Genetic Admixture: A Tool to Identify Diabetic Nephropathy Genes in African Americans

Jasmin Divers, PhD; Shahriar Moossavi, MD, PhD; Carl D. Langefeld, PhD; Barry I. Freedman, MD

Diseases with an inherited component that demonstrate different prevalence in various ancestral populations can now be studied using admixture mapping in an appropriate admixed population. This strategy called mapping by admixture linkage disequilibrium or MALD utilizes polymorphic genetic markers that are spaced throughout the genome to identify genomic regions where the estimated admixture proportion is significantly different than its expected value. These genetic markers are selected based on their ancestry informativeness content. The MALD approach assumes that genomic regions showing excess ancestry from the ancestral population with higher disease prevalence, in the sample of admixed individuals, are more likely to harbor polymorphisms that confer higher risk to disease than others. Certain conditions including essential hypertension, type 2 diabetes mellitus and common complex forms of nephropathy demonstrate clear differences in disease frequency in individuals of African and European descent and appear particularly suited to this type of analysis. Genetic admixture can also cause confounding in association studies conducted on an admixed sample leading to inflated type I error rates and possible loss of power. This manuscript describes the background, methodologies and uses for admixture mapping in the search for genes that underlie type 2 diabetes mellitus and its associated nephropathy in the African American population, and statistical methods to address the confounding issues in genetic association tests. (*Ethn Dis*. 2008;18:384–388)

Key Words: Genetics, Diabetic Nephropathy, Ethnicity, African Americans, Admixture

From the Division of Public Health Sciences, Department of Biostatistical Sciences, Section on Statistical Genetics and Bioinformatics, and Center for Public Health Genomics (JD, CDL) and the Department of Internal Medicine, Section on Nephrology (SM, BIF), Wake Forest University School of Medicine, Winston-Salem, North Carolina.

Address correspondence and reprint requests to Barry I. Freedman, MD; Section on Nephrology; Wake Forest University School of Medicine; Medical Center Boulevard; Winston-Salem, NC 27157-1053; 336-716-6192; 336-716-4318 (fax); bfreedma@wfubmc.edu

Introduction

Approximately 20 million Americans, or one of every 14 individuals has diabetes mellitus, with an additional 6.2 million undiagnosed. Worldwide, the prevalence of diabetes is expected to increase by nearly 50% in this decade alone. Diabetes is currently the leading cause of end-stage renal disease (ESRD) in developed nations. Minority populations such as African Americans (AA) are at substantially greater risk, compared to European Americans (EA), for developing diabetes and diabetic nephropathy. This manuscript reviews the use of mapping by admixture linkage disequilibrium (MALD) for the detection of genes that contribute to the observed ethnic disparities in susceptibility to type 2 diabetes mellitus and diabetic complications.

Admixture: Definition and Implications

Admixture describes a process by which individuals from two (or more) populations who have been separated for long periods of time come together and create offspring. The AA population can best be described genetically as an admixed population formed by the gene flow between mainly individuals with European and African ancestry. Evidence of genetic contribution from other ancestral populations, such as American Indians and Asians, can be observed to lesser degrees. Previous studies report 3%–26% admixture of Caucasian genes in AA, with higher proportions found in northern compared to southern states.[1,2] The admixture process varies widely even within geographic regions. For example, AA in Charleston, South Carolina had an estimated Caucasian admixture rate of 11.6%, while this proportion was estimated at 22.5% for AA living in New Orleans.[3]

Long[4] described two models of admixture: 1) the intermixture and 2) the continuous gene flow. In the ''intermixture model,'' the admixture event occurs once between the ancestral populations. The subsequent generations result from random mating among individuals of the hybrid population, assuming no mutation or selection. The ''continuous gene flow model'' assumes a continuous flow from a donor to a recipient population. Many researchers believe the African American and Hispanic populations are prime examples of the continuous gene flow model. Importantly, this admixture process occurred relatively recently, approximately 20 generations ago.

A recently admixed population offers several experimental advantages for mapping of disease genes, similar to an inbred cross design. The primary advantage of studying admixed populations is derived from linkage disequilibrium (LD) created by the admixture event. Linkage disequilibrium, or less frequent recombination of nearby stretches of genetic material, decays rapidly in human populations. This decay increases with recombination rate between adjacent markers and the number of generations after the admixture event. Because AAs have experienced recent admixture, their genomes have had less recombination (or reshuffling) since population mixing began, and the stretches of identical ancestry often extend over many megabases. Parra et al[3] reported significant LD on chromosomal segments as long as 20 centiMorgans. This important

observation constituted the basis of using recently admixed populations in genetic linkage and association studies. The key advantage of admixture mapping is that it requires substantially fewer genetic markers for a genomic search than do other methods of association mapping (haplotype or direct association studies). The number of markers for a genomic seared required to conduct these studies is greatly reduced by using ancestry informative markers (AIMs) instead of random markers.[5–9] AIMs have allele frequencies, which vary greatly between the two ancestral populations.

DIABETIC NEPHROPATHY IN AFRICAN AMERICANS

More than 45,000 Americans with diabetes began treatment for ESRD in 2004, equating to 44% of the incident dialysis population.[10] The prevalence of diabetes differs across ethnic groups. Among those >20 years of age, the overall US prevalence of diabetes is 9.6%; 8.7% in European Americans (EAs), 13.3% in AAs, 9.5% in Hispanic Americans and 15% in American Indians and Alaskan natives.[11] In addition, the likelihood of developing diabetic nephropathy differs markedly in each ethnic group. African Americans with diabetes have an overall fourfold higher incidence rate of ESRD, relative to EAs. This ethnic disparity, although attenuated slightly, persists among those with equal access to healthcare.[12,13] Marked ethnic differences in the amount of coronary artery calcified plaque have also been observed in hypertensive and diabetic populations.[14]

Diabetes and diabetic nephropathy cluster within families, perhaps more strongly in AA families where 32% of women and 27% of men reported having a first or second degree relative with ESRD, compared to only 15% of EA women and 12% of EA men.[15,16] The risk for early diabetic kidney disease appears to be similar in AAs and EAs.[17,18] This suggests that: diabetic nephropathy may progress more rapidly to ESRD in AAs; a paradoxical higher non-renal death rate may exist in EAs; or a combination of these factors may be present. Ethnic disparities in the incidence rate of diabetes and diabetic nephropathy and familial aggregation strongly suggest that inherited factors contribute to disease susceptibility. Admixture mapping may be a useful tool to assist geneticists in detecting diabetic nephropathy susceptibility genes.

USE OF ADMIXTURE IN LINKAGE AND ASSOCIATION STUDIES

Recently, there has been increased interest in association studies as a useful and powerful approach to map common disease genes.[19,20] The issue of identifying which populations will be best suited for LD mapping has been the subject of much debate.[21,24] The extent of LD is a complex function of a number of genetic and evolutionary factors such as mutation, recombination and gene conversion rates, demographic and selective events, and the age of the mutation itself. Some of these factors affect the whole genome while others only affect particular genomic regions. Because the LD created by the admixture process is relatively recent, it paradoxically offers great promise for gene mapping, especially for traits such as prostate cancer, hypertension, diabetic complications, and multiple sclerosis that have markedly different prevalence among the ancestral populations and cause confounding issues in genetic association studies. The confounding issues result from correlations between markers that are due to the admixed origin of the populations being studied.

In the 1950s, indications suggested that admixture mapping might be beneficial to localize genes underlying ethnic variation in disease. The statistical basis of this approach was first explored by Chakraborty and Weiss,[25] and subsequently by Stephens, Briscoe and O'Brien who termed it "mapping by admixture linkage disequilibrium" (MALD).[26,27] Admixture mapping has been used successfully to identify disease genes or loci associated with prostate cancer,[28] hypertension,[29] and multiple sclerosis.[30] The application of admixture mapping had been limited until the availability of genome-wide sets of highly informative AIMs and adequate statistical tools to successfully conduct these studies. One of the first applications was proposed by McKeigue[31] who demonstrated that conditioning on parental admixture in an association test between a marker and trait locus is equivalent to a test of linkage. He later applied this method[32–34] to find evidence of linkage between FY and AT3, two markers located 22 cM apart from each other, by testing for association of ancestry conditional on parental admixture. Zhu et al[35] refined McKeigue's test by generalizing this method such that it was valid under both the intermixture and continuous gene flow models. Patterson[36] proposed a Bayesian whole-genome statistic, which could be applied in linkage analysis. Hoggart[37] developed an affected only association test for rare diseases that provided, using the same sample size, comparable results to a case-control study and Montana and Pritchard[6,38] presented an association test and genome-wide scan methodology using admixture mapping. Wen-Chung et al[39] found that the long range of linkage disequilibrium observed in admixed individuals made interval transmission disequilibrium tests more powerful than marker-by-marker transmission disequilibrium testing. Montana and Pritchard also showed that random single nucleotide polymorphisms (SNPs) could be an acceptable alternative when AIMs are not available. Smith et al[7] then identified a set of more than 2,000 promising SNPs for MALD analysis. New marker panels have since been developed for admixture

mapping in Hispanic Americans[8,9,40] Assessing the statistical significance of linkage peaks observed in admixture mapping represents an important issue. Earlier methods rely on either the Bonferroni correction or computer-intensive methods to determine whether observed *P* values were statistically significant. Sha et al[41] showed the first order Markov Chain assumptions made about the individual ancestry distribution could also be applied to a statistical test for linkage. They then utilized this assumption to derive analytical significance tests to determine whether an observed linkage peak was significantly significant.

POPULATION STRATIFICATION IN GENETIC ASSOCIATION TESTS AMONG ADMIXED SAMPLES

Investigators have used measures of individual genetic ancestry as covariates in association studies in order to control the type I error which may be inflated when admixture can be a confounder.[42,43] These methods are coined structured association testing (SAT) and can be divided into those that estimate the ancestry proportion of each individual in the sample and use this estimate as a covariate in the test for association,[43–45] and those that rely upon a measure of genetic background obtained by performing a principal component analysis (PCA) on the genotypic data to provide control for population stratification in the test for genetic association.[46–48] The first approaches propose starting by obtaining individual ancestry proportion estimates and using this estimate as a covariate in the test for association. Several algorithms and software have been developed to estimate individual ancestry.[6,35,37,38,49–53] These software programs have been tested to different degrees and some, STRUCTURE for example, widely applied. Several reports demonstrate that these programs provide similar individual estimates when they are run under the appropriate models. However, some tend to cluster individuals closer to the poles (1 or 0) more often than others, when the ancestral allele frequency estimate is biased.

Recent work by Redden and colleagues[54] has shown that simply including the individual admixture estimates as a covariate in the test for association may not be sufficient. Residual confounding may still occur in non-additive models, for example when testing for dominant, recessive or overdominant effects. Divers et al[55] showed that the individual ancestry proportion estimates should only be seen as error contaminated measures of the true individual ancestry proportions and that ignoring these measurement errors may inflate type I errors.

The principal component approach adjusts for population stratification through the inferred genetic background variables obtained through the first (or the first several) principal components computed from the genotyped data. The effect of the genetic background variable on the phenotype of interest is then modeled as a linear function[56,57] or a non-linear function.[48,58]

ADMIXTURE VS SOCIOECONOMIC STATUS

Variation in disease prevalence and severity cannot be explained by either race or ethnicity alone. They should be placed in the context of a rather complex set of criteria that include social, political, biological, and economic aspects, which are unique to each society.[59] The validity of racial/ethnic categories for biomedical and genetic research has recently been challenged. Some researchers propose that a race-neutral approach should be used, with the focus on genetic clustering rather than self-identified ethnicity for human genetic categorization[60,61]; while others claim that there is no biological basis for race. Risch et al[62–63] have provided an epidemiologic perspective on the issue of human categorization in biomedical and genetic research that strongly supports the continued use of self-identified race and ethnicity.

CONCLUSIONS

Genetic variation among races is small, which explains why relatively few diseases can be studied through admixed populations. However, diseases with different risks in populations of Africans and Europeans are among those that can be studied with admixture mapping. Several disorders, including hypertension, diabetes and diabetic nephropathy, are among the leading causes of morbidity and mortality in the US and abroad. Therefore, there is a great need for statistical methods that can combine the information provided by admixture and other indicators of socio-economic status to provide valuable insights regarding the occurrence, progression, treatment and hopefully cure of these diseases.

REFERENCES

1. Glass B, Li CC. The dynamics of racial intermixture; an analysis based on the American Negro. *Am J Hum Genet*. 1953;5:1–20.
2. Reitnauer PJ, Go RCP, Acton RT, et al. Evidence for genetic admixture as a determinant in the occurrence of insulin-dependent diabetes-mellitus in United-States blacks. *Diabetes*. 1982;31:532–537.
3. Parra EJ, Marcini A, Akey L, et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet*. 1998;63:1839–1851.
4. Long JC. The genetic structure of admixed populations. *Genetics*. 1991;127:417–428.
5. McKeigue PM. Prospects for admixture mapping of complex traits. *Am J Hum Genet*. 2005;76:1–7.
6. Montana G, Pritchard JK. Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet*. 2004; 75:771–789.
7. Smith MW, Patterson N, Lautenberger JA, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet*. 2004;74:1001–1013.

8. 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.
9. Price AL, Patterson N, Yu F, et al. A genomewide admixture map for Latino populations. *Am J Hum Genet*. 2007;80: 1024–1036.
10. US Renal Data System. *USRDS 2006 Annual Data Report: Atlas of End Stage Renal Disease in the United States*. Bethesda, MD: National Institute of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2006.
11. Center for Disease Control and Prevention. *National Diabetes Fact Sheet: General Information and National Estimates on Diabetes in the United States, 2005*. Atlanta, GA: U.S Department of Health and Human Services, Center for Disease Control and Prevention; 2005.
12. Karter AJ, Ferrara A, Liu JY, Moffet HH, Ackerson LM, Selby JV. Ethnic disparities in diabetic complications in an insured population. *JAMA*. 2002;287:2519–2527.
13. Young BA, Maynard C, Boyko EJ. Racial differences in diabetic nephropathy, cardiovascular disease, and mortality in a national population of veterans. *Diabetes Care*. 2003;26:2392–2399.
14. Freedman BI, Hsu FC, Langefeld CD, et al. The impact of ethnicity and sex on subclinical cardiovascular disease: the Diabetes Heart Study. *Diabetologia.*. 2005;48:2511–2518.
15. Freedman BI, Volkova NV, Satko SG, et al. Population-based screening for family history of end-stage renal disease among incident dialysis patients. *Am J Nephrol*. 2005;25: 529–535.
16. McClellan W, Speckman R, McClure L, et al. Prevalence and characteristics of a family history of end-stage renal disease among adults in the United States population: Reasons for Geographic and Racial Differences in Stroke (REGARDS) Renal Cohort Study. *J Am Soc Nephrol*. 2007;18:1344–1352.
17. Coresh J, Astor BC, Greene T, Eknoyan G, Levey AS. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *Am J Kidney Dis*. 2003;41:1–12.
18. Trivedi HS, Pang MM. Discrepancy in the epidemiology of nondiabetic chronic renal insufficiency and end-stage renal disease in black and white Americans: the third National Health and Nutrition Examination Survey and United States Renal Data System. *Am J Nephrol*. 2003;23:448–457.
19. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516–1517.
20. Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res*. 2000;10:1435–1444.
21. Nordborg M, Tavare S. Linkage disequilibrium: what history has to tell us. *Trends in Genetics*. 2002;18:83–90.
22. Kaessmann H, Zollner S, Gustafsson AC, et al. Extensive linkage disequilibrium in small human populations in Eurasia. *Am J Hum Genet*. 2002;70:673–685.
23. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet*. 1999;23:397–404.
24. Eaves IA. The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet.*. 2000;25: 320–323.
25. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences of the United States of America*. 1988;85: 9119–9123.
26. Stephens JC, Briscoe D, Obrien SJ. Mapping by admixture linkage disequilibrium in human-populations - limits and guidelines. *Am J Hum Genet*. 1994;55:809–824.
27. Briscoe D, Stephens JC, O'Brien SJ. Linkage disequilibrium in admixed populations: applications in gene mapping. *J Hered*. 1994;85: 59–63.
28. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103:14068–14073.
29. Zhu X, Luke A, Cooper RS, et al. Admixture mapping for hypertension loci with genome-scan markers. *Nature Genetics*. 2005;37: 177–181.
30. Reich D, Patterson N, De Jager PL, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nature Genetics*. 2005;37:1113–1118.
31. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: Methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet*. 1998;63:241–251.
32. McKeigue PM. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admired populations. *Am J Hum Genet*. 1997;60:188–196.
33. McKeigue PM. Multipoint admixture mapping. *Genetic Epidemiology*. 2000;19: 464–465.
34. McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Annals of Human Genetics*. 2000;64:171–186.
35. Zhu X, Cooper RS, Elston RC. Linkage analysis of a complex disease through use of admixed populations. *Am J Hum Genet*. 2004;74:1136–1153.
36. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet*. 2004;74:979–1000.
37. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. *Am J Hum Genet*. 2004;74:965–978.
38. 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.
39. Wen-Chung L, Yun-Chun Y. Admixture mapping using interval transmission/disequilibrium tests. *Annals of Human Genetics*. 2003;580–588.
40. Tian C, Hinds DA, Shigeta R, et al. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet*. 2007;80:1014–1023.
41. Sha Q, Zhang X, Zhu X, Zhang S. Analytical correction for multiple testing in admixture mapping. *Hum Hered*. 2006;62:55–63.
42. Pritchard JK, Stephens M, Donnelly PJ. Correcting for population stratification in linkage disequilibrium mapping studies. *Am J Hum Genet*. 1999;65:A101.
43. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*. 1999;65:220–228.
44. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theoretical Population Biology*. 2001;60:227–237.
45. Ziv E, Burchard EG. Human population structure and genetic association studies. *Pharmacogenomics.*. 2003;4:431–441.
46. Zhang S, Zhu X, Zhao H. On a semiparametric test to detect ssociations between quantitative traits and candidate genes using unrelated individuals. *Genetic Epidemiology*. 2006;24:44–56.
47. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38:904–909.
48. Zhang SL, Zhu XF, Zhao HY. On a semiparametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. *Genetic Epidemiology*. 2003;24:44–56.
49. Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations

in stratified populations. *Am J Hum Genet*. 2003;72:1492–1504.
50. Purcell S, Sham P. Properties of structured association approaches to detecting population stratification. *Hum Hered*. 2004;58:93–107.
51. Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*. 2005;28:289–301.
52. Zhang C, Chen K, Seldin MF, Li HZ. A hidden Markov modeling approach for admixture mapping based on case-control data. *Genetic Epidemiology*. 2004;27:225–239.
53. Patterson N. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet*. 2004;74:979–1000.
54. Redden D, Divers J, Vaughan L, et al. Regional admixture mapping and structured association testing: conceptual unification and an extensible general linear model. *PLoS Genetics*. 2006;2:1254–1264.
55. Divers J, Vaughan LK, Padilla MA, Fernandez JR, Allison DB, Redden DT. Correcting for measurement error in individual ancestry estimates in structured association tests. *Genetics*. 2007;176:1823–1833.
56. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909.
57. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genetics*. 2006;2:2074–2093.
58. Chen HS, Zhu X, Zhao H, Zhang S. Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. *Annals of Human Genetics*. 2003;67:250–264.
59. Goodman AH. Why genes don't count (for racial differences in health). *Am J Public Health*. 2000;90:1699–1702.
60. Schwartz RS. Racial profiling in medical research. *N Engl J Med*. 2001;344:1392–1393.
61. Wilson JF, Weale ME, Smith AC, et al. Population genetic structure of variable drug response. *Nat Genet*. 2001;29:265–269.
62. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol*. 2002;3(7):Epub2002. Jul 1.
63. Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*. 2003;348:1170–1175.

AUTHOR CONTRIBUTIONS

Design concept of study: Freedman, Langefeld, Divers
Acquisition of data: Freedman, Divers
Data analysis and interpretation: Freedman, Divers, Langefeld, Moossavi
Manuscript draft: Moossavi, Freedman, Divers
Statistical expertise: Divers, Langefeld
Acquisition of funding: Freedman, Langefeld
Administrative, technical, or material assistance: Moossavi, Divers, Langefeld, Freedman
Supervision: Freedman and Langefeld