Native American Ancestry Is Associated With Severe Diabetic Retinopathy in Latinos

Xiaoyi Gao,¹ W. James Gauderman,² Paul Marjoram,² Mina Torres,³ Yii-Der I. Chen,⁴ Kent D. Taylor,⁴ Jerome I. Rotter,⁴ and Rohit Varma³

¹Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, Illinois, United States

²Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States

³USC Eye Institute, Department of Ophthalmology, University of Southern California, Los Angeles, California, United States

⁴Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA, Torrance, California, United States

Correspondence: Xiaoyi Gao, Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, 1905 West Taylor Street, Suite 235, Chicago, IL 60612, USA; rgao@uic.edu.

Rohit Varma, USC Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, 1450 San Pablo Street, Suite 4900, Los Angeles, CA 90033, USA; rvarma@usc.edu.

Submitted: June 17, 2014

Accepted: August 12, 2014

Citation: Gao X, Gauderman WJ, Marjoram P, et al. Native American ancestry is associated with severe diabetic retinopathy in Latinos. *Invest Ophthalmol Vis Sci*. 2014;55:6041–6045. DOI:10.1167/iovs.14-15044

PURPOSE. Diabetic retinopathy (DR) is a leading cause of blindness in working age adults. Studies have observed that Latinos have a higher prevalence of DR than whites. The purpose of this study is to test the association between genetic admixture and severe DR in Latinos with type 2 diabetes mellitus (T2DM).

METHODS. We conducted a case-control study using 944 T2DM subjects from the Los Angeles Latino Eye Study. Cases ($n = 135$) were defined as proliferative or severe nonproliferative DR subjects. Controls ($n = 809$) were other diabetic subjects in the cohort. Genotyping was performed on the Illumina OmniExpress BeadChip. We estimated genetic ancestry in Latinos using STRUCTURE with the HapMap reference panels. Univariate and multivariate logistic regression analyses were used to test the relationship between the proportions of genetic ancestry and severe DR.

RESULTS. Native American ancestry (NAA) in Latino T2DM subjects is associated significantly with severe DR ($P = 0.002$). The association remained significant ($P = 0.005$) after adjusting for age, sex, duration of diabetes, hemoglobin A1c, body mass index, systolic blood pressure, education, and income. We also validated the NAA estimates in Latinos using ADMIXTURE with the 1000 Genomes Project reference panels and obtained consistent results.

CONCLUSIONS. Our results demonstrate for the first time to our knowledge that NAA is a significant risk factor for severe DR in Latinos.

Keywords: diabetic retinopathy, Latinos, genetic ancestry, Native American

Diabetic retinopathy (DR) is a leading cause of new cases of blindness in working age adults in the United States, as well as worldwide, and is the most common form of diabetic eye disease.^{1,2} Diabetic retinopathy is a complication of diabetes mellitus characterized by the alteration of small blood vessels in the retina.³ There are several stages of DR, with proliferative diabetic retinopathy (PDR) being the most advanced stage of the disease. If left untreated, it can cause severe vision loss and even blindness. According to the Centers for Disease Control and Prevention, an estimated 4.2 million Americans have DR in the United States, of whom 655,000 are at risk for severe vision loss (available in the public domain at <http://www.cdc.gov/visionhealth/>).

While obesity, diet, physical activity, and genetics all are known to influence diabetes mellitus, the reason that some individuals are susceptible to DR while others are not remains unknown. It also has been reported that Latinos have a higher prevalence of DR than non-Hispanic Whites and African Americans.^{4–8} Several risk factors have been reported for DR in Latinos, such as older age, male sex, longer duration of diabetes (DOD), poor glycemic control, and high blood pressure.^{4,8,9} Reported factors for PDR in Latinos include longer DOD, being on insulin treatment, and higher systolic blood pressure.⁹

Latinos typically are a three-way admixture of Native American (NAA), European, and African ancestry.^{10,11} The proportions of these ancestries vary substantially across individual Latinos.¹² Given the observed differences in the prevalence of DR in different ethnic groups and the strong public health need for the prevention of severe DR, determining whether or not genetic ancestry is associated with severe DR in Latinos may provide additional insight into these ethnic disparities. The purpose of this study is to examine the relationship between the proportions of genetic ancestry and severe DR in Latinos with type 2 diabetes mellitus (T2DM) using data collected from the Los Angeles Latino Eye Study (LALES),¹³ the largest population-based study of ophthalmic disease in Latinos. To our knowledge, this is the first study to investigate the association between genetic ancestry and severe DR in Latinos.

MATERIALS AND METHODS

Ethics Statement

This research was approved by the University of Southern California Health Sciences Campus, the University of Illinois at

Chicago, and Los Angeles Biomedical Research Institute at Harbor-University of California, Los Angeles (UCLA). All clinical investigation was approved by the institutional review board/ethics committee at the University of Southern California and conducted according to the principles expressed in the Declaration of Helsinki.

Diabetic Retinopathy Grading and Study Subjects

All subjects underwent detailed ophthalmologic examinations. Retinopathy levels were classified by the Early Treatment Diabetic Retinopathy Study (ETDRS) grading scale¹⁴ in LALES.¹⁵ The eye with the worse score was used to define the retinopathy level for each subject. We conducted this research using 944 T2DM subjects from LALES. Cases included proliferative or severe nonproliferative DR subjects (ETDRS level 47–85,¹⁵ $n = 135$). Controls included all other diabetic subjects in the cohort ($n = 809$). Details of the DR distribution plot are shown in Supplementary Figure S1. All subjects were 40 years of age and older. Written, informed consent was obtained from all participants.

Genotyping and Quality Control

We initially genotyped 1070 diabetic subjects from the larger LALES cohort using the Illumina OmniExpress BeadChip (730,525 markers; Illumina, Inc., San Diego, CA, USA). We also included 59 duplicate samples to verify genotyping reproducibility. The genotyping was performed at the Genotyping Laboratory of the Institute for Translational Genomics and Population Sciences at the LA-Biomed Research Institute at Harbor-UCLA. Single nucleotide polymorphisms (SNPs) were called using the Illumina GenomeStudio (v2011.1) software. Individuals were excluded if genotyping call rates were less than 97%. The average call rate was greater than 99.32%. The reproducibility was greater than 99.99%. We used the software PLINK (v1.07; available in the public domain at <http://pngu.mgh.harvard.edu/~purcell/plink>)¹⁶ to perform quality control. Individuals with inconsistency between reported sex and genetic sex, unexpected duplicates, with missing DR data, and related individuals were removed. Markers were excluded if minor allele frequency (MAF) < 0.01, call rates < 95%, or if Hardy-Weinberg equilibrium P values < 10^{-6} . The SNPs were coded on the forward strand. After quality control (QC), 809 controls and 135 cases remained and were used in the downstream analysis. For the 944 subjects, the average call rate was greater than 99.35% and 585,765 SNPs remained.

Genetic Ancestry Estimation

Genetic ancestry was assessed using the program STRUCTURE (available in the public domain at <http://pritchardlab.stanford.edu/structure.html>), a Bayesian clustering approach using a Markov Chain Monte Carlo method.^{17,18} To make inferences based on reference populations of known ancestry, we included all the unrelated Northern Europeans (CEU, $n = 60$) and West Africans (YRI, $n = 60$) from the HapMap project¹⁹ and Native Americans ($n = 105$).¹⁰ After merging data sets, we randomly selected 5000 autosomal SNPs for estimating genetic ancestry. Studies have shown that randomly selected SNPs are sufficient to differentiate populations of different ancestry.²⁰ For STRUCTURE analysis, 10,000 burn-in and 10,000 iterations for estimation were used. We also specified the prior information of three ancestral populations in STRUCTURE. The STRUCTURE output of interest included the estimated proportions of European, Native American, and African ancestry, the sum of which equals 1, for each study subject. As a validation of the genetic ancestry estimates, we also ran

the ADMIXTURE²¹ program, a model-based estimation of ancestry that can use a large number of genetic markers, with reference panels (CEU, $n = 87$; YRI, $n = 88$) from the 1000 Genomes Project²² and Native Americans ($n = 105$).¹⁰

Statistical Analysis

We tested the association of several variables; that is, age, sex, glycosylated hemoglobin levels (HbA1c), body mass index (BMI), systolic blood pressure (SBP), DOD, education, income, proportion NAA, and proportion African ancestry (AA), with severe DR. These variables were chosen because they were used in previous investigations of DR (i.e., age, sex, HbA1c, BMI, SBP, and DOD) or because they may be confounding factors in genetic ancestry studies of Latinos (i.e., socioeconomic factors). The proportion of European ancestry was not included because it is equal to one minus the proportions of NAA and AA. Associations were conducted using logistic regression, in univariate and multivariate analyses. Backward logistic regression was used for selecting candidate risk factors. We also performed permutation tests to validate the association of genetic ancestry with severe DR. We randomly shuffled the phenotype and repeated the logistic regression 1000 times. We then recorded the number of times a permutation P value was less than the observed P value. Risk factors with $P \leq 0.05$ were declared significant. Statistical analyses were performed using SAS v9.2 (SAS Institute, Inc., Cary, NC, USA). R software²³ was used for graphing.

RESULTS

Figure 1 plots individual genetic ancestry estimates, and the relationship between the subjects' proportion of NAA and the presence of severe DR. Each individual is represented by a vertical bar on the x -axis, which is composed of three colored segments corresponding to the proportions of genetic ancestry from three ancestry populations. Native American, African, and European ancestries are colored dark yellow, blue, and purple, respectively. The plot is sorted by the proportion of NAA to aid visualization. Ticks on the x -axis are severe DR subjects. It is clear that as the proportion of NAA increases in subjects (as seen on the right hand side of the plot), the density of tick marks increases. This suggests that individuals with a higher proportion of NAA ancestry are at a higher risk for having severe DR than individuals with lower proportions of NAA ancestry (as seen on the left hand side of the plot).

Table 1 shows the characteristics of the study sample and the univariate associations of various candidate risk factors; that is, age, sex, HbA1c, BMI, SBP, DOD, education, income, NAA, and AA. In the univariate analysis, HbA1c ($P = 0.002$), SBP ($P < 0.0001$), DOD ($P < 0.0001$), and NAA ($P = 0.002$) are significantly associated with severe DR at the 0.05 significance level. Age, sex, BMI, education, income, and AA did not have significant associations with severe DR. Compared to controls, the cases had longer DOD (e.g., 58.5% vs. 26.7% for the ≥ 15 years category), higher SBP (144 vs. 135), and higher NAA (49.6% vs. 45.3%).

Table 2 shows the multivariate logistic regression results. Model 1 shows the results for the full model, which includes all the candidate risk factors in the logistic regression. When age, sex, DOD, HbA1c, BMI, SBP, education, income, NAA, and AA are all in the logistic regression model, only age, DOD, HbA1c, SBP, and NAA are significantly associated with severe DR and have P values of 0.002, <0.0001, 0.016, <0.0001, and 0.016, respectively. Model 2 presents the results of the backward logistic regression analysis. The significance level for a variable staying in the model was 0.05. Education, AA, BMI, sex, and

FIGURE 1. Individual genetic ancestry estimates. Each individual is represented by a *vertical bar*, which is composed of three *colored segments* corresponding to the proportions of genetic ancestry from three ancestry populations. Native American, African, and European ancestries are colored *dark yellow*, *blue*, and *purple*, respectively. The *x*-axis and *y*-axis denote subjects and proportions of genetic ancestry, respectively. Subjects are sorted by their NAA to aid visualization. *Ticks* on the *x*-axis are severe DR subjects.

income were removed by the backward elimination, and only age ($P = 0.006$), DOD ($P < 0.0001$), HbA1c ($P = 0.01$), SBP ($P < 0.0001$), and NAA ($P = 0.005$) remained. Compared to the full model, the odds ratio (OR) and 95% confidence interval (CI) in the reduced model for age, DOD, HbA1c, SBP, and NAA showed similar effect size and 95% CI. Moreover, the reduced model with only the effects of DOD, HbA1c, SBP, and NAA had smaller Akaike Information Criterion and Bayes Information Criterion, which indicated better fit. To make the interpretation for NAA easier, we dichotomized NAA at 50% (the mean of NAA in

cases) and categorized our data into subjects with higher and lower NAA. Model 3 shows the multivariate logistic regression using the dichotomized NAA along with age, DOD, HbA1c, and SBP. We see that age, DOD, HbA1c, and SBP remain significant, and showed similar effect size and 95% CI to model 2. The dichotomized NAA also is significant ($P = 0.002$) with OR = 1.87 (95% CI, 1.26, 2.78), which means that the risk of having severe DR in Latino subjects with higher NAA ($\geq 50\%$) is approximately 1.87 times that in those with lower NAA. The empirical P value for NAA from permutation tests was 0.002; that is, of 1000 permutation tests, only 2 had P values less than the observed P value, 0.002.

As a validation of the NAA estimates, we also ran ADMIXTURE,²¹ a model-based method for estimating genetic ancestry that can use a large number of SNPs, with the reference panels from the 1000 Genomes Project. After merging with our data sets, 74,307 SNPs remained and were used in ADMIXTURE for inferring genetic ancestry. Figure 2 shows the pairwise plot of the ADMIXTURE and STRUCTURE NAA estimates. In general, we see that the STRUCTURE and ADMIXTURE NAA estimates agree with each other very well with a correlation coefficient greater than 0.99. The association between ADMIXTURE NAA estimates and severe DR remained significant (data not shown), which demonstrates the robustness of our results.

DISCUSSION

To our knowledge, this is the first study to investigate the contribution of genetic ancestry on DR in Latinos. In this study, we used data from LALES, the largest epidemiologic eye study in Latinos ($n = 6357$). The Latinos used in the study were collected in the city of La Puente, Los Angeles County, CA, USA, and 95% of them were of Mexican origin.¹³ We found a significant association between NAA and severe DR in Latinos with T2DM. Specifically, NAA was associated positively with severe DR in Latinos, even after adjusting for known risk factors of DR. This suggested that at least a partial explanation for the observation that Latinos have a higher prevalence of severe DR than in non-Hispanic Whites and African Americans is due to genes of Native American origin. We validated our

TABLE 1. Characteristics of the Study Sample and Univariate Association Results

	Controls, <i>n</i> = 809	Cases, <i>n</i> = 135	<i>P</i> Value
Age, y	63.1 (9.8)	62.4 (9.4)	0.44
Sex, male	42.4%	43.7%	0.77
HbA1c, %	7.8 (1.8)	8.3 (2.0)	0.002
BMI, kg/m ²	31.9 (6.0)	31.2 (5.9)	0.17
SBP, mm Hg	135 (20.6)	144 (23.0)	<0.0001
Duration of diabetes, y			
<5	32.6%	5.2%	<0.0001
5–9	21.3%	12.6%	
10–14	19.4%	23.7%	
≥15	26.7%	58.5%	
Education level, y			
≤6	47.7%	49.3%	0.8
7–11	21.9%	23.1%	
≥12	30.4%	27.6%	
Income level*			
<\$20,000	52.8%	57.0%	0.22
\$20,000–\$40,000	33.4%	35.1%	
>\$40,000	13.8%	7.9%	
NA ancestry, %	45.3 (14.8)	49.6 (14.8)	0.002
African ancestry, %	2.9 (3.1)	3.0 (2.4)	0.71

All statistics shown for controls and cases are mean (SD) except for sex, duration of diabetes, education level, and income level which are shown as percentages.

* Data missing for 126 subjects.

TABLE 2. Multivariate Association Results

	Model 1*		Model 2		Model 3†	
	OR (95% CI)	P Value	OR (95% CI)	P Value	OR (95% CI)	P Value
Age	0.96 (0.93, 0.99)	0.002	0.97 (0.95, 0.99)	0.006	0.96 (0.94, 0.99)	0.004
Sex		NS				
Duration of diabetes	1.07 (1.05, 1.10)	<0.0001	1.06 (1.05, 1.08)	<0.0001	1.06 (1.05, 1.08)	<0.0001
HbA1c	1.16 (1.03, 1.30)	0.016	1.15 (1.03, 1.27)	0.01	1.15 (1.04, 1.27)	0.009
BMI		NS				
SBP	1.03 (1.02, 1.04)	<0.0001	1.02 (1.01, 1.03)	<0.0001	1.02 (1.01, 1.03)	<0.0001
Education		NS				
Income		NS				
NAA	1.02 (1.00, 1.03)	0.016	1.02 (1.01, 1.03)	0.005	1.87 (1.26, 2.78)	0.002
African ancestry		NS				

NS, not significant.

* Income data missing for 126 subjects.

† The NAA is dichotomized at 50%.

NAA estimates using two independent state-of-the-art software tools.

There have been considerable challenges in analyzing the genetics of DR.^{24,25} At present, there is no consensus on the best control and case groups for studying DR. For example, Fu et al.²⁶ defined controls as subjects who had ETDRS grade 10 to 37 (normal to early nonproliferative DR) and cases as subjects with moderate-to-severe nonproliferative DR and PDR. Sheu et al.²⁷ defined controls as subjects with ≥ 8 years of duration without DR, and Grassi et al.^{28,29} defined controls as all non-PDR diabetic subjects in his cohorts. To increase sample sizes for cases and controls, we defined cases as PDR or severe nonproliferative DR subjects ($n = 135$), and used all the other diabetic subjects as controls ($n = 809$). It is possible that some controls may become cases with longer duration of diabetes. However, the estimates of association should be biased toward the null hypothesis.

FIGURE 2. Pairwise plots of the NAA estimates from STRUCTURE and ADMIXTURE. The x-axis and y-axis denote STRUCTURE NAA and ADMIXTURE NAA estimates, respectively. If the two estimates for the same subjects match with each other perfectly, they would fall on the diagonal line. We see that the STRUCTURE and ADMIXTURE NAA estimates agree with each other very well with a correlation coefficient greater than 0.99.

There are some limitations inherent in our study. The individual ancestry estimates were derived by statistical procedures and may contain errors. However, the errors are likely to be random and unbiased with respect to the phenotype analyzed. We did not include diet or environmental exposures in our analysis. The sample size for severe DR cases is not very large, which is likely to be the limitation of any single cohort study. We emphasize the need to replicate our results in an independent Latino cohort. Through STRUCTURE and ADMIXTURE, we obtained global ancestry of our study subjects. A more detailed lookup of each individual's ancestry can be obtained from local ancestry estimation.³⁰ However, Latinos are an admixed population and typically are viewed as a three-way admixture of European, Native American, and African, which has imposed challenges in the statistical inference of accurate local ancestry. Moreover, the accuracy of local ancestry depends on the accuracy of haplotypes in the reference panel. Considerable efforts have been made to create public genetic resources for subjects of European and African ancestries; for example, the 1000 Genomes Project.³¹ At present, public resources for Native American haplotypes still are very limited. With the increase in the genetic research in minority populations, the resources for Native Americans are likely to improve in the future.

In conclusion, we discovered a novel risk factor for severe DR in Latinos and our study shows for the first time to our knowledge that NAA is associated positively with severe DR in Latinos, even after adjusting for social economic factors (education and income) and known risk factors. Therefore, genetic ancestry should be included in the risk assessment of DR in research studies of Latinos with T2DM.

Acknowledgments

The authors thank the study participants in LALES, and study staff who helped with the data collection.

Supported in part by National Institutes of Health (NIH; Bethesda, MD, USA) Grants U10EY011753 (RV), R01EY022651 (XG), and P30EY001792 (departmental core grant), and an unrestricted departmental grant from Research to Prevent Blindness. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences (CTSI) Grant UL1TR000124 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) Grant DK063491. The authors alone are responsible for the content and writing of the paper.

Disclosure: **X. Gao**, None; **W.J. Gauderman**, None; **P. Marjoram**, None; **M. Torres**, None; **Y.-D.I. Chen**, None; **K.D. Taylor**, None; **J.I. Rotter**, None; **R. Varma**, None

References

1. Fong DS, Aiello L, Gardner TW, et al. Retinopathy in diabetes. *Diabetes Care*. 2004;27(suppl 1):S84-S87.
2. Foster A, Resnikoff S. The impact of Vision 2020 on global blindness. *Eye (Lond)*. 2005;19:1133-1135.
3. Klein R, Klein B. Vision disorders in diabetes. In: Harris MI, Hamman RF, eds. *Diabetes in America*. Bethesda, MD: National Institutes of Health; 1985:293-338.
4. Haffner SM, Fong D, Stern MP, et al. Diabetic retinopathy in Mexican Americans and non-Hispanic whites. *Diabetes*. 1988;37:878-884.
5. Harris MI, Klein R, Cowie CC, Rowland M, Byrd-Holt DD. Is the risk of diabetic retinopathy greater in non-Hispanic blacks and Mexican Americans than in non-Hispanic whites with type 2 diabetes? A U.S. population study. *Diabetes Care*. 1998;21:1230-1235.
6. Klein R, Klein BE, Moss SE. Epidemiology of proliferative diabetic retinopathy. *Diabetes Care*. 1992;15:1875-1891.
7. Varma R, Torres M, Pena F, Klein R, Azen SP. Prevalence of diabetic retinopathy in adult Latinos: the Los Angeles Latino Eye Study. *Ophthalmology*. 2004;111:1298-1306.
8. West SK, Klein R, Rodriguez J, et al. Diabetes and diabetic retinopathy in a Mexican-American population: Proyecto VER. *Diabetes Care*. 2001;24:1204-1209.
9. Varma R, Macias GL, Torres M, Klein R, Pena FY, Azen SP. Biologic risk factors associated with diabetic retinopathy: the Los Angeles Latino Eye Study. *Ophthalmology*. 2007;114:1332-1340.
10. Mao X, Bigham AW, Mei R, et al. A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet*. 2007;80:1171-1178.
11. Price AL, Patterson N, Yu F, et al. A genomewide admixture map for Latino populations. *Am J Hum Genet*. 2007;80:1024-1036.
12. Salari K, Choudhry S, Tang H, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genet Epidemiol*. 2005;29:76-86.
13. Varma R, Paz SH, Azen SP, et al. The Los Angeles Latino Eye Study: design, methods, and baseline data. *Ophthalmology*. 2004;111:1121-1131.
14. Early Treatment Diabetic Retinopathy Study Research Group. Early photocoagulation for diabetic retinopathy: ETDRS report number 9. *Ophthalmology*. 1991;98(suppl):766-785.
15. Varma R, Choudhury F, Klein R, Chung J, Torres M, Azen SP. Four-year incidence and progression of diabetic retinopathy and macular edema: the Los Angeles Latino Eye Study. *Am J Ophthalmol*. 2010;149:752-761.
16. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
17. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567-1587.
18. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945-959.
19. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467:52-58.
20. Gao X, Starmer J. Human population structure detection via multilocus genotype clustering. *BMC Genet*. 2007;8:34.
21. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655-1664.
22. 1000 Genomes Project. A deep catalog of human genetic variation. Available at: <http://www.1000genomes.org/>, Accessed March 10, 2012.
23. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2009.
24. Kuo JZ, Wong TY, Rotter JI. Challenges in elucidating the genetics of diabetic retinopathy. *JAMA Ophthalmol*. 2014;132:96-107.
25. Cooke Bailey JN, Sobrin L, Pericak-Vance MA, Haines JL, Hammond CJ, Wiggs JL. Advances in the genomics of common eye diseases. *Hum Mol Genet*. 2013;22:R59-R65.
26. Fu YP, Hallman DM, Gonzalez VH, et al. Identification of diabetic retinopathy genes through a genome-wide association study among Mexican-Americans from Starr County, Texas. *J Ophthalmol*. 2010;2010:861291.
27. Sheu WH, Kuo JZ, Lee IT, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. *Hum Mol Genet*. 2013;22:3165-3173.
28. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet*. 2011;20:2472-2481.
29. Grassi MA, Tikhomirov A, Ramalingam S, et al. Replication analysis for severe diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 2012;53:2377-2381.
30. Edwards TL, Gao X. Methods for detecting and correcting for population stratification. *Curr Protoc Hum Genet*. 2012;1:Unit 1.22.1-1.22.14.
31. Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061-1073.