Human Molecular Genetics, 2010, Vol. 19, No. 14 ***2877–2885***
doi:10.1093/hmg/ddq173
Advance Access published on April 23, 2010

A functional *ABCA1* gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans

Víctor Acuña-Alonzo[1,5,†], Teresa Flores-Dorantes[1,†], Janine K. Kruit[6], Teresa Villarreal-Molina[1], Olimpia Arellano-Campos[2], Tábita Hünemeier[7], Andrés Moreno-Estrada[8], Ma Guadalupe Ortiz-López[9,10], Hugo Villamil-Ramírez[1], Paola León-Mimila[1], Marisela Villalobos-Comparan[1], Leonor Jacobo-Albavera[1], Salvador Ramírez-Jiménez[1], Martin Sikora[11], Lin-Hua Zhang[6], Terry D. Pape[6], Ma de Ángeles Granados-Silvestre[9,10], Isela Montufar-Robles[9,10], Ana M. Tito-Alvarez[12], Camilo Zurita-Salinas[12], José Bustos-Arriaga[13], Leticia Cedillo-Barrón[13], Celta Gómez-Trejo[5], Rodrigo Barquera-Lozano[5], Joao P. Vieira-Filho[14], Julio Granados[3], Sandra Romero-Hidalgo[15], Adriana Huertas-Vázquez[16], Antonio González-Martín[17], Amaya Gorostiza[5,17], Sandro L. Bonatto[18], Maricela Rodríguez-Cruz[19], Li Wang[20], Teresa Tusié-Luna[1], Carlos A. Aguilar-Salinas[2], Ruben Lisker[4], Regina S. Moises[14], Marta Menjivar[9], Francisco M. Salzano[7], William C. Knowler[21], M. Cátira Bortolini[7], Michael R. Hayden[6], Leslie J. Baier[21] and Samuel Canizales-Quinteros[1,*]

[1]Unit of Molecular Biology and Genomic Medicine, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ), Universidad Nacional Autónoma de México (UNAM), Mexico City 14000, Mexico, [2]Department of Endocrinology and Metabolism, [3]Department of Immunology and Rheumatology and [4]Direction of Research, INCMNSZ, Mexico City 14000, Mexico, [5]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Mexico City 14030, Mexico, [6]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada V6T1Z4, [7]Departamento de Genetica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil, [8]Department of Genetics, Stanford University, Stanford, CA 94305, USA, [9]Department of Biology, Faculty of Chemistry, Universidad Nacional Autónoma de México (UNAM), Mexico City 04510, Mexico, [10]Laboratory of Molecular Endocrinology, Hospital Juárez de México, Mexico City, Mexico, [11]Institute of Evolutionary Biology (UPF-CSIC), CEXS-UPF-PRBB, Barcelona 37-49 08003, Spain, [12]Laboratorios Zurita & Zurita, Quito N12-120, Ecuador, [13]Department of Molecular Biomedicine, Cinvestav, IPN, Mexico City 07360, Mexico, [14]Division of Endocrinology, Department of Medicine, Federal University of São Paulo, São Paulo 04034-970, Brazil, [15]Department of Computational Genomics, Instituto Nacional de Medicina Genómica, Mexico City 01900, México, [16]Cedars-Sinai Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048-5909, USA, [17]Department of Zoology and Physical Anthropology, Faculty of Biology, Complutense University of Madrid, Madrid 28040, Spain, [18]Faculty of Biological Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre 90619-900 RS, Brazil, [19]Unit of Medical Research in Nutrition, Pediatrics Hospital, National Medical Center XXI Century, Mexican Institute of Social Security, Mexico City 06725, Mexico, [20]Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China, [21]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ 85014, USA

Received January 7, 2010; Revised March 19, 2010; Accepted April 21, 2010

*To whom correspondence should be addressed at: Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Vasco de Quiroga, No. 15 Colonia Sección 16, Tlalpan 14000, México D.F., Mexico. Tel/Fax: +52 5556550011; Email: cani@servidor.unam.mx
†These authors contributed equally to this work.

It has been suggested that the higher susceptibility of Hispanics to metabolic disease is related to their Native American heritage. A frequent cholesterol transporter *ABCA1* (ATP-binding cassette transporter A1) gene variant (*R230C*, rs9282541) apparently exclusive to Native American individuals was associated with low high-density lipoprotein cholesterol (HDL-C) levels, obesity and type 2 diabetes in Mexican Mestizos. We performed a more extensive analysis of this variant in 4405 Native Americans and 863 individuals from other ethnic groups to investigate genetic evidence of positive selection, to assess its functional effect *in vitro* and to explore associations with HDL-C levels and other metabolic traits. The *C230* allele was found in 29 of 36 Native American groups, but not in European, Asian or African individuals. *C230* was observed on a single haplotype, and *C230*-bearing chromosomes showed longer relative haplotype extension compared with other haplotypes in the Americas. Additionally, single-nucleotide polymorphism data from the Human Genome Diversity Panel Native American populations were enriched in significant integrated haplotype score values in the region upstream of the *ABCA1* gene. Cells expressing the *C230* allele showed a 27% cholesterol efflux reduction ($P < 0.001$), confirming this variant has a functional effect *in vitro*. Moreover, the *C230* allele was associated with lower HDL-C levels ($P = 1.77 \times 10^{-11}$) and with higher body mass index ($P = 0.0001$) in the combined analysis of Native American populations. This is the first report of a common functional variant exclusive to Native American and descent populations, which is a major determinant of HDL-C levels and may have contributed to the adaptive evolution of Native American populations.

INTRODUCTION

It has been suggested that genetic susceptibility of Hispanics to type 2 diabetes (T2D), obesity and dyslipidemia is related to their Native American heritage (1–3). We recently found a frequent non-synonymous variant (*R230C*, rs9282541) within the ATP-binding cassette transporter A1 gene (*ABCA1*) associated with low high-density lipoprotein cholesterol (HDL-C) levels (the most common dyslipidemia in populations with Native American ancestry), obesity and T2D in Mexican Mestizos (4,5). ABCA1 plays a key role in cholesterol efflux and transfer from peripheral cells to lipid-poor apolipoprotein A1 (ApoA1), the first step in HDL particle formation (6,7).

The *R230C* variant was initially described in the Oji-Cree population (8). To date, it has been found only in Native American and Mexican-Mestizo populations (4). We performed a large-scale analysis including individuals from 36 Native North and South American groups and assessed the effect of this variant on anthropometric and metabolic traits. Because it was previously suggested that *R230C* may have conferred selective advantage as a thrifty gene and/or resistance against certain infectious diseases (4), we performed a more thorough analysis seeking evidence of positive selection.

RESULTS

The *C230* allele was present in the majority of the Native American populations at an average frequency of 12% (range 0–31%) (Fig. 1; Supplementary Material, Table S1), but was absent from 863 additional individuals belonging to different European (Spaniard and Dutch) and Asian groups (Han Chinese, Manchu, Mongolian, Siberians and Eskimos) (Fig. 1). The distribution of this allele was not structured according to language or geographic groups (North versus South America), as evidenced by analysis of molecular variance (AMOVA) ($P = 0.978$ and 0.895, respectively). However, the *C230* allele frequency increased at tropical latitudes (between the tropics of Cancer and Capricorn) and gradually decreased at higher latitudes both to the North and South (Fig. 1), showing a significant correlation ($r^2 = 0.328$; $P = 0.02$).

The C230 allele is located on a single genetic haplotype

To perform a phylogenetic reconstruction of the evolutionary relationships, 15 additional single-nucleotide polymorphisms (SNPs) within a 50 kb block were analyzed in 20 Native American and 25 Mexican-Mestizo trios to define haplotype blocks within the region. Together with data from HapMap populations, a total of 58 haplotypes were identified (Supplementary Material, Table S2). Seven haplotypes were found in Native Americans, and the *C230* allele was clearly found in only one genetic block (haplotype 32) in all Mexican-Mestizo, North and South American native individuals analyzed. Maximum parsimony (MP)-based network analysis is shown in Figure 2. The phylogenetic reconstruction uncovered two major lineages (haplogroups A and B) defined by the non-synonymous polymorphism *R219K* within the *ABCA1* gene. The *C230* allele occurred in haplogroup B characterized by the ancestral *R219* allele.

Positive selection testing

The long-range haplotype (LRH) test showed that the extension of linkage disequilibrium (LD) was much longer in *C230* than in non-*C230* chromosomes [relative extended haplotype homozygosity (REHH) = 9.8 and 5.9 at 365 and −434 kb from the core, respectively; Fig. 3]. The former value remained significant compared with REHH values from additional regions elsewhere in the genome, genotyped in a similar set of Native American populations ($P = 0.036$ at 300 kb from the core and $P = 0.021$ at a marker H of 0.04). REHH values were significant in both Kichwa ($P = 0.018$ at a distance of 300 kb and $P = 0.007$ at a marker H of 0.04) and Nahua populations ($P = 0.043$ at a distance of 300 kb and $P = 0.021$ at a marker H of 0.04). The only other two significant core haplotypes were located upstream

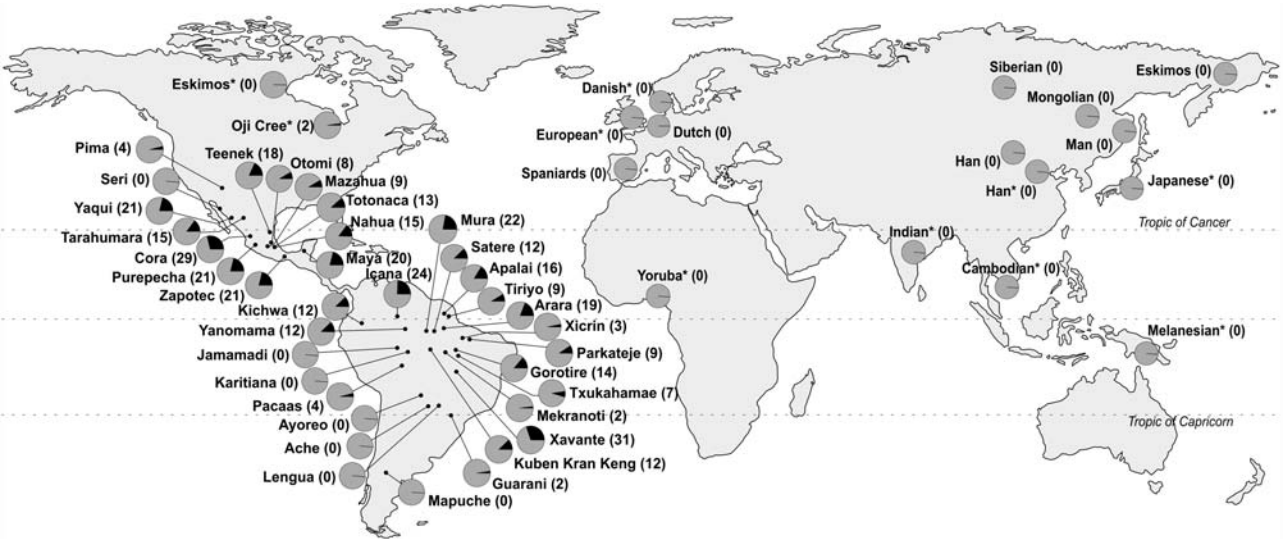

Figure 1. Frequency distribution of the *C230* allele [(%) black-shaded area] in Native American, European, Asian and African populations. *C230* frequency data from populations with an asterisk were obtained from previous reports (8–11 and HapMap and SNP500CANCER databases).

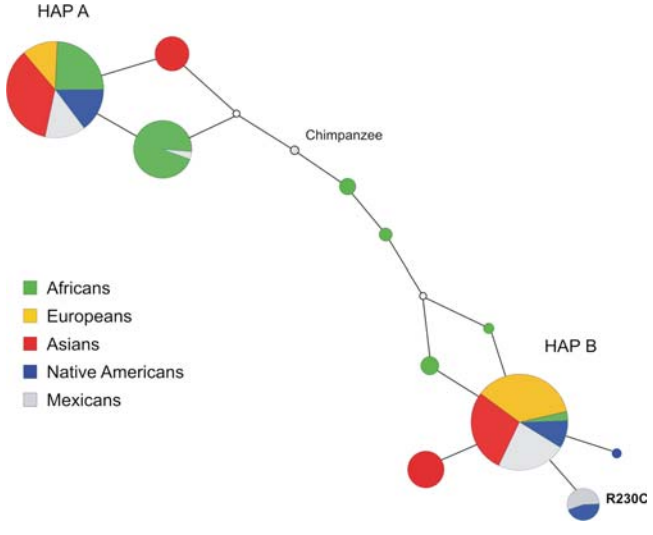

Figure 2. MP-based network describing the evolutionary relationships of 11 distinct haplotypes. Native American and Mexican-Mestizo haplotypes were established in trios; the remainder were inferred from HapMap groups. Each haplotype is represented by a circle whose area reflects the overall number of copies observed and whose color-coding indicates the frequency of the haplotype in the HapMap groups and Native Americans and Mexican Mestizos. Line length is proportional to the number of differences between haplotypes. Non-filled circles represent non-sampled haplotypes reconstructed by the MP algorithm as evolutionary intermediaries between observed haplotypes. The phylogenetic reconstruction uncovered two major lineages (haplogroups A and B) defined by the non-synonymous polymorphism *R219K*. The *R230C* variant occurred on haplogroup B characterized by the ancestral *R219* allele, which is frequent in Europe, Asia and America but infrequent in African populations. The *C230* allele was found in only one genetic block in all Mexican-Mestizo and Native American individuals analyzed.

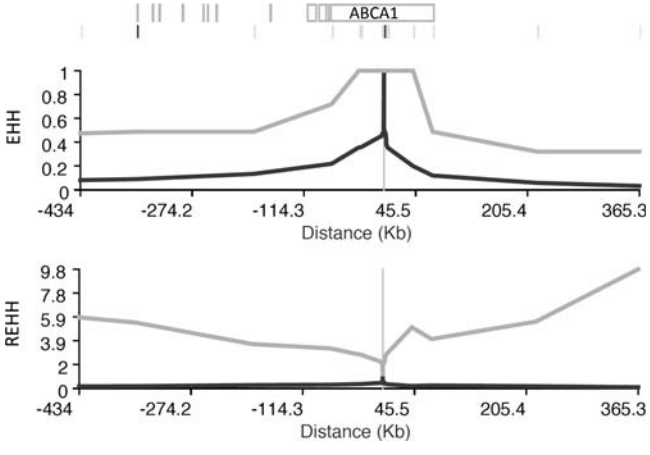

Figure 3. EHH and REHH of *ABCA1/R230C* and 23 additional SNPs × physical distance in Native American individuals. *C230*-bearing chromosomes (in gray) appear to have greatly extended LD compared with non-*C230*-bearing chromosomes (in black). Gray vertical lines in the upper track represent genotyped SNPs, and boxes indicate annotated genes.

the *ABCA1* gene (Supplementary Material, Table S3). Furthermore, in Native Americans from the Human Genome Diversity Panel (HGDP) (*R230C* genotypes not available), the *ABCA1* 5′ region (∼75 kb upstream *R230C*) was clearly enriched for outliers of the integrated haplotype score (iHS) statistic genome-wide distribution (iHS > 2.5) (Fig. 4). In agreement with the REHH analysis, the highest iHS values are clustered upstream *ABCA1*.

Association of R230C with HDL-C levels and other metabolic traits

Overall, the prevalence of hypoalphalipoproteinemia (HA) was the most common dyslipidemia (65% in Mexican and South American natives; Supplementary Material, Table S4). Table 1 shows the effect of *R230C* on HDL-C and total cholesterol levels and body mass index (BMI). The *R230C*/*C230C* genotypes were significantly associated with low HDL-C levels in Pimas ($P = 6.4 \times 10^{-5}$) and in the combined analysis of eight Mexican native groups ($P = 5.3 \times 10^{-8}$). In South

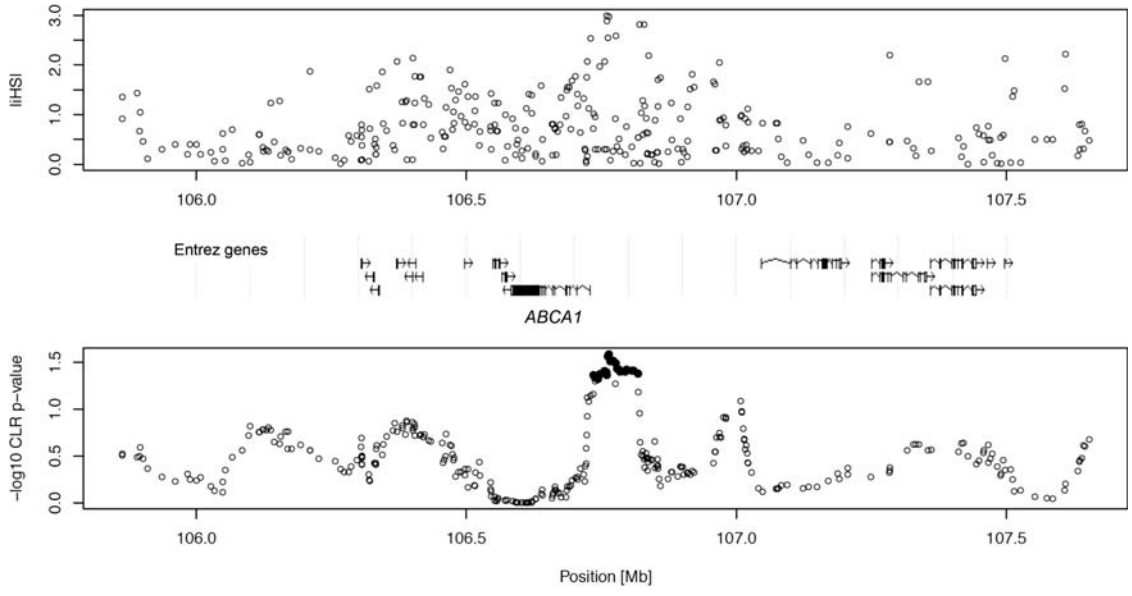

Figure 4. iHS values for individual SNPs flanking the *ABCA1* region (2 Mb) × physical distance (top panel) and *P*-values for the composite likelihood ratio (CLR) test based on a 31-SNP sliding window analysis to detect local regions enriched for high iHS values (bottom panel) in the combined Native American sample. Filled circles indicate *P*-values <0.05 genome-wide significance level.

Table 1. Association of *R230C* with lipid levels and BMI in Native American populations

Native American population (*n*)	HDL-C levels Effect (SE)	*P*-value	Total cholesterol Effect (SE)	*P*-value	BMI Effect (SE)	*P*-value
North America						
USA						
Pima (2563)[a]	−0.075 (0.019)	6.4×10^{-5}	−0.071 (0.014)	1.8×10^{-7}	0.008 (0.015)	0.586
Mexico						
Yaquis (45)	—	—	—	—	0.044 (0.018)	0.012
Teenek (67)	−0.051 (0.026)	0.057	−0.010 (0.024)	0.671	0.001 (0.016)	0.978
Coras (123)	−0.033 (0.014)	0.021	−0.006 (0.013)	0.681	0.032 (0.011)	0.006
Purepechas (15)	−0.040 (0.074)	0.603	−0.048 (0.046)	0.333	0.034 (0.019)	0.097
Mazahuas (83)	−0.039 (0.036)	0.281	−0.031 (0.041)	0.444	0.006 (0.019)	0.758
Nahuas (267)	−0.040 (0.014)	0.014	−0.014 (0.020)	0.470	−0.004 (0.008)	0.617
Totonacas (113)	−0.028 (0.021)	0.180	−0.031 (0.015)	0.042	0.022 (0.013)	0.085
Zapotecs (106)	−0.047 (0.022)	0.038	0.007 (0.019)	0.723	0.007 (0.013)	0.605
Mayans (110)	−0.040 (0.017)	0.023	−0.043 (0.015)	0.004	−0.007 (0.013)	0.554
Mexican natives combined	−0.038 (0.007)	5.3×10^{-8}	−0.019 (0.007)	0.027	0.010 (0.004)	0.012
South America						
Kichwas (79)	−0.043 (0.030)	0.153	−0.005 (0.020)	0.791	0.024 (0.012)	0.050
Parkatejé (78)	−0.029 (0.026)	0.270	−0.002 (0.030)	0.945	0.046 (0.012)	0.0003
All Native Americans combined	−0.042 (0.006)	1.77×10^{-11}	−0.021 (0.006)	7.15×10^{-5}	0.011 (0.003)	0.0001

Effect values are presented as effect size per C230 allele copy, standard error (SE). Linear regression was performed on the basis of log-transformed values for HDL-C levels (mg/dl), total cholesterol levels (mg/dl) and BMI (kg/m^2), adjusting for age, gender and diabetes status. HDL-C and total cholesterol levels were also adjusted for BMI.

[a]*P*-value adjusted by age, gender, birth year, diabetes status and family membership.

American native groups, biochemical data were available from two populations (Parkatejés and Kichwas), and although HDL-C levels were lower in *C230* carriers, the differences did not reach statistical significance. Altogether, the combined results of all Native American groups showed a highly significant effect of the *C230* allele (−4.2% per copy, $P = 1.77 \times 10^{-11}$). Interestingly, differences in the effect of *R230C* on lipid profiles were observed in some Native American populations. *R230C* was strongly associated with low total cholesterol and triglyceride levels in Pimas ($P = 1.8 \times 10^{-7}$ and $P = 7.0 \times 10^{-4}$, respectively) and Mayans ($P = 0.004$ and 0.010, respectively). Although the combined analysis in Native American groups also showed a significant association with low total cholesterol levels ($P = 7.15 \times 10^{-5}$), it was clearly not as significant as the association with low HDL-C levels ($P = 1.77 \times 10^{-11}$). The *C230* allele was associated with higher BMI in the combined analysis of Mexican native groups ($P = 0.012$), in Native South American populations ($P = 0.0003$ and 0.050 for Parkatejé and Kichwas, respectively), but not in Pima Indians. Altogether, combined

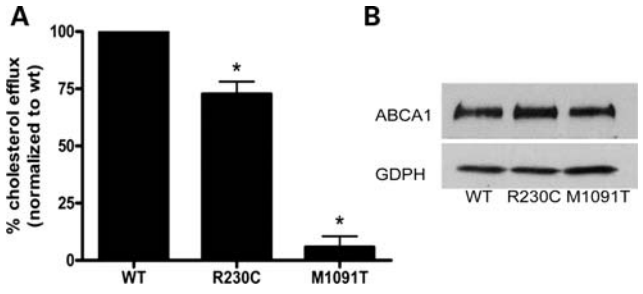

Figure 5. Functional characterization of the ABCA1/*R230C* variant by lipid efflux assay. (**A**) Polyclonal stable cell lines expressing the ABCA1 wild-type (WT), variant *R230C* and mutant *M1091T* (known defective in lipid efflux) were generated, and efflux activity for cholesterol was performed as described in Materials and Methods. Data represent mean ± SD of two to five experiments as a percent of the ApoA1-dependent efflux induced by wild-type ABCA1. Each assay was performed in triplicate. $^{*}P < 0.001$. (**B**) The expression of WT, variant *R230C* and mutant *M1091T* ABCA1 protein in Flp-In cells was assessed by western immunoblotting.

results showed that the *C230* allele was associated with higher BMI ($P = 0.0001$), and this association was evidently more significant in male than in female individuals ($P = 4.05 \times 10^{-6}$ versus $P = 0.02$).

Sequencing and *in vitro* functional analysis

To rule out the presence of another possible causal variant in LD with *C230*, all 50 exons and the promoter region of *ABCA1* were sequenced in a limited number of individuals (2 of each genotype); however, no promoter or coding variant in LD with *R230C* was found. Cholesterol efflux from Flp-In cell lines expressing the *ABCA1 C230* allele was significantly lower (27%) than that of cells expressing the wild-type *R230* allele ($P < 0.001$) (Fig. 5A), confirming that this variant has a functional effect *in vitro*. In contrast, the *C230* and *R230* cell lines showed no differences in phospholipid efflux. The *C230* allele expressed ABCA1 protein at levels comparable with that of *T1091* and wild-type alleles (Fig. 5B).

DISCUSSION

R230C, a private allele to the Americas

The *R230C* allele first identified in Oji-Crees and Mexican Mestizos was found in most Amerindian groups throughout the Americas, but not in any ethnic group from other continents (4,8). This is in agreement with previous studies that have not found this allele in 7717 Caucasian, Asian, African and South-Pacific Rim individuals (HapMap and SNP500CANCER databases) (8–11). This absence strongly suggests that *C230* is a private allele (exclusive to Native American and Native American-derived populations), although it may be present in some non-Amerindian populations not included in this analysis. Its presence on the same haplotype in both North and South Americans suggests that it may have arisen among Native American founders in Beringia or North-East Asia. This is in agreement with recent studies suggesting that founder populations stayed in Beringia long enough to give rise to exclusive genetic variants (12–14). Although private alleles in the HLA system and other genes have been previously reported in some Native American populations (15,16), there is only one previous report of a common private autosomal allele (microsatellite *D9S1120 9RA*) ubiquitous in the Americas (17). It is noteworthy that *D9S1120* and *R230C* (*ABCA1*) are both located on chromosome 9q, although separated by a 19 Mb distance. We genotyped *DS9S1120* in 16 *C230C* Native American and Mestizo homozygotes and found no allele in LD, indicating that the two ancestry informative markers are independent.

The *C230* allele distribution varied among different Amerindian populations (0–0.31). The complex demographic processes that these populations have gone through must have played a crucial role in this distribution. Initially, the moderate bottleneck in the out of Beringia process led to a relatively small effective population size, so genetic drift could have been one of the main causes of fluctuation (18,19). In groups that later expanded demographically to constitute societies formed by thousands of individuals such as Mesoamericans, genetic drift was less likely to cause differences in the distribution of *C230* frequencies.

Evidence suggesting R230C underwent positive selection

Understanding the impact of natural selection acting on particular genes in human populations can provide insights into the genetic etiology of human disease. Interestingly, *ABCA1* has been recognized as one of the genes most likely to have been subject to positive selection in humans since the divergence from the common ancestor of our lineage and that of chimpanzees (20,21). The results of the REHH and iHS analyses for the *ABCA1* gene region in Native Americans are not compatible with a simple neutral evolutionary model, but are consistent with the hypothesis that the *R230C* variant resides on a haplotype which is the target of an ongoing directional selective sweep. It must be acknowledged, however, that with the currently available genotyping data, it is not possible to define whether the *R230C* haplotype is also responsible for the signal resulting from the iHS test.

The geographical distribution of the *C230* allele clearly differs from the North-to-South gradient described for genome-wide neutral markers (22), suggesting the possibility of a climate-related adaptive process, as has been previously described for other genes involved in energy metabolism (23). In the context of Neel's hypothesis (24), *R230C* carriers could have had a selective advantage. Because the C230 ABCA1 protein shows decreased cholesterol efflux, the presence of this variant could favor intracellular cholesterol and energy storage. Specifically, adipose tissue benefits various biological functions including the ability to accommodate fluctuations in energy supply such as severe famine, the regulation of reproductive function and providing energy for the immune system now known to have a significant energy cost (25,26). However, under current westernized lifestyle changes, this allele may have become a major susceptibility allele for low HDL-C levels and other metabolic traits, which is consistent with the association of the *R230C* variant with higher BMI in Native American populations, and with obesity, T2D and metabolic syndrome in Mexican Mestizos (4,5). However, other environmental factors may also be involved in *C230* allele frequency distribution. For

instance, cholesterol plays an important role in various infectious processes such as the entry and replication of Dengue virus type 2 and flaviviral infection (27). The ABCA1 transporter is known to participate in infectious and/or thrombotic disorders involving vesiculation, since homozygous *ABCA1* gene deletions confer complete resistance against cerebral malaria in mice (28,29). Interestingly, areas with higher *C230* allele frequencies correspond to dengue, yellow fever and malaria distributions in the Americas (30). Altogether, these different lines of evidence suggest that the *ABCA1 C230* allele may have been important for survival throughout the colonization of the Americas.

Association of R230C with HDL-C levels and other metabolic traits

Overall, the prevalence of low HDL-C levels was not only higher in Native Americans than in European, Asian and African individuals (3), but also the most common dyslipidemia (65% in Mexican and South American natives). The Pima population is known to have much lower HDL-C and total cholesterol levels than US Caucasians (31). *R230C*/*C230C* genotypes were strongly associated with low HDL-C levels in Native American rural populations, Pimas and urban Mexican Mestizos (4,9). In fact, the sole presence of the *C230* allele explains ~4% of the HDL-C level variation in these populations, which is higher than the variation explained by any other SNP associated with HDL-C levels identified through genome-wide scans in Europeans and Indian Asians (9). This is consistent with both *in silico* (PANTHER subPSEC score −4.27) and *in vitro* evidence confirming that the *R230C* variant is functional (27% decrease in cholesterol efflux) (22). The functional effect is significant, but mild compared with the *T1091* allele previously identified in Tangier patients (32).

Environmental factors and further genetic variation (within *ABCA1* or other genes) may play a relevant role in the association of the *C230* allele with other metabolic traits. Lower total cholesterol and triglyceride levels were found in *C230* carriers only in Pimas and Mayans. In addition, the *C230* allele was associated with higher BMI in Mexican native groups, but not in the Pima Indians. Moreover, a gender effect was observed, as the association of *R230C* with higher BMI was more significant in males. Interestingly, in a previous study, a transcription factor 7-like 2 (*TCF7L2*) gene haplotype (HapA) with evidence of positive selection was also associated with higher BMI only in male individuals (33). Further studies are required to confirm the role of *R230C* in these metabolic and other fat storage-related traits such as non-alcoholic fatty liver disease, which is highly prevalent in Hispanic populations (34,35).

The *C230* allele has also been associated with T2D in the Mexican-Mestizo population (5). The overall frequency of T2D in most Mexican native groups was also high (11.3%); however, the study design was not appropriate for a case–control association in these groups. Interestingly, the *R230C* was only marginally associated with T2D in Pimas ($P = 0.06$) despite the large sample size and the previous finding that HDL-C concentrations in non-diabetic Pima Indian women were negatively associated with the development of T2D (36). Impaired ABCA1 function causes cholesterol accumulation in beta cells in animal models, suggesting that beneficial reductions in plasma lipids may limit the extent of beta cell damage and could partially mask glucose homeostasis disturbances (37). The highly significant association of *R230C* with reduced total cholesterol and triglyceride serum levels observed in Pimas may be one of the factors explaining this marginal association. The role of *R230C* as a risk allele for T2D in Mexican native groups and its interaction with environmental factors requires further analysis.

In conclusion, to the best of our knowledge, this is the first report of a common functional variant exclusive to Native American and descent populations associated with low HDL-C levels and other metabolic traits. We present several lines of evidence in favor of positive selection for the *R230C* allele possibly contributing to the adaptive evolution of Native American populations and providing insight into the genetic etiology of currently prevalent metabolic disease.

MATERIALS AND METHODS

Subjects

The study included a total of 4405 adult individuals from 36 different Native American groups and 863 Europeans and Asians. All Mexican and South American natives and their ancestors (two generations) were born in the same community and spoke their own native language. Field research was conducted by multidisciplinary teams.

Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the Ethics Committees of all participant institutions. Participants provided written informed consent. Local authorities gave their approval to participate in the study, and a translator was used as needed.

Anthropometric and biochemical analyses

Anthropometric and metabolic parameters were available for 2563 Pimas, 1016 Mexican and 157 South American native individuals (Supplementary Material, Table S4). The Pima subjects included are part of an ongoing longitudinal study of the etiology of T2D in the Gila River Indian community in Central Arizona (38). All biochemical measurements in 1050 Mexican natives and Kichwas (from Ecuador) were performed by the INCMNSZ with commercially available standardized methods as described by Villarreal-Molina *et al.* (4). Biochemical parameters of Parkatejé individuals have been previously described (39). T2D and HA were defined according to the American Diabetes Association and National Cholesterol Education Program (NCEP) criteria, respectively (40,41).

DNA sequencing of the *ABCA1* gene

Genomic DNA was extracted from peripheral blood leukocytes. The 50 exons and proximal promoter region of the

ABCA1 gene were amplified in samples from six individuals (two *R230R*, two *R230C* and two *C230C*) as described previously (8). Amplicons were sequenced using ABI PRISM BigDye Terminators version 3.1 on an ABI 3100 automated sequencer according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA).

SNP genotyping

The *R230C* variant and 23 SNPs spanning an 800 kb region were genotyped using TaqMan assays (ABI Prism 7900HT Sequence Detection System; Applied Biosystems). The 23 SNPs were selected from the HGDP for being informative in Native American populations and were genotyped in 10 Kichwa, 7 Nahua and 3 Zapotec trios (mother–father–offspring). The names and chromosomal position of all SNPs analyzed are given in Supplementary Material, Table S5. Genotyping call rate exceeded 95% per SNP, and no discordant genotypes were observed in 40 duplicate samples. Deviation from Hardy–Weinberg equilibrium was not observed for any SNP.

Generation of R230C variant constructs and cell lines

Polyclonal stable cell lines expressing the *ABCA1 R230C* variant were generated using the Flp-In system (Invitrogen, Carlsbad, CA, USA) as described previously (42). The generation and detailed biochemical characterization of many of these cell lines are described elsewhere (43). Briefly, the *R230C* variant was generated by PCR-based site-directed mutagenesis using the primers 230F,5′-GAGCGAGTACTT TGTTCCAACATG and 230R,5′-CATGTTGGAACAAAGT ACTCGCTC and cloned into pcDNA5/FRT (Invitrogen). The plasmid was completely sequenced prior to transfection into 293 Flp-In cells. The *M1091T ABCA1* mutation previously identified in Tangier patients was used as control (32).

Cholesterol and phospholipid efflux assay

Efflux experiments were performed as described previously (42). Briefly, cells were loaded with 1 μCi of [^{3}H] cholesterol or 2 μCi of [^{3}H]-choline (Amersham Biosciences, Little Chalfont, Buckingham, UK) for 24 h. The following day, the medium was removed and replaced with serum-free medium containing 5 mg/ml delipidated bovine serum albumin (Sigma, St Louis, MO, USA). After 1 h of incubation, 20 μg/ml human ApoA1 (Athens Research and Technology, Athens, GA, USA) was added for 4 h. For cholesterol efflux, the medium was collected and cells were lysed in 0.1 N NaOH/0.1% SDS. For phospholipid efflux, the [^{3}H]-choline/phospholipids/ApoA1 in the medium was collected by immunoprecipitation with ApoA1 antibody and cells were digested for protein assay. The radioactivity in the samples was quantified by scintillation counting. Cholesterol efflux is expressed as a percent of counts in medium over total (medium + cells). Phospholipid efflux is expressed as counts of the immunocollected [^{3}H]-choline/phospholipid/ApoA1 normalized to cell protein. Data are expressed as percent of the ApoA1-dependent efflux induced by wild-type *ABCA1*. Significance was calculated using a one-way ANOVA test with a Newman–Keuls post-test using GraphPad Prism 4 software (San Diego, CA, USA). ABCA1 expression was determined by western blotting as described previously (44), using anti-ABCA1 or anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibodies (Chemicon, Temecula, CA, USA).

Statistical analyses

Population genetics. Allele and genotype frequencies, Hardy–Weinberg equilibrium and AMOVA were calculated using Arlequin 3.11 software (45). The linguistic classification of Native American languages was adopted from Campbell (46). Network analyses of haplotypes from HapMap, Mexican-Mestizo and Native American population data were performed using a median-joining and maximum-parsimony method (Network 4.510 software) (47).

Positive selection tests. The LRH test was applied to examine the decay of LD (Sweep software) (48) within an 800 kb region flanking *R230C* using data obtained from the 20 Native American trios described earlier. LD decay was then compared with LRH data generated elsewhere covering a total of 24 Mb of the genome in Native American populations (49). To further explore the presence of positive selection signatures in the *ABCA1* region, the iHS was estimated as described previously (50,51), using publicly available genotype data for ~650 000 SNPs genome-wide distributed in five Native American groups from the HGDP (52). With an approach similar to that described by Nielsen *et al*. (53) to detect regions with aberrant allele frequency spectra (test 1), we applied a composite likelihood test to detect regions with aberrant 'iHS spectra'. We first categorized |iHS| in bins of size 0.1 and then estimated the probability of observing an SNP in each bin, both in the whole-genome data set (background distribution) and in each 31-SNP sliding window over the whole genome. Two composite likelihoods were estimated for each window, multiplying the probability of observing each SNP in the window, by either the probability of observing the SNP in the genome-wide background or that estimated from the window. A log-likelihood ratio was then estimated comparing both likelihoods, where extreme values indicate unusual iHS patterns compared with the rest of the genome.

Associations with HDL-C and metabolic traits. Associations of *R230C* genotypes with HDL-C and other metabolic traits were tested using linear regression models (assuming an additive model) adjusting for covariates including age, sex and BMI (SPSS, version 15.0, statistical package; Chicago, IL, USA). All variables tested were log-transformed for the analysis. Combined association tests were conducted using a Mantel–Haenszel-like model (54). The combined estimated effect was computed as a weighted average of the individual estimated effects using weights proportional to the inverse of the standard errors squared (33).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the volunteers from all native populations of North and South America for their participation; Luz E. Guillén-Pineda for her technical assistance; Gastón Macín for his participation in sample collection in Nahua and Totonaca communities; and the staff of the Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH, for conducting the examinations of the Pima Indian participants.

Conflict of Interest statement: None declared.

FUNDING

This research was supported by grant 69856 from the Consejo Nacional de Ciencia y Tecnología (CONACYT) México, and partly supported by grant 660 from the Fundación Mexicana para la Salud-Silanes; by the Intramural Research Program of the National Institute of Digestive and Kidney Diseases (NIDDK), NIH; and by a grant of the Canadian Institutes of Health Research (CIHR). V.A.-A., T.F.-D., M.V.-C. and L.J.-A. were supported by fellowships from CONACYT, México. J.K.K. was supported by fellowships of the Michael Smith Foundation for Health Research (MSFHR) and the CIHR. M.R.H. holds a Canada Research Chair in Human Genetics.

WEB RESOURCES

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

HapMap database, http://www.hapmap.org/

SNP500CANCER, http://snp500cancer.nci.nih.gov

HGDP-CEPH database, http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP

Arlequin 3.11 software, http://cmpg.unibe.ch/software/arlequin3/

Network 4.510 software, http://www.fluxus-engineering.com

Sweep software, http://www.broad.mit.edu/mpg/sweep/index.html

REFERENCES

1. Lorenzo, C., Serrano-Rios, M., Martinez-Larrad, M.T., Gabriel, R., Williams, K., Gonzalez-Villalpando, C., Stern, M.P., Hazuda, H.P. and Haffner, S.M. (2001) Was the historic contribution of Spain to the Mexican gene pool partially responsible for the higher prevalence of type 2 diabetes in Mexican-origin populations? The Spanish Insulin Resistance Study Group, the San Antonio Heart Study, and the Mexico City Diabetes Study. *Diabetes Care*, **24**, 2059–2064.
2. Cossrow, N. and Falkner, B. (2004) Race/ethnic issues in obesity and obesity-related comorbidities. *J. Clin. Endocrinol. Metab.*, **89**, 2590–2594.
3. Aguilar-Salinas, C.A., Canizales-Quinteros, S., Rojas-Martínez, R., Mehta, R., Villarreal-Molina, M.T., Arellano-Campos, O., Riba, L., Gómez-Pérez, F.J. and Tusié-Luna, M.T. (2009) Hypoalphalipoproteinemia in the populations with Native American origin: an opportunity to assess the interaction of genes and environment. *Curr. Opin. Lipidol.*, **20**, 92–97.
4. Villarreal-Molina, M.T., Aguilar-Salinas, C.A., Rodriguez-Cruz, M., Riaño, D., Villalobos-Comparan, M., Coral-Vazquez, R., Menjivar, M., Yescas-Gomez, P., Königsoerg-Fainstein, M., Romero-Hidalgo, S. *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.
5. Villarreal-Molina, M.T., Flores-Dorantes, M.T., Arellano-Campos, O., Villalobos-Comparan, M., Rodríguez-Cruz, M., Miliar-García, A., Huertas-Vazquez, A., Menjivar, M., Romero-Hidalgo, S., Wacher, N.H. *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.
6. Attie, A.D., Kastelein, J.P. and Hayden, M.R. (2001) Pivotal role of ABCA1 in reverse cholesterol transport influencing HDL levels and susceptibility to atherosclerosis. *J. Lipid Res.*, **42**, 1717–1726.
7. Oram, J.F. and Vaughan, A.M. (2006) ATP-binding cassette cholesterol transporters and cardiovascular disease. *Circ. Res.*, **99**, 1031–1043.
8. Wang, J., Burnett, J.R., Near, S., Young, K., Zinman, B., Hanley, A.J., Connelly, P.W., Harris, S.B. and Hegele, R.A. (2000) Common and rare ABCA1 variants affecting plasma HDL cholesterol. *Arterioscler. Thromb. Vasc. Biol.*, **20**, 1983–1989.
9. Kooner, J.S., Chambers, J.C., Aguilar-Salinas, C.A., Hinds, D.A., Hyde, C.L., Warnes, G.R., Gómez-Pérez, F.J., Frazer, K.A., Elliott, P., Scott, J. *et al.* (2008) Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.*, **4**, 149–151.
10. Hu, S., Zhong, Y., Hao, Y., Luo, M., Zhou, Y., Guo, H., Liao, W., Wan, D., Wei, H., Gao, Y. *et al.* (2009) Novel rare alleles of ABCA1 are exclusively associated with extreme high-density lipoprotein-cholesterol levels among the Han Chinese. *Clin. Chem. Lab. Med.*, **47**, 1239–1245.
11. Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B. and Tybjaerg-Hansen, A. (2004) Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J. Clin. Invest.*, **114**, 1343–1353.
12. González-José, R., Bortolini, M.C., Santos, F.R. and Bonatto, S.L. (2008) The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am. J. Phys. Anthropol.*, **137**, 175–187.
13. Kitchen, A., Miyamoto, M.M. and Mulligan, C.J. (2008) A three-stage colonization model for the peopling of the Americas. *PLoS ONE*, **3**, e1596.
14. Fagundes, N.J.R., Kanitz, R. and Bonatto, S.L. (2008) A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS ONE*, **3**, e3157.
15. Williams, R.C. and McAuley, J.E. (1992) HLA class I variation controlled for genetic admixture in the Gila River Indian community of Arizona: a model for the Paleo-Indians. *Hum. Immunol.*, **33**, 39–46.
16. Salzano, F.M. (2002) Molecular variability in Amerindians: widespread but uneven information. *An. Acad. Bras. Cienc.*, **74**, 223–263.
17. Schroeder, K.B., Jakobsson, M., Crawford, M.H., Schurr, T.G., Boca, S.M., Conrad, D.F., Tito, R.Y., Osipova, L.P., Tarskaia, L.A., Zhadanov, S.I. *et al.* (2009) Haplotypic background of a private allele at high frequency in the Americas. *Mol. Biol. Evol.*, **26**, 995–1016.
18. Fagundes, N.J.R., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L. and Excoffier, L. (2007) Statistical evaluation of alternative models of human evolution. *Proc. Natl Acad. Sci. USA*, **104**, 17614–17619.
19. Fagundes, N.J.R., Kanitz, R., Eckert, R., Valls, A.C., Bogo, M.R., Salzano, F.M., Smith, D.G., Silva, W.A. Jr, Zago, M.A., Ribeiro-dos-Santos, A.K. *et al.* (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.*, **8**, 583–592.
20. Brunham, L.R., Singaraja, R.R., Pape, T.D., Kejariwal, A., Thomas, P.D. and Hayden, M.R. (2005) Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet.*, **1**, e83.
21. Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B. *et al.* (2003) Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science*, **302**, 1960–1963.
22. Wang, S., Lewis, C.M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C. *et al.* (2007) Genetic variation and population structure in native Americans. *PLoS Genet.*, **3**, e185.
23. Hancock, A.M., Witonsky, D.B., Gordon, A.S., Eshel, G., Pritchard, J.K., Coop, G. and Di Rienzo, A. (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.*, **4**, e32.

24. Neel, J.V. (1962) Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am. J. Hum. Genet.*, **14**, 353–362.
25. Prentice, A.M., Rayco-Solon, P. and Moore, S.E. (2005) Insights from the developing world: thrifty genotypes and thrifty phenotypes. *Proc. Nutr. Soc.*, **64**, 153–161.
26. Wells, J.C. (2009) Ethnic variability in adiposity and cardiovascular risk: the variable disease selection hypothesis. *Int. J. Epidemiol.*, **38**, 63–71.
27. Lee, C.J., Lin, H.R., Liao, C.L. and Lin, Y.L. (2008) Cholesterol effectively blocks entry of flavivirus. *J. Virol.*, **82**, 6470–6480.
28. Simons, K. and Toomre, D. (2000) Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell. Biol.*, **1**, 31–39.
29. Combes, V., Coltel, N., Alibert, M., van Eck, M., Raymond, C., Juhan-Vague, I., Grau, G.E. and Chimini, G. (2005) ABCA1 gene deletion protects against cerebral malaria: potential pathogenic role of microparticles in neuropathology. *Am. J. Pathol.*, **166**, 295–302.
30. Rogers, D.J. and Randolph, S.E. (2000) The global spread of malaria in a future, warmer world. *Science.*, **289**, 1763–1766.
31. Howard, B.V., Davis, M.P., Pettitt, D.J., Knowler, W.C. and Bennett, P.H. (1983) Plasma and lipoprotein cholesterol and triglyceride concentrations in the Pima Indians: distributions differing from those of Caucasians. *Circulation.*, **68**, 714–724.
32. Clee, S.M., Kastelein, J.J., van Dam, M., Marcil, M., Roomp, K., Zwarts, K.Y., Collins, J.A., Roelants, R., Tamasawa, N., Stulc, T. *et al.* (2000) Age and residual cholesterol efflux affect HDL cholesterol levels and coronary artery disease in ABCA1 heterozygotes. *J. Clin. Invest.*, **106**, 1263–1270.
33. Helgason, A., Pálsson, S., Thorleifsson, G., Grant, S.F., Emilsson, V., Gunnarsdottir, S., Adeyemo, A., Chen, Y., Chen, G., Reynisdottir, I. *et al.* (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.*, **39**, 218–225.
34. Browning, J.D., Szczepaniak, L.S., Dobbins, R., Nuremberg, P., Horton, J.D., Cohen, J.C., Grundy, S.M. and Hobbs, H.H. (2004) Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology*, **40**, 1387–1395.
35. Guerrero, R., Vega, G.L., Grundy, S.M. and Browning, J.D. (2009) Ethic differences in hepatic steatosis: an insulin resistance paradox? *Hepatology*, **49**, 791–801.
36. Fagot-Campagna, A., Narayan, K.M., Hanson, R.L., Imperatore, G., Howard, B.V., Nelson, R.G., Pettitt, D.J. and Knowler, W.C. (1997) Plasma lipoproteins and incidence of non-insulin-dependent diabetes mellitus in Pima Indians: protective effect of HDL cholesterol in women. *Atherosclerosis*, **128**, 113–119.
37. Brunham, L.R., Kruit, J.K., Verchere, C.B. and Hayden, M.R. (2008) Cholesterol in islet dysfunction and type 2 diabetes. *J. Clin. Invest.*, **118**, 403–408.
38. Knowler, W.C., Bennett, P.H., Hamman, R.F. and Miller, M. (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am. J. Epidemiol.*, **108**, 497–505.
39. Vieira-Filho, J.P., Reis, A.F., Kasamatsu, T.S., Tavares, E.F., Franco, L.J., Matioli, S.R. and Moisés, R.S. (2004) Influence of the polymorphisms Tpr64Arg in the beta 3-adrenergic receptor gene and Pro12Ala in the PPAR gamma 2 gene on metabolic syndrome-related phenotypes in an indigenous population of the Brazilian Amazon. *Diabetes Care*, **27**, 621–622.
40. The Expert Committee on the Diagnosis Classification of Diabetes Mellitus (2003) Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, **26**, S5–S20.
41. Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) (2001) Executive summary of the Third Report of the National Cholesterol Education Program (NCEP). *JAMA*, **285**, 2486–2497.
42. See, R.H., Caday-Malcolm, R.A., Singaraja, R.R., Zhou, S., Silverston, A., Huber, M.T., Moran, J., James, E.R., Janoo, R., Savill, J.M. *et al.* (2002) Protein kinase A site-specific phosphorylation regulates ATP-binding cassette A1 (ABCA1)-mediated phospholipid efflux. *J. Biol. Chem.*, **277**, 41835–41842.
43. Singaraja, R.R., Van Eck, M., Bissada, N., Zimetti, F., Collins, H.L., Brunham, L.R., Kang, M.H., Zannis, V.I., Chimini, G. and Hayden, M.R. (2006) Specific mutations in ABCA1 have discrete effects on ABCA1 function and lipid phenotypes both *in vivo* and *in vitro*. *Circ. Res.*, **99**, 389–397.
44. Wellington, C.L. (2002) ABCA1 mRNA and protein distribution patterns predict multiple different roles and levels of regulation. *Lab. Invest.*, **82**, 273–283.
45. Excoffier, L., Laval, G. and Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, **1**, 47–50.
46. Campbell, L. (1997) *American Indian languages: the historical linguistics of Native America*. Oxford University Press, New York.
47. Bandelt, H.J., Forster, P. and Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37–48.
48. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.
49. Moreno-Estrada, A., Tang, K., Sikora, M., Marquès-Bonet, T., Casals, F., Navarro, A., Calafell, F., Bertranpetit, J., Stoneking, M. and Bosch, E. (2009) Interrogating 11 fast-evolving genes for signatures of recent positive selection in worldwide human populations. *Mol. Biol. Evol.*, **26**, 2285–2297.
50. Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol.*, **4**, e72.
51. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W. *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.*, **19**, 826–837.
52. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **22**, 1100–1104.
53. Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. and Bustamante, C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res.*, **15**, 1566–1575.
54. Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl Cancer Inst.*, **22**, 719–748.