0.017$ ; DG10S478, OR = 1.62,  $p = 0.041$ ]. The SNP rs7903146 shows similar trends, but its association with T2D is not as strong (OR = 1.39,  $p = 0.152$ ). Analysis of the parental samples, as well as other available data, indicates that there are substantial population frequency differences for these polymorphisms: The frequencies of the T2D risk factors are more than 20% higher in European and West African populations than in East Asian and Native American populations.

**EJ Parra<sup>a</sup>, E Cameron<sup>a</sup>,  
L Simmonds<sup>a</sup>, A Valladares<sup>b</sup>,  
P McKeigue<sup>c</sup>, M Shriver<sup>d</sup>,  
N Wachter<sup>e</sup>, J Kumate<sup>f</sup>,  
R Kittles<sup>g</sup> and M Cruz<sup>b</sup>**

<sup>a</sup>Department of Anthropology, University of Toronto at Mississauga, Ontario, Canada, <sup>b</sup>Biochemistry Research Unit, Hospital de Especialidades, Centro Médico 'Siglo XXI,' Instituto Mexicano del Seguro Social, Mexico DF, Mexico, <sup>c</sup>Conway Institute, University College Dublin, Ireland, <sup>d</sup>Department of Anthropology, Penn State University, State College, PA, USA, <sup>e</sup>Clinical Epidemiology Research Unit, Hospital de Especialidades, Centro Médico 'Siglo XXI,' Instituto Mexicano del Seguro Social, Mexico DF, Mexico, <sup>f</sup>Fundación IMSS, Mexico DF, Mexico, and <sup>g</sup>Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL, USA

Key words: admixture – association – Mexico City – polymorphisms – *TCF7L2* – type 2 diabetes

Corresponding author: Esteban J Parra, Department of Anthropology, University of Toronto at Mississauga, 3359 Mississauga Road, Room 4026, South Building Mississauga, ON L5L 1C6, Canada.  
Tel.: +905-828-3889;  
fax: +905-828-3837;  
e-mail: eparra@utm.utoronto.ca

Received 30 October 2006, revised and accepted for publication 22 January 2007

A recent study by Grant et al. (1) reported the association between the microsatellite DG10S478 located within intron 3 of the transcription factor 7-like 2 gene (*TCF7L2*) with type 2 diabetes (T2D) in Icelandic individuals. Two single-nucleotide polymorphisms (SNPs) within this gene, rs12255372 and rs7903146, were tightly linked with the microsatellite and showed comparable levels of association with T2D (1). These results

were replicated in two additional cohorts from Denmark and the United States (1). Since the publication of this study, several other reports have confirmed the original findings (2–9). The available evidence indicates that *TCF7L2* is the most important T2D susceptibility gene described to date; in European populations, each copy of the susceptibility alleles increases T2D risk 1.4- to 1.5-fold, and given the frequencies of these alleles,

the corresponding population attributable risk (PAR) is quite large, around 15–20%. Unfortunately, there is insufficient data for other population groups. Presently, the mechanism by which *TCF7L2* increases diabetes risk remains unknown, but it has been speculated that it could be through regulation of the hormone glucagon-like peptide (GLP-1), which has a primary role in glucose homeostasis (1, 10). Interestingly, evidence obtained in the Diabetes Prevention Program (DPP) study and a recent large study of more than 8000 individuals in family-based and case-control designs in Scandinavia, Poland and the United States indicates that the risk-conferring genotypes are associated with measures of insulin secretion (impaired beta-cell function) but not with insulin resistance (2, 6).

We analyzed the microsatellite DG10S478 and the SNPs rs12255372 and rs7903146 in a sample of 286 unrelated patients with T2D and 275 controls from Mexico City. This is an admixed sample with a genetic background derived from Native American, European and to a lesser extent, West African populations. It is important to point out that evidence of linkage of T2D to a broad region on chromosome 10q, where *TCF7L2* is located, has been described in Mexican Americans (11) and homozygotes for the TT genotype at rs7903146 were associated with an increased T2D risk in the DPP Hispanic sample (2). In order to minimize potential confounding because of the presence of population stratification in the admixed sample (12–14), we evaluated the association of the three *TCF7L2* polymorphisms with T2D by using the program ADMIXMAP to fit a logistic regression model incorporating individual ancestry, sex, age, body mass index (BMI) and education. We also genotyped these polymorphisms in three parental samples (Native Americans from the state of Guerrero in Mexico, Spanish and Nigerians) and discuss the implications of the observed allele distributions for T2D risk.

## Subjects, materials and methods

### Samples

Samples from 561 unrelated individuals from Mexico City were collected between the years 2000 and 2005 by the Biochemistry and Clinical Epidemiology Research Units of the Medical Center ‘Siglo XXI,’ which is a major hospital complex that belongs to the Mexican Institute of Social Security. Patients with T2D were recruited from their Family Medicine (Primary Care) clinics within the first 2 years of diagnosis, which was made according to the American Diabetes Association (ADA) criteria.

The sample consisted of 286 patients with T2D (198 women and 88 men) and 275 controls (86 women and 189 men). The controls were selected based on: (i) absence of family history of diabetes in the previous two generations and (ii) negative results for the glucose tolerance test, according to the ADA criteria. The sample of patients with T2D and controls come from the area of Mexico City served by the Medical Center ‘Siglo XXI.’ Data on sex, age and BMI were also available. Information on years of education was also collected for the majority of the subjects and was used as a proxy for socioeconomic status. Although no information was available for other measures related to socioeconomic status, such as income or occupation, available data indicate that in Mexico education accounts for a major proportion of the variation observed in earnings (15). Table 1 provides additional details about the sample. Informed consent was obtained from each participant, and the research was approved by the ethical research board (Comité Local de Investigación) of the Medical Center ‘Siglo XXI.’ This study was also approved by the Ethics Review Office at the University of Toronto. We also included in the analysis three samples representative of the Mexican parental populations: a Native American sample from the state of Guerrero (Mexico), including Nahuatl, Mixtec and Tlapanec individuals, a sample from Spain (Valencia) and a sample from Nigeria (Bini from the Edo region).

### Characterization of admixture using ancestry informative markers

We used 69 ancestry informative markers (AIMs) to estimate admixture proportions in the sample from Mexico City. The average frequency differences

Table 1. Details about the sample of patients with T2D and controls from Mexico City

	Controls	Patients with T2D
Number	275	286
Sex (% women)	31.2	69.2
Age (SD)	40.18 (9.272)	55.50 (10.069)
BMI (SD)	27.648 (4.310)	29.673 (4.692)
Education <sup>a</sup> (SD)	2.57 (1.131)	1.63 (0.937)
Native American admixture (SD)	0.642 (0.126)	0.655 (0.112)
European admixture (SD)	0.318 (0.123)	0.301 (0.105)
West African admixture (SD)	0.040 (0.022)	0.044 (0.028)

BMI, body mass index; SD, standard deviation; T2D, type 2 diabetes.

<sup>a</sup>Primary school = 1, secondary school = 2, preparatory school = 3, university degree and/or postgraduate = 4.

between the parental populations (delta) for this panel of markers are 44% for European/Native American populations, 41% for European/West African populations and 51% for Native American/West African populations. Therefore, this panel of AIMs is extremely informative for estimating admixture at the individual level. A detailed description of admixture in this sample has been presented elsewhere (12).

Characterization of the SNPs rs12255372 and rs7903146 and the microsatellite DG10S478

SNP genotyping for rs12255372 and rs7903146 was carried out by conventional 3% agarose gel electrophoresis. Polymerase chain reaction (PCR) cycling for rs12255372 was performed at a temperature of 64.7°C using primers 5'-CCCAGGAATATC-CAGGCAAGGAT-3' (forward) and 5'-CAAA-TGGAGGCTGAATCTGGCA-3' (reverse). Prior to electrophoresis, PCR products were digested with the restriction enzyme *BseGI*. PCR amplification for rs7903146 was carried out at a temperature of 58.9°C using primers 5'-TTAGAGAGCTAAG-CACTTTT TAGGTA-3' (forward) and 5' ACTA-AGTTACTTGCCTTCCCTG-3' (reverse). PCR products were then digested with the restriction enzyme *RsaI*. The microsatellite DG10S478 was amplified using the primers described by Grant et al. (1) at a temperature of 61.4°C and genotyped on a Beckman Coulter CEQ8000 automated capillary sequencer. The genotyping success rates were 97.9% for rs12255372, 98.8% for rs7903146 and 99.1% for DG10S478.

#### Power analysis

Given the frequencies observed in the Mexican population for the putative high-risk alleles and assuming genotype relative risks of 1.45 for the heterozygotes and 2.41 for the homozygotes for the high-risk alleles (1), our case-control study has 73%, 84% and 79% power to detect significant differences at an alpha level of 5% for the markers rs12255372, rs7903146 and DG10S478, respectively. Assuming lower relative risks of 1.3 for the heterozygotes and 1.66 for the homozygotes for the high-risk alleles (3), the power of our study to detect significant differences is 38%, 45% and 42% for the markers rs12255372, rs7903146 and DG10S478, respectively. The power calculations were carried out as described by Purcell et al. (16).

#### Statistical analysis

Deviations of genotype proportions from Hardy-Weinberg expectations were evaluated using an

exact test implemented in the program GDA (17). The extent of linkage disequilibrium (LD) between genetic markers was estimated using an expectation-maximization algorithm implemented in the program 3LOCUS, made available to us by Dr Jeff Long. We report the normalized disequilibrium or  $D'$  coefficient (18) and, for diallelic markers, the squared correlation measure or  $r^2$  (19). Deviations from the theoretical haplotype frequencies were tested using the likelihood ratio statistic ( $G$ ), which has a  $\chi^2$  distribution for large sample sizes.

In order to control for potential confounding because of population stratification as a result of admixture, we used the program ADMIXMAP to build a logistic regression model with T2D as an outcome variable, including individual ancestry, age, sex, BMI and education in the model. ADMIXMAP is a general-purpose program for modeling population admixture with genotype and phenotype data, based on a combination of Bayesian and classical methods. Score tests for allelic association with T2D, conditional on individual admixture and age, sex, BMI and education, were carried out using ADMIXMAP. The parameter tested is the coefficient  $b$  for the effect of the allele under study (coded as 0, 1, or 2 copies) in a regression model that includes admixture and other covariates. We run the program using 1000 iterations for the burn-in period and 25,000 iterations to collect parameter data. We used frequency data obtained from the parental population samples (Native Americans from Mexico, Spanish, and Nigerians) as a prior allele frequency file. ADMIXMAP uses this information to estimate ancestry-specific allele frequencies from the unadmixed and admixed population samples, simultaneously, allowing for sampling error. The program ADMIXMAP has been described in detail by Hoggart et al. (14) and it is freely available at the Genetic Epidemiology Unit (University College, Dublin) Website (20). We also estimated conventional odds ratios (ORs) and 95% confidence intervals for alleles and genotypes, using tests adapted from Sasieni (21) and available at the Institut für Humangenetik Web site (22).

#### Results

Table 2 depicts the genotype and allele frequencies for the SNPs rs12255372 and rs7903146, and the microsatellite DG10S478, in the sample of T2D individuals and controls from Mexico City and in the three relevant parental samples (Native Americans from Mexico, Spanish and Nigerians). The frequency of the rs12255372 T allele was 0.116 in the controls and 0.155 in the cases. The frequency

Table 2. Distribution of genotype and allele frequencies for three polymorphisms within the *TCF7L2* gene (rs12255372, rs7903146 and the microsatellite DG10S478) in an admixed sample of patients with T2D and controls from Mexico City and in three parental population samples (Nahua from Mexico, Spanish and Nigerians)

Marker	Mexico City		Parental samples		
	Controls	T2D cases	Nahua (Mexico)	Spanish	Nigerians
rs12255372					
GG	210	205	87	17	43
GT	54	65	1	19	28
TT	4	11	0	1	9
Allele G	0.884	0.845	0.994	0.716	0.712
Allele T	0.116	0.155	0.006	0.284	0.288
rs7903146					
CC	191	185	76	17	37
CT	73	87	13	17	37
TT	7	11	0	2	7
Allele C	0.839	0.807	0.927	0.708	0.685
Allele T	0.161	0.193	0.073	0.292	0.315
DG10S478					
0/0	203	202	89	17	50
0/4	7	3	0	0	15
0/8	14	11	0	4	2
0/12	41	44	1	16	3
0/16	4	5	0	0	0
0/other	0	5	0	0	3
4/4	0	0	0	0	1
4/8	1	0	0	0	3
4/12	1	2	0	0	1
8/12	0	2	0	0	0
12/12	2	6	0	1	0
12/16	1	2	0	0	0
Other/other	0	0	0	0	0
Allele 0	0.861	0.837	0.994	0.710	0.788
Allele 4	0.017	0.009	0.000	0.000	0.135
Allele 8	0.027	0.023	0.000	0.053	0.032
Allele 12	0.086	0.110	0.006	0.237	0.026
Allele 16	0.009	0.012	0.000	0.000	0.000
Other alleles	0.000	0.009	0.000	0.000	0.019

of the rs7903146 T allele was slightly higher, 0.161 in the controls and 0.193 in the cases. We identified seven alleles at the microsatellite DG10S478 in our sample. The most common microsatellite allele, labeled as 0 in accordance with Grant et al. (1) nomenclature, is present at a higher frequency in the control group (0.861) than the T2D group (0.837). Conversely, the second most frequent allele, allele 12, has a higher frequency in the cases than the controls (0.110 vs 0.086, respectively). The other alleles are present in relatively low frequencies (<3%) in the admixed sample from Mexico City. There were no significant deviations from Hardy–Weinberg proportions for any of the markers. With regard to the parental samples, in the Native American sample from Mexico ( $n = 89$ ) the frequency of the rs7903146 C allele was 0.927 and the frequency of the rs7903146 T allele was 0.073. Markers rs12255372 and DG10S478 showed very little variation in the Native American sample. Most individuals were homozygotes for the

rs12255372 G or DG10S478 0 allele, and we only identified one heterozygote rs12255372 G–T and one heterozygote DG10S478 0–12. The frequency of the rs12255372 T allele and the DG10S478 12 allele was lower than 1% (Table 2). In the sample from Spain ( $n = 38$ ), the frequency of the rs12255372 T allele was 0.314, the frequency of the rs7903146 T allele was 0.281, and we observed three alleles in the microsatellite DG10S478: allele 0 (0.689), allele 8 (0.054) and allele 12 (0.257). Finally, in the sample from Nigeria ( $n = 81$ ), the frequency of the rs12255372 T allele was 0.288, the frequency of the rs7903146 T allele was 0.315, and there were five alleles in the microsatellite DG10S478: allele –4 (0.019), allele 0 (0.788), allele 4 (0.135), allele 8 (0.032) and allele 12 (0.026).

We also evaluated the extent of LD in the sample from Mexico City and the three parental samples. The  $D'$  coefficients for all possible pairs of markers are shown in Table 3. For the diallelic markers, we also report the  $r^2$  measure of allelic association. The

Table 3. Patterns of LD between pairs of markers within the *TCF7L2* gene

Sample	rs12255372/rs7903146	rs12255372/DG10S478	rs7903146/DG10S478
	D' (r <sup>2</sup> )	D' (r <sup>2</sup> )	D' (r <sup>2</sup> )
Mexico City	T/T: 0.871 (0.562)	G/-16: 1.000 G/0: 0.936 T/4: 0.172 T/8: 0.875 T/12: 0.923 T/16: 1.000 T/20: 1.000	C/-16: 1.000 C/0: 0.773 C/4: 0.316 T/8: 0.783 T/12: 0.889 T/16: 1.000 C/20: 0.348
Nahua	<b>p &lt; 0.0001</b> T/T: 1.000 (0.068)	<b>p &lt; 0.0001</b> T/12: 1.000 (0.850)	<b>p &lt; 0.0001</b> T/12: 1.000 (0.085)
Spanish	<b>p = 0.047</b> T/T: 0.927 (0.860)	<b>p = 0.0009</b> G/0: 0.929 T/8: 0.647 T/12: 1.000	<b>p = 0.042</b> C/0: 1.000 T/8: 1.000 T/12: 1.000
Nigerians	<b>p &lt; 0.0001</b> T/C: 0.324 (0.021)	<b>p &lt; 0.0001</b> G/-4: 1.000 G/0: 0.426 T/4: 0.351 T/8: 0.625 T/12: 1.000	<b>p &lt; 0.0001</b> T/-4: 0.289 C/0: 0.302 T/4: 0.199 T/8: 0.501 T/12: 0.602
	p = 0.122	<b>p &lt; 0.0001</b>	<b>p = 0.012</b>

LD, linkage disequilibrium. The extent of LD is reported as D' (normalized LD) and r<sup>2</sup>. Significant departures from LD are indicated in bold.

three markers were in tight LD in the samples from Mexico City and Spain, with the rs12255372 T allele strongly associated with the rs7903146 T allele. Similarly, the rs12255372 T allele and the rs7903146 T alleles show strong LD with DG10S478 alleles 8 and 12, while the rs12255372 G and the rs7903146 C alleles are strongly associated with the microsatellite 0 allele. Overall, the associations between these markers are highly significant (p < 0.0001 in all pairwise comparisons). However, in the Nigerian sample, the pattern of LD was different from what is observed in the samples from Spain and Mexico City. Marker rs12255372 shows significant LD with the microsatellite DG10S478 (p < 0.0001), with similar allele associations as found in the other two populations. However, the locus rs7903146 has a much weaker LD with rs12255372 and DG10S478 (p = 0.122 and p = 0.012, respectively). This is consistent with the LD patterns between rs7903146 and rs12255372 recently described in the African American DPP sample (2). The LD patterns observed in the Native American sample that we analyzed need to be interpreted with caution, because of the very low frequency of the rs12255372 T and the DG10S478 12 alleles (0.006). In fact, the D' and r<sup>2</sup> measures are very divergent in this sample, in agreement with previous reports indicating that D' tends to be inflated for loci with rare alleles (23).

Table 4 shows the results of the analysis of the effect of the three *TCF7L2* polymorphisms on diabetes risk in the sample from Mexico City. We

report the ORs estimated using the program ADMIXMAP, which incorporates individual ancestry and covariates such as sex, age, BMI and education in the analysis. In agreement with previous studies, the analysis using ADMIXMAP shows evidence of association of the rs12255372 T allele with T2D. The OR associated with each copy of the rs12255372 T allele was estimated as 1.78 (p = 0.017). The association of the DG10S478 X allele (a category comprising all non-zero alleles) with T2D was also significant (OR = 1.62, p = 0.041). Similar results were obtained when analyzing all the common microsatellite alleles separately (data not shown). The results for the rs7903146 T allele were not significant (OR = 1.39, p = 0.152). Table 4 also shows the conventional OR statistics for alleles and genotypes without controlling for relevant covariates. The analysis using the conventional OR statistics for the alleles of each marker shows the same trends previously described for ADMIXMAP, but with somewhat lower ORs and p-values (rs12255372 T allele, OR = 1.40, p = 0.058; DG10S478 X allele, OR = 1.21, p = 0.255; rs7903146 T allele, OR = 1.25, p = 0.162).

### Discussion

In this article, we show that, in agreement with previous studies in other populations (1–9), polymorphisms within the *TCF7L2* gene show

Table 4. Analysis of the effect of three polymorphisms within the *TCFL72* gene on diabetes risk in a sample from Mexico City, using the program ADMIXMAP

	OR	95% CI	p-value
rs12255372			
ADMIXMAP (effect of each copy of allele T)	<b>1.78</b>	<b>1.11–2.88</b>	<b>0.017</b>
Conventional ORs			
Allele T vs allele G	1.40	0.99–1.99	0.058
GT vs GG	1.23	0.82–1.86	0.315
TT vs GG	2.82	0.88–8.99	0.069
rs7903146			
ADMIXMAP (effect of each copy of allele T)	1.39	0.89–2.17	0.152
Conventional ORs			
Allele T vs allele G	1.25	0.92–1.70	0.162
CT vs CC	1.23	0.85–1.78	0.273
TT vs CC	1.62	0.62–4.28	0.324
DG10S478			
ADMIXMAP (effect of each copy of allele X <sup>a</sup> )	<b>1.62</b>	<b>1.02–2.57</b>	<b>0.041</b>
Conventional ORs			
Allele 0 vs allele X	1.21	0.87–1.68	0.255
OX vs 00	1.04	0.70–1.53	0.861
XX vs 00	2.41	0.84–6.97	0.094

OR, odds ratio, CI, confidence interval. The analysis incorporates individual ancestry and sex, age, BMI and education as covariates, minimizing potential confounding because of admixture stratification in the Mexico City sample. Significant values are indicated in bold. We also report conventional OR estimates for alleles and genotypes.

<sup>a</sup>All non-zero alleles.

a significant association with T2D in an admixed sample from Mexico City. We have previously evaluated the admixture proportions in this sample using a panel of 69 AIMs and estimated that the Native American, European and West African genetic contributions are 65%, 30% and 5%, respectively (12). Others and we have previously shown that in admixed populations the presence of admixture stratification can be an important confounder in population-based association studies (12–14). In fact, the sample from Mexico City shows strong evidence of stratification because of admixture. There is substantial dispersion in the individual admixture proportions within the sample, and a test for residual allelic association between unlinked loci implemented in the program ADMIXMAP indicates significant evidence for residual stratification (12). Similarly, we observed that 23% of unlinked AIMs show significant associations in this sample. This percentage is substantially higher than the 5% expected and indicates that there is substantial admixture stratification (12). In admixed populations, the effect of population stratification in candidate gene association studies can be controlled for by including admixture proportions as a variable in a regression analysis (12). In this study, we used the program ADMIXMAP to evaluate the association of three *TCFL72* polymorphisms (rs12255372, rs7903146 and the microsatellite DG10S478) with T2D controlling for individual ancestry and other covariates having

a significant effect on T2D risk (sex, age, BMI and education) (12).

The analysis of the rs12255372, rs7903146 and DG10S478 loci in the sample of controls and patients with T2D from Mexico City indicates that the rs12255372 T allele and the DG10S478 X allele (all non-zero alleles) are significantly associated with diabetes risk. The OR for the rs12255372 T allele is 1.78 ( $p = 0.017$ ), which is slightly higher than previously reported results in large samples that indicate an effect of approximately 1.3–1.6 (1–7). The OR for the DG10S478 X allele is 1.62 ( $p = 0.041$ ). Most of the signal in this marker is because of allele 12, which is present at a frequency of 0.086 in the controls and 0.110 in the patients with T2D. In agreement with previous studies, the rs7903146 T allele also has a higher frequency in the patients with T2D than the controls (0.193 vs 0.161), but after controlling for individual ancestry, sex, age, BMI and education, the difference is not significant (OR = 1.39,  $p = 0.152$ ). Analysis of the data using conventional OR statistics, which do not take into account the effect of individual ancestry and other covariates in the analysis, shows similar trends but with lower OR and p-values (Table 4). It is important to note that, given the presence of population stratification in the Mexico City sample, it becomes critical to control for potential confounders, and the analysis using ADMIXMAP is the preferred type of analysis. The program ADMIXMAP takes into account the differences in admixture proportions between the

control and the case groups. As shown in Table 1, the control group has a higher European ancestry than the group of patients with T2D (1.7% higher). Therefore, we would expect that, given the substantially higher frequency of the risk alleles rs12255372 T and DG10S478 X alleles in the European sample with respect to the Native American sample (more than 25% higher, see Table 2), the frequency of these alleles should be higher in the control sample than in the sample of patients with T2D. In fact, we observed the reversed pattern, which explains the higher OR values obtained using ADMIXMAP and strongly supports previous results indicating that rs12255372 T and DG10S478 X increase diabetes risk. In accordance to previous studies (1–9), our data are also consistent with a multiplicative or additive genetic model, instead of a dominant or recessive model (Table 4).

The analysis of the allele frequency and LD data in the parental samples (Native Americans from Mexico, Spanish, and Nigerians) offers interesting insights on the potential impact of *TCF7L2* polymorphisms on T2D risk in different populations. In this sense, it is important to mention that the frequencies of the putative risk alleles (rs12255372 T, rs7903146 T and DG10S478 X) are substantially higher in European and West African populations than in Native American populations. In the Native American sample ( $n = 90$ ) from the state of Guerrero in Mexico, comprised of individuals of Nahua, Tlapanec and Mixtec ancestry, the frequency of the rs12255372 T and DG10S478 X allele (in this case, allele 12) was 0.006 (Table 2). The frequency of the rs7903146 T allele was slightly higher (0.073). In agreement with our data, the frequencies of the rs12255372 T allele (0.05) and the rs7903146 T allele (0.12) were also substantially lower in the American Indian DPP sample than in the European and African American DPP samples (2). Interestingly, the alleles rs12255372 T and rs7903146 T are also absent or present at very low frequency (less than 2.5%) in Chinese and Japanese populations (24). Therefore, the available data indicate that there are substantial frequency differences between populations for these three polymorphisms, and these differences may have important epidemiological implications. For the locus rs12255372, assuming risk ratios of 1.30 for the heterozygotes GT and 1.66 for the homozygotes TT [based on the study with the largest sample size, (3)] the PAR in the Spanish population and the Nigerian populations would be around 15%, while in the Japanese population (24) would be around 1% and in the Chinese (24) and Mexican Native American populations would

be lower than 1%. Similarly, there are large differences in PAR for the markers DG10S478 and rs7903146. The frequency distributions observed in the sample from Mexico City for the three *TCF7L2* polymorphisms reflects the history of recent admixture of this population, with intermediate frequencies between the Native American and European samples. Finally, it is important to discuss the pattern of LD observed in the samples analyzed in this study. Markers rs12255372, rs7903146 and DG10S478 X show strong LD in the population of European ancestry and the sample from Mexico City, but this is not the case in the sample from West Africa (see Table 2). This is not surprising given the well-known fact that LD is typically lower in African than non-African populations (25, 26). The pattern of LD that we observed in the Nigerian sample is similar to the pattern described by Florez et al. (2) in the DPP African American sample. The lower LD observed in West African groups implies a higher resolution to identify the causal variants. Additional studies in populations of African ancestry could provide further insights on the relative roles of the rs12255372 and rs7903146 loci, or yet unidentified variants in LD with these two polymorphisms, in T2D risk.

### **Acknowledgements**

We would like to thank all the individuals from Mexico City for their participation in this study. We are also grateful to the personnel at the National Blood Bank, Medical Center Siglo XXI (Instituto Mexicano del Seguro Social, IMSS) in Mexico. This work was supported by the Fondo para el Fomento de Investigacion and the IMSS in Mexico. Estoban J. Parra was supported by the Canadian Institutes of Health Research, the Canada Foundation for Innovation, the Ontario Innovation Trust and by the Banting and Best Diabetes Centre at the University of Toronto. Mark D. Shriver and Paul McKeigue are supported by a grant from the National Institutes of Health/ National Human Genome Research Institute (HG02154).

### **References**

1. Grant SF, Thorleifsson G, Reynisdottir I et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet* 2006; 38: 320–323.
2. Florez JC, Jablonski KA, Bayley N et al. *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 2006; 355: 241–250.
3. Groves CJ, Zeggini E, Minton J et al. Association analysis of 6,736 U.K. subjects provides replication and confirms *TCF7L2* as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 2006; 55: 2640–2644.
4. Zhang C, Qi L, Hunter DJ et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* 2006; 55: 2645–2648.

5. Scott LJ, Bonnycastle LL, Willer CJ et al. Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. *Diabetes* 2006; 55: 2649–2653.
6. Saxena R, Gianniny L, Burtt NP et al. Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce insulin response to glucose in nondiabetic individuals. *Diabetes* 2006; 55: 2890–2895.
7. Cauchi S, Meyre D, Dina C et al. Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. *Diabetes* 2006; 55: 2903–2908.
8. Damcott CM, Pollin TI, Reinhart LJ et al. Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish. *Diabetes* 2006; 55: 2654–2659.
9. Vliet-Ostaptchouk JV, Shiri-Sverdlov R, Zhernakova A et al. Association of variants of transcription factor 7-like 2 (TCF7L2) with susceptibility to type 2 diabetes in the Dutch Breda cohort. *Diabetologia* 2007; 50: 59–62.
10. Yi F, Brubaker PL, Jin T. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. *J Biol Chem* 2005; 280: 1457–1464.
11. Duggirala R, Blangero J, Almasy L et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 1999; 64: 1127–1140.
12. Martinez-Marignac VL, Valladares A, Cameron E et al. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 2007; 120: 807–819.
13. Choudhry S, Coyle NE, Tang H et al. Population stratification confounds genetic association studies among latinos. *Hum Genet* 2006; 118: 652–664.
14. Hoggart CJ, Parra EJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003; 72: 1492–1504.
15. Lopez-Acevedo G. Evolution of earnings and rates of returns to education in Mexico. World Bank Policy Research Working Paper No. 2601. October 2001.
16. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19: 149–150.
17. Lewis PO, Zaykin D. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0, 2001.
18. Lewontin RC. The interaction of selection and linkage. I. General considerations, heterotic models. *Genetics* 1964; 49: 49–67.
19. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968; 38: 226–231.
20. <http://www.ucd.ie/genepi/admixmap/>. October 2007.
21. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics* 1997; 53: 1253–1261.
22. <http://ihg.gsf.de/cgi-bin/hw/hwa1.pl>. October 2007.
23. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; 3: 299–309.
24. <http://www.hapmap.org>. October 2007.
25. Reich DE, Cargill M, Bolk S et al. Linkage disequilibrium in the human genome. *Nature* 2001; 411: 199–204.
26. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003; 4: 587–597.