Editorial

Local Ancestry Association, Admixture Mapping, and Ongoing Challenges

Arya Mani, MD

Genome-wide association is the most widely used methodology for mapping of disease genes for complex traits. Among many limitations of this approach is the requirement for investigating large disease populations and use of sizeable number of genetic markers, which necessitates correction for multiple comparisons and is a reason for its reduced power for signal detection. These limitations have prompted alternative approaches to identify novel disease loci. One such approach is the admixture mapping, which provides a more powered alternative for gene discovery in admixed populations for a trait that is differentially distributed in the ancestral populations. One such example is the African American population, which is at increased risk for chronic and end-stage renal disease, HIV-associated nephropathy, and focal segmental glomerulosclerosis, which are believed to originate from African ancestry.

See Article by Shendre et al

In the classical sense, admixture is the result of breeding between individuals from ≥2 isolated populations. The consequence of the gene flow is temporary generation of long haplotype blocks that contains genetic variants of one or another population. The approach in admixture mapping is based on the assumption that in the admixed populations the disease-causing alleles will be more frequent on chromosome segments derived from the ancestral population with the higher prevalence for the disease. The admixture mapping associates ancestry of haplotype block with a specific trait using differences in allele frequencies of the ancestral population. The haplotype blocks become shorter with increasing age of the admixed population through higher recombination events during each meiosis. Shorter admixed linkage disequilibrium (LD) blocks require a higher density of markers to differentiate chromosome ancestry transition. In admixed populations such as African Americans or Hispanics, the gene flow has originated within the last several hundred years, resulting in linked alleles that show extended LD relative to the ancestral populations. In the African American population, gametes are

roughly 80% derived from African and from 20% European ancestries.[1] Admixed LD in African Americans is in average 30-cM regions[2] with statistically strongest admixed LD spanning a 17-cM regions, reflecting 6 to 7 generations of admixture.[3] Admixture has also occurred in many other world populations such as between the Spanish and Amerindians and with considerable geographical heterogeneity.

Ancestry at any given locus, known as local ancestry, is inferred based on haplotype transitions from 1 parental population to another. Hidden Markov Model algorithms is used to identify markers that have a significantly greater than chance likelihood of being on the same segment. Admixture mapping is performed using 2 distinct case–control and case-only methodologies. For case–control studies, locus-specific ancestry at each ancestry-informative marker between cases and controls is calculated. The admixture mapping has the power to detect association with relatively modest odds ratios with 2500 or fewer cases.[3] A moderate number (≈1500–2500) of ancestry-informative marker is initially used for gene mapping, followed by dense mapping to identify the functional allele. For case-only studies, the initial mapping is performed by comparing extent of ancestry in each locus to genome-wide average ancestry.[4,5] Using statistical analysis, deviation of local ancestry away from the genome-wide average will be shown as distinct peaks. Use of far fewer markers compared with GWAS and hence the requirement of modest correction for multiple comparisons is an obvious advantage of the admixture mapping over GWAS.

The major requirement for admixture mapping panels is a set of genetic markers that are informative, that is, have allele frequencies widely different between ancestral populations (shown as δ) and are not in LD with each other. Over the last decade, there has been a rapid growth in development of panels of ancestry-informative markers for admixture typing[6,7] and software programs and statistical methods for the analysis. Ideal is an allele that is fixed in a population (fixation index [Fst]=1), but given the scarcity of such markers alleles with Fst >0.5 are acceptable for the mapping. Local Ancestry Inference in Admixed Populations Using LD is a popular software for admixture mapping in unrelated individuals and Local Ancestry Inference in Admixed Populations Using Haplotype Data for nuclear trios. High-density genome-wide admixture mapping panels have been for obvious reasons first constructed for African American, Latino/Hispanics, and Uyghurs. Although admixture between 2 populations has been modeled in most admixture studies, 3-way admixture is not uncommon. Latinos/Hispanics for instance have ancestral contributions that based on their geographical distribution ranges from different Native American populations and

Europeans to Africans. Like every association study, replication of the data is required in independent population.

Most recent development is the joint use of admixture mapping and association studies that combines the power of detecting disease susceptibility loci or trait loci with fewer markers and smaller study population in admixture mapping with that of an association testing that improves the resolution.[8] For related individuals, transmission–disequilibrium test has been combined with local ancestry mapping.[9]

In the study by Shendre et al,[10] authors examine the association of local European ancestry (LEA) with common carotid artery intima-media thickness (cCIMT) among African Americans using a well-defined African American population of MESA (Multi-Ethnic Study of Atherosclerosis). MESA was designed to characterize subclinical cardiovascular disease in 45 to 84 years old men and women without prevalent cardiovascular disease. African Americans comprised 28% (n=1891). A total of 1554 African Americans were genotyped using the Affymetrix Human Single-Nucleotide Polymorphism (SNP) array 6.0. The details of the ultrasound method used to measure cCIMT in the MESA cohort have been described elsewhere.[11] Incident events had to be documented and meet established criteria that were reviewed by 2 physicians and determined over an average follow-up of 9.3 years from baseline. After extensive filtering, 611 449 autosomal SNPs (579 847 SNPs for the ARIC study [Atherosclerosis Risk in Communities]) remained for inclusion in the admixture estimation and association analyses. In addition, 595 SNPs or ancestry-informative markers that had been associated with cCIMT in previous studies were included. LD was used to determine LEA. HapMap phase II and III data of Central European and Yoruba trios from Ibadan, Nigeria were used as reference population and as parameters for the Hidden Markov Model to determine the local ancestry within a 300 SNPs long window-based framework. In this method, ancestry at each SNP is determined for each individual based on haplotype sets from ancestral populations using a dense set of markers. Local ancestry was graded based on the number of European ancestry alleles at each SNP, and their average was used to measure global ancestry for each individual. Ancestry association with cCIMT was tested using the PLINK. Linear regression model was used after adjustment for global ancestry and cardiovascular risk factors. The total number of independent tests in the analyses was estimated to be 148.6 based on a method proposed by Shriner et al[12] that calculates the total number of effective independent tests; this resulted in a threshold significance level of $3.36×10^{-4}$. This threshold was considerably lower than those used for GWAS, providing considerably greater power for signal detection. LEA regions associated with cCIMT were evaluated for associations with stroke and cardiovascular disease events in the MESA cohort.

The analysis identified the LEA gene region in the SERGEF gene on chromosome 11, which was also associated with higher odds of stroke as the only region that achieved genome-wide significance (β=0.0137; $P=2.98×10^{-4}$). The LEA gene region in the TPH1 gene

on chromosome 11 showed subthreshold association with cCIMT ($P=3.42×10^{-4}$). A previously associated SNP rs2081015 in the coding region of the GALNT10 gene was reproduced in this study. Although nominally significant

Table. Loci identified by Admixture Mapping

Phenotype	Study Population	Chromosome	Peak SNP/Marker in the region
Hypertension	African American	6q24.1	
	African American	21q21	
	African American	6q21	
	MA	6q21	rs2272996
	African American	6q21	rs2272996
	African American	6q24.3	rs2272996
ESRD	African American	22q12	rs482181
Nondiabetic ESRD		22q13.1	rs482181(MYH9)
HIVAN		22q13.1	rs482181
FSGS		22q13.1	rs482181
FSGS and HIVAN	African American	22q13.1	
	African American	2p14	rs6724395
Obesity	African American	2q23.3	
BMI	African American	Xq25	
	African American	Xq13	
	African American	5q13.3	
	African American	3q29	D3S1311
	African American	5q14–5q25	D5S2501
	African American	15q26	D15S816
Type II DM	African American	12	rs1565728
Breast cancer(ER+PR+ vs ER−PR−)	African American		
Localized tumors	African American		
White blood cell count	African American	1q	rs2814778
Peripheral arterial disease: ankle-brachial index	African American	11	rs9665943
	African American		rs9665943
LDL-C levels	African American	3q4q	D3S2427, D4S2367
HDL-C levels	African American	8q	D8S1136
		14q	GATA193A07
		9q	
TG levels	African American	8q15q	D8S1179D15S652
Multiple sclerosis	African American	1	Centromeric
Prostate cancer		8q24	3.8 Mb

BMI indicates body mass index; DM, diabetes melitus; ER, estrogen; ESRD indicates end-stage renal disease; FSGS, focal segmental glomerulosclerosis; HDL-C, high-density lipoprotein cholesterol; HIVAN, HIV-associated nephropathy; LDL-C, low-density lipoprotein cholesterol; MA, Mexican-American; PR, progesterone; and TG, triglyceride.

regions associated with cCIMT ($P<0.05$) in the MESA cohort were replicated in 3000 African Americans from the ARIC cohort, the genome-wide significant association of LEA with cCIMT in the *SERGEF* gene could not be replicated. There were many subthreshold associations between some of the LEA regions associated with cCIMT and clinical cardiovascular disease events, that is, protective association in relation to the cardiovascular events and opposite thereof in case of stroke. Interestingly, regions of LEA associated with cCIMT at previously significant SNPs, that is, rs2081015, also showed subthreshold levels of significance ($\beta=-0.0118$, $P=2.45\times10^{-3}$).

Genome-wide admixture association studies for cCIMT are not novel approaches. The novelty here is in the use of local ancestry association approach to potentially improve power, while reducing false discovery rate. Overall, the study was able to identify novel associations in relation to cCIMT and clinical events. None of these LEA regions, however, met the threshold of genome-wide significance. Whether the signal on chromosome 11 in MESA population is a robust or false-positive finding should hence be examined in future admixture mappings. The result of the study should by no means undermine the power of admixture mapping in finding disease-associated loci in admixed population. Over last decade, a number of admixture mapping studies have led to identification of novel genetic loci for hypertension, end-stage and chronic renal disease, and diabetes mellitus among others (Table). The admixture mapping also holds promise for identifying the genetic contribution to type II diabetes mellitus and obesity in Amerindians. This study, however, reminds us all of the limitations of all currently available mapping methods and the constant need for their improvement.

Disclosures

References

1. Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet*. 2005;76:463–477. doi: 10.1086/428654.
2. Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW. Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am J Hum Genet*. 2000;66:969–978. doi: 10.1086/302820.
3. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet*. 2004;74:979–1000. doi: 10.1086/420871.
4. 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. doi: 10.1086/420855.
5. Montana G, Pritchard JK. Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet*. 2004;75:771–789. doi: 10.1086/425281.
6. Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, et al. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet*. 2004;114:263–271. doi: 10.1007/s00439-003-1058-6.
7. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet*. 2004;74:1001–1013. doi: 10.1086/420856.
8. Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WH, et al. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARe and a Breast Cancer Consortium. *PLoS Genet*. 2011;7:e1001371. doi: 10.1371/journal.pgen.1001371.
9. Tang H, Siegmund DO, Johnson NA, Romieu I, London SJ. Joint testing of genotype and ancestry association in admixed families. *Genet Epidemiol*. 2010;34:783–791. doi: 10.1002/gepi.20520.
10. Shendre A, Wiener H, Irvin MR, Zhi D, Limdi NA, Overton ET, et al. Admixture mapping of subclinical atherosclerosis and subsequent clinical events among blacks in 2 large cohort studies. *Circ Cardiovasc Genet*. 2017;10:xxx–xxx. doi: 10.1161/CIRCGENETICS.116.001569.
11. Soliman EZ, Ding J, Hsu FC, Carr JJ, Polak JF, Goff DC Jr. Association between carotid intima-media thickness and pericardial fat in the Multi-Ethnic Study of Atherosclerosis (MESA). *J Stroke Cerebrovasc Dis*. 2010;19:58–65. doi: 10.1016/j.jstrokecerebrovasdis.2009.03.008.
12. Shriner D, Adeyemo A, Rotimi CN. Joint ancestry and association testing in admixed individuals. *PLoS Comput Biol*. 2011;7:e1002325. doi: 10.1371/journal.pcbi.1002325.

Circulation
Genomic and Precision Medicine

American Heart Association

Local Ancestry Association, Admixture Mapping, and Ongoing Challenges

Arya Mani

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://circgenetics.ahajournals.org/content/10/2/e001747