



Published in final edited form as:

*Genet Epidemiol.* 2013 January ; 37(1): 38–47. doi:10.1002/gepi.21687.

## Testing Genetic Association With Rare Variants in Admixed Populations

Xianyun Mao<sup>1</sup>, Yun Li<sup>2,3,4</sup>, Yichuan Liu<sup>1</sup>, Leslie Lange<sup>2</sup>, and Mingyao Li<sup>1,\*</sup>

<sup>1</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

<sup>2</sup>Department of Genetics, University of North Carolina, Chapel Hill, North Carolina

<sup>3</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina

<sup>4</sup>Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina

### Abstract

Recent studies suggest that rare variants play an important role in the etiology of many traits. Although a number of methods have been developed for genetic association analysis of rare variants, they all assume a relatively homogeneous population under study. Such an assumption may not be valid for samples collected from admixed populations such as African Americans and Hispanic Americans as there is a great extent of local variation in ancestry in these populations. To ensure valid and more powerful rare variant association tests performed in admixed populations, we have developed a local ancestry-based weighted dosage test, which is able to take into account local ancestry of rare alleles, uncertainties in rare variant imputation when imputed data are included, and the direction of effect that rare variants exert on phenotypic outcome. We used simulated sequence data to show that our proposed test has controlled type I error rates, whereas naïve application of existing rare variants tests and tests that adjust for global ancestry lead to inflated type I error rates. We showed that our test has higher power than tests without proper adjustment of ancestry. We also applied the proposed method to a candidate gene study on low-density lipoprotein cholesterol. Our results suggest that it is important to appropriately control for potential population stratification induced by local ancestry difference in the analysis of rare variants in admixed populations.

### Keywords

admixed population; rare variants; population stratification

## INTRODUCTION

Common complex diseases have been widely studied in genome-wide association studies (GWAS) involving millions of single nucleotide polymorphisms (SNPs). Although many common variants have been identified for disease association, a large portion of the disease heritability is still missing [Manolio et al., 2009]. Interest in rare variants naturally arises when searching in common variants exhausts. Early studies and recent studies [Dickson et

al., 2010; Hershberger et al., 2010; Wessel et al., 2010; Zawistowski et al., 2010] have shown the benefits of studying rare variants in common disorders. Advances in next-generation sequencing technologies and studies such as the 1000 Genomes Project Consortium [2010] provide a platform for studying rare variants. However, due to their low frequencies, rare variants do not fit well in the established genetic association regime, especially not in a single marker test fashion. To achieve sufficient power, analysis of rare variants often requires collapsing information across multiple markers in a genomic region.

In the past few years, various methods have been developed to analyze rare variants. Early methods such as the Cohort Allelic Sums Test [Morgenthaler and Thilly, 2007] and the Combined Multivariate and Collapsing approach [Li and Leal, 2008] compare the number of rare variants between cases and controls and examine whether they have equal burden. Madsen and Browning [2009] introduced a weighted sum statistic in which each variant is weighted by a function of the allele heterozygosity. These earlier methods use an arbitrary single minor allele frequency threshold to define rare variants. In contrast, Price et al. [2010] proposed a variable threshold approach to optimize the inclusion threshold of rare variants. However, none of the abovementioned methods differentiate rare variants that are associated with the disease in different directions and thus may lose power when both risk and protective variants exist. Recently, methods that explicitly model the risk and protective variants have been developed. The Weighted Haplotype and Imputation-based tests (WHaIT) [Li et al., 2010a] and the replication-based method [Ionita-Laza et al., 2011] separate the effect of protective variants from the risk ones based on the data and then combine their information into a single test statistic. The C-alpha method [Neale et al., 2011] circumvents the association direction problem by directly modeling the variance of allele counts. In contrast, the sequence kernel association test (SKAT) [Wu et al., 2011] uses a logistic kernel machine model to allow for complex relationships between rare variants and the phenotypic trait. The tests implemented in C-alpha and SKAT can both be categorized as score tests. Recently, Bayesian-based methods have also been developed for rare variants analysis. Yi and Zhi [2011] proposed a novel Bayesian generalized linear model approach that allows for disparate effects and uses different weights to different variants based on their contributions to phenotype. Yi et al. [2011] further extended this approach to jointly estimate group and individual-variant effects for both common and rare variants by utilizing a hierarchical generalized linear model framework.

Although the above methods have shown promise in the analysis of rare variants, none of them explicitly model population stratification, an issue that is well recognized in genetic association studies. Population stratification emerges when there is a systematic difference in allele frequencies among study subjects due to ancestry difference. Unrecognized population stratification can lead to both false-positive and false-negative findings and can obscure true association signals if not appropriately corrected. Population stratification is often observed in genetic association studies of common variants and thus controlling population stratification has become a routine in GWAS focusing on common variants [Epstein et al., 2007; Li et al., 2010a; Price et al., 2006]. However, the impact of population stratification on the analysis of rare variants has not been systematically evaluated, especially not in admixed populations. Because rare variants arose more recently than common variants, they are more likely to be population specific. This is supported by findings from the 1000 Genomes Project Consortium [2010], which has shown that the number of rare variants differs significantly among different populations. Additionally, Mathieson and McVean [2012] showed that rare variants typically demonstrate different and stronger stratification than common variants, and such stratification cannot be corrected by existing methods. These empirical findings have important implications for the analysis of admixed populations, such as African Americans and Hispanic Americans, who are recently admixed and have inherited ancestry from more than one continent. Due to the allelic

spectrum differences among the original ancestral populations, it is possible that there is unrecognized hidden population structure in certain genomic regions of admixed samples.

In this paper, we show that substantial differences in ancestry exist in certain regions of the genome between cases and controls even when their overall ancestry proportions are comparable. This suggests that local ancestry difference could lead to population stratification that cannot be corrected by traditional methods such as EIGENSTRAT [Mathieson and McVean, 2012; Price et al., 2006] because these methods aim to capture global population structure using genetic markers across the entire genome, whereas subtle differences in local ancestry might be diluted due to the inclusion of markers from other genomic regions [Qin et al., 2010; Wang et al., 2011]. To correct for population stratification induced by local ancestry difference, we develop an Ancestry-Based Weighted Dosage Score (AWDS) for the analysis of rare variants. Our AWDS test shares similarity with the weighted dosage score (WDS) test in WHaIT [Li et al., 2010a] in that both take into account the direction of association to allow for risk and protective alleles. The key difference is that AWDS is able to adjust the WDS by local ancestry of the test region, and thus allowing the local ancestry difference among study subjects to be appropriately controlled. Through simulations using whole-genome sequence data, we confirm that AWDS has controlled type I error rates when population stratification is present, whereas the naïve application of existing rare variants tests such as WDS and modified versions of WDS that include either global principal components (PCs; obtained from markers distributed across the genome) or local PCs (obtained from markers within a 20 Mb window of the test region) as covariates all yield inflated type I error rates. We demonstrate that our new test has greater power than tests without proper adjustment of ancestry. We also applied our new test to a candidate gene study on low-density lipoprotein cholesterol (LDL-C). Our results suggest that it is important to appropriately control for population stratification induced by local ancestry difference in the analysis of rare variants in admixed populations.

## METHODS

We will develop an ancestry-based weighted dosage test to examine the effect of multiple rare variants in a region. Our test is able to take into account local ancestry of rare alleles, uncertainties in rare variant imputation when imputed data are included, and the direction of effect (risk or protective) that rare variants exert on phenotypic outcome.

## NOTATION

We assume that admixture has occurred between two ancestral populations, denoted by  $X$  and  $Y$ , in a recently admixed population. We assume a set of markers with known genetic locations are available for ancestry estimation. The genotype of individual  $i$  across  $m$  markers is denoted by  $G_i = (G_{i1}, \dots, G_{im})$  where  $G_{ij} = a_{ij1}/a_{ij2}$  is the genotype of individual  $i$  at marker  $j$ , and  $a_{ij1}$  and  $a_{ij2}$  are the alleles at marker  $j$ . The order of alleles in  $G_{ij}$  may be arbitrary because our test does not rely on phased haplotype information. We denote  $P_{ijk}^s$  to be the probability of the  $k$ th ( $k = 1, 2$ ) allele of SNP  $j$  of individual  $i$  being from ancestral population  $s (= X, Y)$ , given flanking marker genotypes. This probability can be estimated using HapMix [Price et al., 2009], which has an option (DIPOID for HAPMIX MODE) that gives 16 probabilities of all  $4 \times 4$  values of ancestry and genotype for each individual, from which one can infer the corresponding local ancestry for each allele by calculating the corresponding conditional probabilities. Let

$$d_{ijk}^s = \begin{cases} 1, & \text{if } a_{ijk} \text{ is the minor allele of population } s \\ 0, & \text{if } a_{ijk} \text{ is the major allele of population } s \end{cases}$$

When imputed data are included,  $d_{ijk}^s$  can be a fractional number and its value can be obtained based on allele-specific imputation dosage score, which incorporates imputation uncertainty into account. We define the overall minor allele dosage from population  $s$  for individual  $i$  at marker  $j$  as

$$D_{ij}^s = \sum_{k=1}^2 d_{ijk}^s p_{ijk}^s I\{p_{ijk}^s > t\},$$

where  $t$  is a threshold value used to filter out alleles that do not have accurately inferred ancestry and we set the value of  $t$  at 0.9. For a sample of  $n$  individuals, the adjusted minor allele frequency of population  $s$  ancestry at marker  $j$  can be calculated as

$$f_j^s = \frac{\sum_{i=1}^n \left\{ \sum_{k=1}^2 d_{ijk}^s p_{ijk}^s I\{p_{ijk}^s > t\} \right\} + 1}{\sum_{i=1}^n \left\{ \sum_{k=1}^2 p_{ijk}^s I\{p_{ijk}^s > t\} \right\} + 2},$$

where the numerator approximates the total number of minor alleles, and the denominator approximates the total number of alleles including both minor and major alleles, with population  $s$  ancestry, across all study subjects.

### ANCESTRY-BASED WEIGHTED DOSAGE SCORE

The first step in our test is to partition the  $m$  markers in the test region into three categories with respect to their effects on disease risk in population  $s$ . Let  $M_R^s$  represent the set of markers whose rare alleles increase disease risk in population  $s$ ,  $M_P^s$  represent the set of markers whose rare alleles decrease disease risk in population  $s$ , and  $M_N^s$  represents the set of markers having no effects in population  $s$ . We can partition the markers into different categories based on a training set. For example, we can randomly choose 30% of the cases and controls to form the CaSe TRaining set (CSTR) and the ConTrol TRaining set (CTTR), and leave the rest of samples as the testing set. We assign the markers into three groups by the following rules:

$$\left\{ \begin{array}{l} \text{marker } j \in M_R^s \text{ if } f_{CSTR,j}^s - f_{CTTR,j}^s > \mu \sqrt{\frac{f_{CTTR,j}^s(1-f_{CTTR,j}^s)}{2n_{CTTR,j}^s}} \\ \text{marker } j \in M_P^s \text{ if } f_{CSTR,j}^s - f_{CTTR,j}^s < -\mu \sqrt{\frac{f_{CTTR,j}^s(1-f_{CTTR,j}^s)}{2n_{CTTR,j}^s}} \\ \text{marker } j \in M_N^s \text{ if } |f_{CSTR,j}^s - f_{CTTR,j}^s| < \mu \sqrt{\frac{f_{CTTR,j}^s(1-f_{CTTR,j}^s)}{2n_{CTTR,j}^s}} \end{array} \right. ,$$

where  $n_{CTTR,j}^s = \sum_{i=1}^{n_{CTTR}} \left\{ \sum_{k=1}^2 p_{ijk}^s I\{p_{ijk}^s > t\} \right\}$  is the total number of alleles of population  $s$  ancestry in CTTR at marker  $j$ , and  $\mu$  is a constant that is determined by a prespecified type I error rate. For example,  $\mu = 1.28$  (1.64) corresponds to a type I error of 0.2 (0.1). Following Li et al. [2010a], we set  $\mu$  at 1.28 and randomly selected 30% of the samples for training in the analysis. Subsequently, we define the Ancestry-based Weighed Dosage Score (AWDS) for population  $s$  for individual  $i$  as

$$AWDS_i^s = \frac{\sum_{j=1}^m D_{ij}^s \cdot w_j^s (I\{j \in M_R^s\} - I\{j \in M_P^s\})}{\sum_{j=1}^m \sum_{k=1}^2 I\{p_{ijk}^s > t\}},$$

where  $w_j^s = 1 / \sqrt{n_{CTTR,j}^s \cdot f_{CTTR,j}^s \cdot (1 - f_{CTTR,j}^s)}$  is the weight for marker  $j$ , and the denominator standardizes the summation of the WDS by the total number of alleles that are inferred with sufficient accuracy. Here, we adjust for the local ancestry of each SNP by counting the number of rare alleles inherited from population  $s$ .

## TEST OF GENETIC ASSOCIATION WITH RARE VARIANTS

Once the ancestry-based weighted dosage scores are calculated for the cases and controls, we can then test for association with disease status either using a two-sample  $t$ -test, a Wilcoxon rank sum test, or a logistic regression with the AWDS included as an independent variable for each of the two ancestral populations. We choose to use logistic regression to test for association due to its flexibility with covariate adjustment. For individual  $i$ , we denote the disease status as  $Z_i$  ( $0 = \text{control}$ ;  $1 = \text{case}$ ). The logistic regression model for ancestral population is

$$\text{logit}[P(Z_i=1)] = \beta_0^s + \beta_1^s AWDS_i^s.$$

Covariates can be included in the above logit model if they are believed to be important confounders. We are interested in testing  $H_0: \beta_1^s = 0$ . Significant result indicates that the test region is associated with the disease in population  $s$ . The significance level can be assessed by permutations in which the case-control labels are randomly shuffled. However, in this procedure, the test statistic is obtained based on a single split of the data, and splitting samples into training and testing sets may introduce large variation to  $\widehat{\beta}_1^s$ . To reduce the impact of such variation on significance estimation, we resort to use a strategy based on bootstrap confidence interval. Bootstrap confidence interval [DiCiccio and Efron, 1996] has been previously employed for hypothesis testing for survival analysis [Cordell and Carpenter, 2000], for locating genes and recombination events [Dorman et al., 2002; Suthers and Wilson, 1990], and for quantitative trait locus (QTL) and expression QTL (eQTL) analyses [Bennewitz et al., 2002; McRae et al., 2005]. In the bootstrap procedure, we generate many training/testing split datasets, and obtain a  $\widehat{\beta}_1^s$  based on each split of the data. To approximate the distribution of  $\widehat{\beta}_1^s$ , we perform 1,000 random split of the data and this allows us to generate a  $(100 - \alpha)\%$  percentile interval of  $\beta_1^s$ . If the interval does not cover zero, we conclude that  $\beta_1^s$  is significantly different from zero at the  $\alpha$  level in population  $s$ .

To improve the computation efficiency, we adopt an adaptive sampling scheme analogous to sequential testing [Wald, 1947]. For a given dataset, we keep generating samples of different splits of the data until a predefined stopping rule is met or until 1,000 splits are generated and analyzed. Specifically, in the first 20 splits, if at least three positive and three negative  $\widehat{\beta}_1^s$  are observed, we stop the procedure and conclude there is no disease association. Otherwise, we keep resampling until the 100th split, and we stop the procedure if at least seven positive and seven negative  $\widehat{\beta}_1^s$  are observed. If the stopping rule is not met at the 100th split, we continue the resampling procedure until a total of 1,000 splits are obtained

and we then use the corresponding percentile confidence interval to assess the significance  $\widehat{\beta}_1^s$  as aforementioned. The above-described sampling strategy is for significance level of 0.05. Similar strategies can be easily derived for other significance levels.

## SIMULATION OF ADMIXED SAMPLES

**Simulation of Genotypes**—We simulated admixed individuals with African and Caucasian ancestry by constructing their genotypes from AFR (78 YRI + 67 LWK + 24 ASW + 5 PUR) and EUR (90 CEU + 92 TSI + 43 GBR + 36 FIN + 17 MXL + 5 PUR) phased haplotypes based on data released by the 1000 Genomes Project (<http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G-2010-08.html>). The 1000 Genomes data include 348 AFR-phased haplotypes and 566 EUR-phased haplotypes, and the total number of overlapping SNPs between AFR and EUR is 8,952,982. We simulated 20,000 admixed haplotypes following the two-stage procedure described by Price et al. [2009]. In the first stage, we obtained the ancestry state of each marker along the chromosome. Because an admixed individual's genome resembles a mosaic of chromosomal segments or ancestry blocks, and markers within the same block have the same ancestry state, this suggests that we need to partition the chromosome into ancestry blocks and then determine the ancestry state for each block. The breakpoints of the ancestry blocks were determined by recombination events that are sampled from a Bernoulli distribution with probability  $1 - e^{-\lambda g}$ , where  $g$  is the genetic distance (in Morgans) between the previous SNP and the current one and  $\lambda$  ( $\approx 6$ ) is the number of generations since admixture. The genetic distance is approximated by assuming that 1 Mb  $\approx$  1 centi Morgan. For each block, we sampled AFR ancestry with probability  $\alpha$  and EUR ancestry with probability  $1 - \alpha$  and markers within the same block were assigned with the same ancestry state. We determined the value of  $\alpha$  from a beta distribution with mean 0.80 and standard deviation 0.10 (typical for African Americans [Smith et al. 2004]). In the second stage, we filled in the genotypes for each admixed haplotype. For any given ancestry block, we randomly sampled a haplotype from the haplotype pool of the same ancestry as the given block and assigned the sequence of the sampled haplotype in that block to the admixed haplotype. We repeated this procedure and generated a pool of 20,000 admixed haplotypes for each of the 22 autosomal chromosomes. Pairs of admixed haplotypes were then merged to create 10,000 diploid admixed individuals.

**Assignment of Disease Status**—We first partitioned the genome into 44,620 nonoverlapping segments with each segment containing 200 SNPs. We then randomly selected  $c$  SNPs (minor allele frequencies (MAFs) between 0.001 and 0.05) as causal sites. We denote the population attributable risks (PARs) for AFR and EUR as  $PAR_{AFR}$  and  $PAR_{EUR}$ , respectively. For a given individual, the genotype relative risk (GRR) of population  $s$  (AFR or EUR) descent at site  $j$  is defined as

$$GRR_j^s = \left[ \frac{PAR_j^s}{(1 - PAR_j^s) f_j^s} + 1 \right]^{(-1)^{\{\xi_j^s=1\}}},$$

where  $\xi_j^s=1$  indicates that the rare allele at marker  $j$  decreases disease risk in population  $s$ . The values of GRR for noncausal sites were set as 1.

Our GRR calculation is similar to Madsen and Browning [2009] and Li et al. [2010a]. The difference is that we differentiate the risk of rare variants based on their ancestry state. In other words, two individuals carrying the same rare allele may have different risk if the

alleles are from different ancestral populations. Let  $I_{ijk}$  ( $1 = \text{AFR}$ ;  $0 = \text{EUR}$ ) denote the ancestry state for allele  $k$  of individual  $i$  at marker  $j$ . The disease status of individual  $i$  given its population-specific dosages,  $d_{ijk}^{\text{AFR}}$ ,  $d_{ijk}^{\text{EUR}}$ , and ancestry states was assigned according to

$$P(Z_i=1 \mid d_{ijk}^{\text{AFR}}, d_{ijk}^{\text{EUR}}, I_{ijk}) = b_0 \times \prod_{j=1}^{200} \prod_{k=1}^2 (GRR_j^{\text{AFR}})^{I_{ijk}^{\text{AFR}}=1, I_{ijk}=1} \times (GRR_j^{\text{EUR}})^{I_{ijk}^{\text{EUR}}=1, I_{ijk}=0},$$

where  $b_0$  is the baseline penetrance and was fixed at 10%. Sampling was repeated until the desired number of cases and controls was reached.

To evaluate type I error, we randomly selected 1,000 individuals from the pool as cases and 1,000 individuals as controls. Because there is no systematic difference among cases and controls, these selected individuals have similar patterns in terms of their global ancestry. This allows us to evaluate whether substantial local ancestry differences exist when the global ancestry patterns are similar. To evaluate power, we considered two scenarios for causal variants: (1) risk variants only, i.e., all variants increase disease risk, and (2) mixture of risk and protective variants. We assumed the causal variants have different risks in EUR and AFR, and for each ancestral population, the PARs for all causal variants are the same. For the risk-only scenario, we consider both one-sided disease model and two-sided disease model with respect to whether causal variants are from only one ancestral population or both. For the one-sided model, we selected 20 causal variants and the PARs were set at 0.002, 0.003, 0.004, and 0.005 for AFR, and at 0.008, 0.012, 0.016, and 0.02 for EUR. Of note, the power of detecting AFR-specific effect is higher in an 80% AFR and 20% EUR admixed population. The reason is that, the effective sample size in AFR, defined as the number of alleles of AFR descent, is greater than that in EUR. As such, we used a higher PAR for EUR because of its smaller effective sample size. In the two-sided model, 20 causal variants are selected from each ancestral population (a total of 40 causal variants) with different PARs. Four scenarios were simulated,  $\text{PAR}_{\text{AFR}} = 0.003$  with  $\text{PAR}_{\text{EUR}} = 0.004$ ,  $\text{PAR}_{\text{AFR}} = 0.003$  with  $\text{PAR}_{\text{EUR}} = 0.008$ ,  $\text{PAR}_{\text{AFR}} = 0.001$  with  $\text{PAR}_{\text{EUR}} = 0.012$ , and  $\text{PAR}_{\text{AFR}} = 0.002$  with  $\text{PAR}_{\text{EUR}} = 0.012$ . For the mixture scenario, we assumed there are 20 causal variants in the region, and the number of protective variants range from 5 to 10 to 20. Without loss of generality, we assume causal variants of one ancestral origin contribute to disease risk.

## RESULTS

### COMPARISON OF TYPE I ERROR

We are interested in assessing the impact of local ancestry difference on type I errors of our method along with existing methods. We generated 10,000 diploid admixed individuals as previously described. We randomly selected 1,000 individuals as cases and 1,000 individuals as controls. For each selected individual, we partitioned the genome into 44,620 segments of 200 markers, and estimated the local ancestry of each segment as the average of ancestry proportions across all markers within the segment. The 44,620 segments were then ranked by the mean difference of local ancestry between cases and controls. Five hundred segments were retained each from the bottom, the middle and the top of the ranked list. The 1,500 selected segments were then analyzed by both the WDS test [Li et al., 2010a] and the AWDS test.

To make a fair comparison, in addition to the original WDS test, we also considered modified versions of WDS by either adjusting for global PCs obtained from the linkage disequilibrium (LD) pruned marker list across the entire genome or adjusting for local

ancestry based on local PCs obtained from the pruned markers in the 10 Mb downstream and 10 Mb upstream of the test region per Qin et al. [2010]. We used PLINK to perform LD-based marker pruning with the pairwise option of window size 50, increment 5, and LD threshold 0.1. The number of SNPs used for global PCs is 52,822. We denoted these two modified versions of WDS test as  $WDS_{GPC}$  and  $WDS_{LPC}$ , respectively. We repeated the simulation procedure 20 times and obtained a total of 10,000 segments for each of the three local ancestry categories.

We assessed the type I error rates in terms of local ancestry disparity between cases and controls. The histogram of local ancestry differences from one simulation run is summarized in Figure 1. We considered three groups of 200-SNP segments: category I (local ancestry difference 0.001 ~ 0.005), category II (0.01 ~ 0.02), and category III (0.03 ~ 0.05). In each category, we analyzed 10,000 segments using AWDS, the original WDS, and WDS with adjustment of either global ancestry or local ancestry. The type I error rates are summarized in Table I. Of note, the type I error rates were evaluated for each ancestral population separately. As expected, both  $AWDS_{AFR}$  and  $AWDS_{EUR}$  have controlled type I error rates under all scenarios, whereas all WDS-type tests have inflated type I error rates when local ancestry difference is large regardless whether PC adjustment was included or not. Our results also indicate that the type I error rate of WDS was improved only marginally by adjusting for global PCs (from 0.173 to 0.157), but improved noticeably by local PC adjustment (from 0.173 to 0.069). These results strongly suggest that the commonly used global adjustment procedures are ineffective when local ancestry differs from the global ancestry proportion. Additionally, simple adjustment with local PCs cannot guarantee valid association result despite its noticeable improvement. We note that our purpose of the analysis is to show that for regions where local ancestry differs noticeably between cases and controls (e.g., the top 500 segments), the type I error rates can be very high if local ancestry is not appropriately adjusted, although the genome-wide type I error rate is still close to the nominal level.

## COMPARISON OF POWER

First, we present the risk-only one-sided disease model in Figure 2. We used larger PAR values for EUR to make the EUR-side tests have comparable power with the AFR-side tests. In the range of PARs we considered, for each ancestry population (AFR or EUR), the corresponding AWDS test consistently has higher power than WDS-type tests including WDS with adjustment for global PCs and local PCs. For example, when all causal variants come from AFR,  $AWDS_{AFR}$  is at least 10% more powerful than WDS-type tests. Similarly,  $AWDS_{EUR}$  is consistently more powerful than WDS-type tests when all causal variants are from EUR. In addition to the one-sided disease model, we also simulated four two-sided scenarios in which rare variants of both ancestral origins contributed to disease risk (Figure 3). In the two-sided model, WDS-type tests still have less power than AWDS for the side with larger contribution to disease risk but have more power than AWDS for the other side. This pattern is expected because WDS is designed for the collective effect of causal variants ignoring their ancestry although AWDS devises separate tests for different ancestral populations.

When both risk variants and protective variants are present, we compared the power of AWDS with WDS-type tests in a range of PAR values as well as with different number of protective variants. For simplicity, we only chose moderate PARs for AFR (0.003, 0.004, and 0.005) and for EUR (0.012, 0.016, 0.02). With the number of causal variants fixed at 20 and the number of protective variants ranging from 5 to 10 to 20, we considered three PAR values for risk variants along with three PAR values for protective variants. We summarize the power of AWDS and WDS-type tests for AFR (Figure 4A–C) and EUR (Figure 4D–F), separately. Although the power of all tests is negatively correlated with the number of



protective variants, for AFR, the corresponding AWDS test still outperforms all WDS-type tests despite that the latter tests have inflated type I error rates.

## UNCERTAINTY IN ANCESTRY PROBABILITY ESTIMATION

In the previous analyses, we have assumed that the ancestry states in the test regions are known. In real studies, the ancestry states have to be estimated using genetic marker data. To evaluate the impact of estimation uncertainty on our method, we assessed the performance of the tests when ancestry states are estimated. We chose to use HAPMIX to estimate local ancestries as this program represents one of the most accurate algorithms for allele-specific ancestry estimation [Price et al., 2009]. We tested 200 individuals across 5,000 consecutive SNPs on chromosome 22 using all EUR and AFR haplotypes from the 1000 Genome Project as reference. As expected, HAPMIX yielded highly accurate ancestry estimates with more than 98% of the inferred ancestry states being identical to the true states. Ideally, we would like to use HAPMIX to estimate the local ancestry for all of the 10,000 simulated genomes. However, the computational time would be tremendous. For the testing dataset we considered, it took about 55 hr on a four-core AMD Opteron™ Processor 6212 (AMD, Sunnyvale, CA) with 4G memory. It is computationally prohibitive to run HAPMIX on a genome-wide scale in a large number of simulations. To evaluate the performance of our tests when estimation uncertainty exist although being practical, we considered an error model derived based on the empirical estimation results obtained from HAPMIX. From the HAPMIX estimates on the testing data, we observed that (1) the local ancestry states for genetic markers that are close to a recombination breakpoint between different ancestral populations are often inferred with greater uncertainty than those far away from such recombination breakpoints; (2) the uncertainty estimates in the local ancestries of the two haplotypes from the same individual are nearly independent; (3) a small percentage (~1%) of the haplotypes without recombinations between different ancestral populations exhibit above-average uncertainty in their estimated local ancestries. Based on these observations, we derived an error model for ancestry probability (Table II), which induces similar patterns of uncertainties for local ancestry estimates as HAPMIX.

We introduced random errors in local ancestry estimates in the simulated datasets following the error model in Table II, and reevaluated the type I error rates and power for two risk-variant only one-sided disease models. As shown in Table III, the type I error rates of AWDS are still under control when ancestry probabilities instead of the true ancestral states were used. The power is decreased slightly for AFR (from 0.597 to 0.551 for  $PAR_{AFR} = 0.003$ ) but moderately for EUR (from 0.791 to 0.702 for  $PAR_{EUR} = 0.012$ ). The noticeable power loss for EUR is due to the fact that the effective sample size for EUR is much smaller than AFR and thus local ancestries for EUR haplotypes are generally estimated with greater uncertainty than haplotypes with AFR ancestry. Because of the threshold value  $t$  we used in computing the dosage scores, many haplotypes with EUR ancestries were excluded from the analysis, leading to reduced sample size and thus loss of power.

## APPLICATION TO THE LDL DATASET

We applied our approach to a combined dataset on LDL-C from Candidate-gene Association Resource (CARE) and Women's Health Initiative (WHI) [Reiner et al., 2011]. All samples from the two cohorts were genotyped using Affymetrix 6.0 SNP array (Affymetrix, Santa Clara, CA). Previous GWAS analysis revealed strong association between LDL and a common SNP rs1367117 in *APOB* [Teslovich et al., 2010;  $P$ -value =  $4 \times 10^{-114}$ ], a gene located on chromosome 2. To evaluate the performance of AWDS in real data, we imputed rare variants in this gene using haplotypes from the 1000 Genomes Project Consortium [2010] as reference and then tested for association between *APOB* and LDL. We assigned individuals with LDL > 160 mg/dL as cases ( $n = 1,319$ ) and those with LDL < 129 mg/dL as

controls ( $n = 3,532$ ). We retained 66 SNPs with imputed dosage  $r^2 > 0.9$  in the analysis, including 35 common SNPs ( $MAF > 0.05$ ) and 31 less common SNPs ( $MAF < 0.05$ ). The allele-specific population ancestry in the *APOB* region was inferred using HAPMIX for all cases and controls. The average African ancestries for cases and controls are 0.696 (SD = 0.098) and 0.712 (SD = 0.059), respectively. A total of 1,256 cases and 3,413 controls were inferred with accurate ancestry (defined as  $p_{ijk}^{AFR} > 0.9$  or  $p_{ijk}^{AFR} < 0.1$  for at least one-third of the heterozygous sites in an individual). We included age, gender, BMI, and study site as covariates. We only considered those SNPs with  $MAF < 0.05$  in our analysis. As shown in Table IV, AWDS revealed marginal association for the African side of the test, suggesting that there might be residual association explained by rare variants in *APOB* for the surveyed samples.

## DISCUSSION

We have shown that for samples collected from recently admixed populations, there exist noticeable local ancestry differences among study subjects even when their global ancestry patterns are similar. Such local ancestry difference can either lead to spurious association or diminished power for the analysis of rare variants if population stratification is not appropriately controlled. Through simulations with sequence data, we showed that adjusting for global ancestry was only marginally better than a naïve test without ancestry adjustment, and even a small disparity of local ancestry between cases and controls could lead to severely inflated type I errors. The simulated situation we considered is an ideal admixture scenario, where both the switch of ancestry states due to recombination and the assignment of ancestral haplotypes are random. However, factors that are not yet modeled here, such as demographic histories, subpopulations within a meta-population and natural selection on particular gene regions, may lead to additional local ancestry disparity and thus the effect of local ancestry difference induced population stratification may be even stronger in real studies. Our findings underscore the importance of local ancestry adjustment [Qin et al., 2010; Wang et al., 2011], and reinforce that population stratification in rare variants is not necessarily corrected by existing methods [Mathieson and McVean, 2012]. When analyzing a real dataset on LDL, we were able to confirm the original finding [Teslovich et al., 2010] and identified the association signal mainly due to common variants of African ancestry.

To incorporate local ancestry in the analysis, we need to overcome a series of technical challenges including but are not limited to how to efficiently estimate local ancestry, how to construct test statistics using local ancestry and whether tests for different ancestral populations should be combined. Our AWDS test relies on accurate estimation of local ancestry, in particular, the ancestry state of each allele. As shown in our testing dataset, with the coverage of whole genome sequencing data, programs such as HAPMIX can produce highly accurate local ancestry estimates. However, we note that computational challenges persist. If one uses a large number of haplotypes from the reference panel, computation may take months or years. For sequence data, where the MAFs of many SNPs may be much smaller, currently available reference panels may not contain some of the rare variants observed in admixed populations. In light of these issues, one may carefully select a subset of SNPs or ancestral informative markers covering the whole genome and estimate the ancestry states of the preselected SNPs using HAPMIX. When there is a switch of ancestry states between two preselected SNPs, one can analyze the original set of SNPs in that region and pinpoint the switch point of ancestry states. For consecutive SNPs with no ancestry state switch, one can interpolate these segments by the inferred ancestry states of the flanking SNPs.

We included local ancestry information in AWDS to standardize ancestry-specific allele dosages. Recall that we have filtered out low ancestry probability by the cutoff  $t$  in

frequency calculation and dosage calculation. In practice, one can adjust  $t$  according to the inferred ancestry probabilities. Extensive simulations with true ancestry states indicate that using AWDS, we have improved power over the original WDS test, WDS<sub>GPC</sub>, and WDS<sub>LPC</sub>. When there is uncertainty in ancestry probability estimation, the type I error rate is roughly unchanged but the power decreases marginally for the AFR-side disease model and decreases moderately for the EUR-side disease model. We interpret the decrease of power as the result of introducing uncertainty in ancestry probability estimation.

We calculated separate AWDS for EUR and AFR to provide population-specific tests. Such separate tests allow us to differentiate the effect of rare variants based on their ancestry origin so that signals of true risk effect from one ancestral population background will not be diluted by the other ancestral population. Moreover, these tests allow a cross-ethnicity replication because evidence of disease association in the two ancestry populations can be directly compared [Risch and Tang, 2006]. Additionally, by computing separate test statistic for each ancestral population, we can easily accommodate admixed populations with more than two-way admixture. For example, for Hispanic Americans, who have ancestry from Africans, Native Americans, and Caucasians, we can include three ancestral populations in our procedure and test for association in each of the three populations separately. We recognize that AWDS tests for different ancestral populations may be combined by using methods such as Fisher's combined probability test. However, we note that conducting such an test requires accurate estimation of  $P$ -values and the adaptive  $P$ -value estimation scheme employed in the AWDS procedure may not be suitable for this purpose. Further work is warranted to find heuristics for fast calculation of  $P$ -values and combining them for inference.

Our method was developed for the analysis of sequence data, but we showed that the proposed framework could incorporate imputed data as well. For example, widely used genotype imputation methods can be readily modified to generate allele-specific imputation dosage scores (i.e., expected number of the minor or major allele). Although by default the allele dosages are aggregated over all ancestral/reference populations, most of the methods, relying on copying alleles from similar haplotypes, can easily generate population-specific allele dosages by recording the population label of the copied haplotypes, as implemented in our MaCH-Admix [Liu et al., 2012]. AWDS tests can then be performed based on these population-specific allele dosages.

In summary, we have proposed a local ancestry-based rare variant association test for admixed populations. Our method can be applied to both whole-genome sequencing data as well as imputed data obtained from GWAS. Our test adjusts local ancestry in burden-based rare variant tests and is able to differentiate risk variants and protective variants. Simulations demonstrate the proposed test is protected against population stratification induced by local ancestry difference and is more powerful than available methods.

## Acknowledgments

This research was supported by NIH grants R01HG004517, R01HG005854 (to M.L.), and R01HG006292 and R01HG006703 (to Y.L.).

## REFERENCES

- 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
- Bennewitz J, Reinsch N, Kalm E. Improved confidence intervals in quantitative trait loci mapping by permutation bootstrapping. *Genetics*. 2002; 160:1673–1686. [PubMed: 11973320]

- Cordell HJ, Carpenter JR. Bootstrap confidence intervals for relative risk parameters in affected-sib-pair data. *Genet Epidemiol.* 2000; 18:157–172. [PubMed: 10642428]
- DiCiccio T, Efron B. Bootstrap confidence intervals. *Statist Sci.* 1996; 11:189–228.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol.* 2010; 8:e1000294. [PubMed: 20126254]
- Dorman KS, Kaplan AH, Sinsheimer JS. Bootstrap confidence levels for HIV-1 recombinants. *J Mol Evol.* 2002; 54:200–209. [PubMed: 11821913]
- Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case-control studies. *Am J Hum Genet.* 2007; 80:921–930. [PubMed: 17436246]
- Hershberger R, Norton N, Morales A, Li D, Siegfried J, Gonzalez-Quintana J. Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1, and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet.* 2010; 3:155–161. [PubMed: 20215591]
- Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genet.* 2011; 7:e1001289. [PubMed: 21304886]
- Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008; 83:311–321. [PubMed: 18691683]
- Li M, Reilly MP, Rader DJ, Wang L-S. Correcting population stratification in genetic association studies using a phylogenetic approach. *Bioinformatics.* 2010a; 26:798–806. [PubMed: 20097913]
- Li Y, Byrnes AE, Li M. To identify associations with rare variants, just WHaIT: Weighted haplotype and imputation-based tests. *Am J Hum Genet.* 2010b; 87:728–735. [PubMed: 21055717]
- Liu EY, Li M, Wang W, Liu Y. MaCH-Admix: genotype imputation for admixed populations. *Genet Epidemiol.* 12-0124.R1. 2012 Manuscript submitted for publication.
- Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009; 5:e1000384. [PubMed: 19214210]
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature.* 2009; 461:747–753. [PubMed: 19812666]
- Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet.* 2012; 44:243–246. [PubMed: 22306651]
- McRae AF, Bishop SC, Walling GA, Wilson AD, Visscher PM. Mapping of multiple quantitative trait loci for growth and carcass traits in a complex commercial sheep pedigree. *Anim Sci.* 2005; 80:135–141.
- Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007; 615:28–56. [PubMed: 17101154]
- Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melandar M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011; 7:e1001322. [PubMed: 21408211]
- 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. [PubMed: 16862161]
- Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 2009; 5:e1000519. [PubMed: 19543370]
- Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exonresequencing studies. *Am J Hum Genet.* 2010; 86:832–838. [PubMed: 20471002]
- Qin H, Morris N, Kang SJ, Li M, Tayo B, Lyon H, Hirschhorn J, Cooper RS, Zhu X. Integrating local population structure for fine mapping in genome-wide association studies. *Bioinformatics.* 2010; 26:2961–2968. [PubMed: 20889494]

- Reiner AP, Lettre G, Nalls MA, Ganesh SK, Mathias R, Austin MA, Dean E, Arepalli S, Britton A, Chen Z, Couper D, Curb JD, Eaton CB, Fornage M, Grant SF, Harris TB, Hernandez D, Kamatini N, Keating BJ, Kubo M, LaCroix A, Lange LA, Liu S, Lohman K, Meng Y, Mohler ER 3rd, Musani S, Nakamura Y, O'Donnell CJ, Okada Y, Palmer CD, Papanicolaou GJ, Patel KV, Singleton AB, Takahashi A, Tang H, Taylor HA Jr, Taylor K, Thomson C, Yanek LR, Yang L, Ziv E, Zonderman AB, Folsom AR, Evans MK, Liu Y, Becker DM, Snively BM, Wilson JG. Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). *PLoS Genet.* 2011; 7:e1002108. [PubMed: 21738479]
- Risch N, Tang H. Whole genome association studies in admixed populations. *Am J Hum Genet.* 2006; 79:254.
- Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, De The G, Essex M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich D. A high-density admixture map for gene discovery in African Americans. *Am J Hum Genet.* 2004; 74:1001–1013. [PubMed: 15088270]
- Suthers GK, Wilson SR. Genetic counseling in rare syndromes: a resampling method for determining an approximate confidence interval for gene location with linkage data from a single pedigree. *Am J Hum Genet.* 1990; 47:53–61. [PubMed: 2349950]
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllenstein U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010; 466:707–713. [PubMed: 20686565]
- Wald, A. *Sequential Analysis.* John Wiley and Sons; New York: 1947.
- Wang X, Zhu X, Qing H, Cooper R, Ewens W, Li C, Li M. Adjustment for local ancestry in genetic association analysis of admixed populations. *Bioinformatics.* 2011; 27:670–677. [PubMed: 21169375]
- Wessel J, McDonald SM, Hinds DA, Stokowski RP, Javitz HS, Kennemer M, Krasnow R, Dirks W, Hardin J, Pitts SJ, Michel M, Jack L, Ballinger DG, McClure JB, Swan GE, Bergen AW. Resequencing of nicotinic acetylcholine receptor genes and association of common and rare

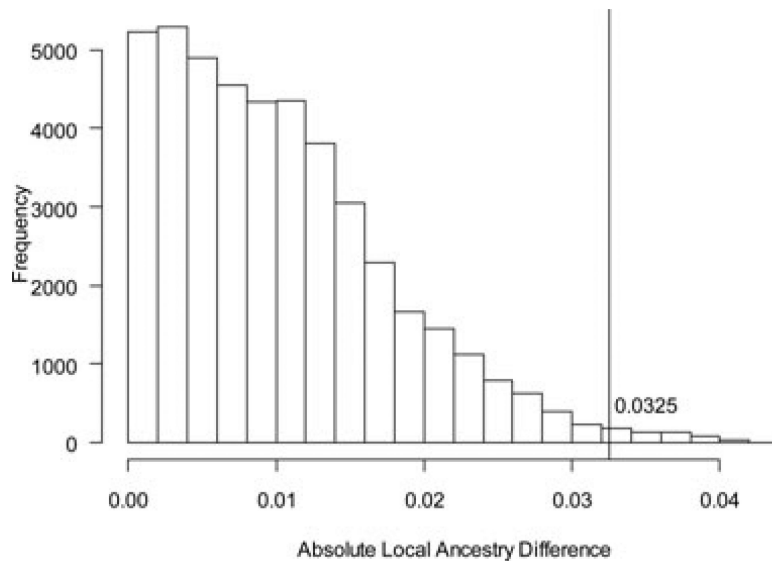
variants with nicotine dependence. *Neuropsychopharmacology*. 2010; 35:2392–2402. [PubMed: 20736995]

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare variant association testing for sequencing data with the sequence kernel association test (SKAT). *Am J Hum Genet*. 2011; 89:82–93. [PubMed: 21737059]

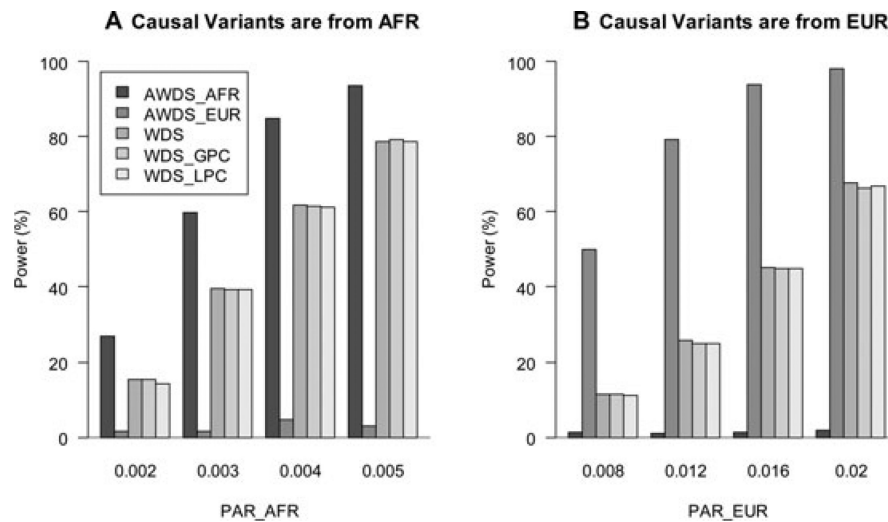
Yi N, Zhi D. Bayesian analysis of rare variants in genetic association studies. *Genet Epidemiol*. 2011; 35:57–69. [PubMed: 21181897]

Yi N, Liu N, Zhi D, Li J. Hierarchical generalized linear models for multiple groups of rare and common variants: jointly estimating group and individual-variant effects. *PLoS Genet*. 2011; 7:e1002382. [PubMed: 22144906]

Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zöllner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am J Hum Genet*. 2010; 87:604–617. [PubMed: 21070896]

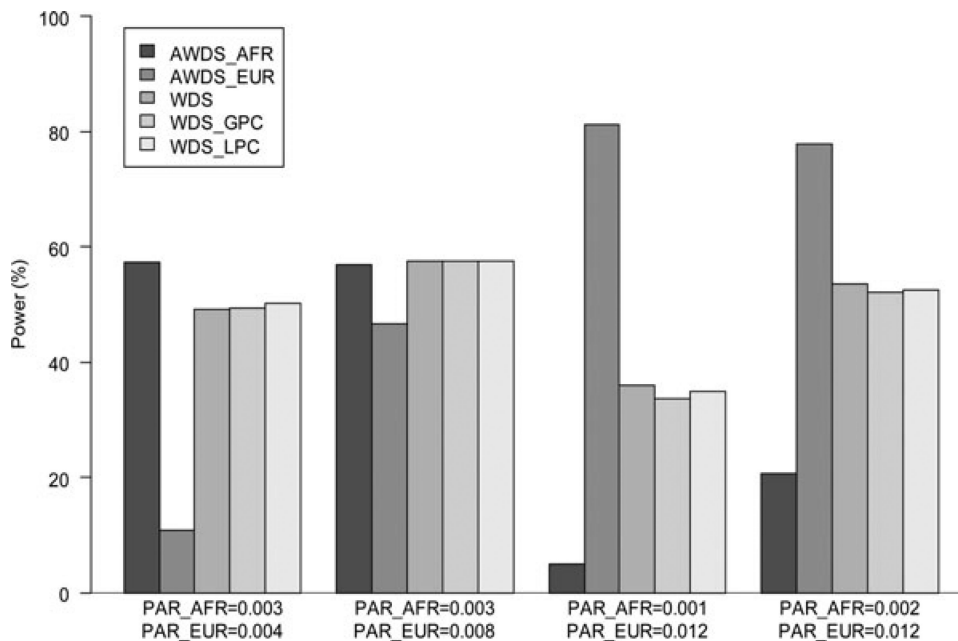


**Fig. 1.** Local ancestry differences between randomly chosen cases and controls in 44,620 200-SNP bins across the genome based on one simulation run. The highest 500 values range between 0.0325 to 0.0421.

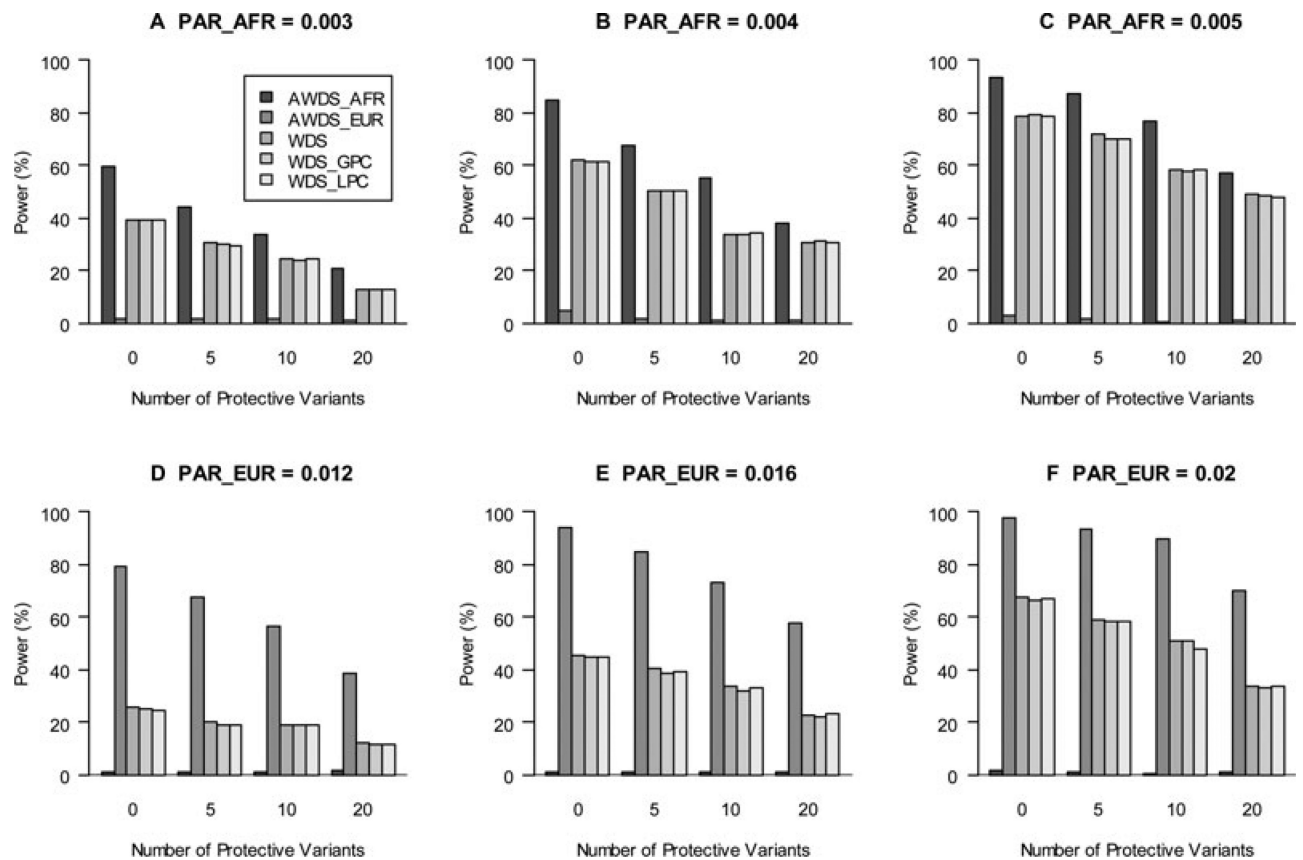


**Fig. 2.** Comparison of power under the risk-variant only one-sided disease model. (A) Power of  $AWDS_{AFR}$ ,  $AWDS_{EUR}$ , WDS,  $WDS_{GPC}$ , and  $WDS_{LPC}$  under risk-variant only model assuming all causal variants come from AFR. Note that the bars for  $AWDS_{EUR}$  are in fact the type I errors for  $AWDS_{EUR}$  under AFR-side disease model. (B) Power of  $AWDS_{AFR}$ ,  $AWDS_{EUR}$ , WDS,  $WDS_{GPC}$ , and  $WDS_{LPC}$  under risk-variant only model assuming all causal variants come from EUR. Note that the bars for  $AWDS_{AFR}$  are in fact the type I errors for  $AWDS_{AFR}$  under EUR-side disease model.





**Fig. 3.** Comparison of power under the risk-variant only two-sided disease model. Power of  $AWDS_{AFR}$ ,  $AWDS_{EUR}$ ,  $WDS$ ,  $WDS_{GPC}$ , and  $WDS_{LPC}$  under risk-variant only model assuming a total of 40 causal variants coming from both AFR and EUR with 20 causal variants each.



**Fig. 4.** Comparison of power under the mixture of risk and protective variants model. The number of causal variants is fixed at 20 for each scenario and we assume causal variants are all from AFR ancestry background (A–C) and all from EUR (D–F). So the bars in (A–C) for AWDS\_EUR are in fact type I errors; the bars in (D–F) for AWDS\_AFR are in fact type I errors.

**TABLE I**

Type I error rates of AWDS and WDS with and without adjustment of population stratification

Category (mean local ancestry difference between cases and controls)	Type I error				
	AWDS <sub>AFR</sub>	AWDS <sub>EUR</sub>	WDS	WDS <sub>GPC</sub>	WDS <sub>LPC</sub>
I (0.001 ~ 0.003)	0.013	0.011	0.016	0.018	0.023
II (0.01 ~ 0.02)	0.014	0.014	0.021	0.022	0.025
III (0.03 ~ 0.05)	0.016	0.011	0.173	0.157	0.069

Significance was assessed at 5% level. AWDS<sub>AFR</sub> and AWDS<sub>EUR</sub> are for separate AWDS tests for AFR and EUR, respectively. WDS is the original WDS test from WHaIT. WDS<sub>GPC</sub> and WDS<sub>LPC</sub> are the WDS tests with adjustment of population stratification using global PCs (based on 52,822 LD-pruned SNPs across the genome) and using local PCs (LD-pruned SNPs from 10 Mb upstream and downstream of the test region).

TABLE II

## Uncertainty model for ancestry probability estimation

	True ancestry probability	Probability with uncertainty
An adjacent recombination event between different ancestral populations <i>or</i> 1% randomly chosen haplotypes without such events	$P(I = \text{AFR}) = 0$	$\epsilon_2$
	$P(I = \text{EUR}) = 1$	$1 - \epsilon_2$
	$P(I = \text{AFR}) = 1$	$1 - \epsilon_2$
	$P(I = \text{EUR}) = 0$	$\epsilon_2$
99% of the haplotypes without nearby recombination events between ancestral populations	$P(I = \text{AFR}) = 0$	$\epsilon_1$
	$P(I = \text{EUR}) = 1$	$1 - \epsilon_1$
	$P(I = \text{AFR}) = 1$	$1 - \epsilon_1$
	$P(I = \text{EUR}) = 0$	$\epsilon_1$

$\epsilon_1$  is the uncertainty parameter when there is no recombination between different ancestral populations and is generated from uniform(0,0.01).  $\epsilon_2$  is the uncertainty parameter when there is an adjacent recombination event between different ancestral populations or the admixed haplotype is chosen as 1% of the haplotypes having high uncertainty in ancestry probability.  $\epsilon_2$  is generated from uniform(0.3, 0.7).  $I$  is the ancestry state of an allele.

TABLE III

Type I error and power of AWDS tests when true ancestry is known and when ancestry is estimated with uncertainty

	True ancestry			Estimated ancestry		
	PAR <sub>AFR</sub>	PAR <sub>EUR</sub>	AWDS <sub>AFR</sub>	AWDS <sub>EUR</sub>	AWDS <sub>AFR</sub>	AWDS <sub>EUR</sub>
Type I	0	0	0.013	0.011	0.017	0.013
Category I						
Category II	0	0	0.014	0.014	0.015	0.011
Category III	0	0	0.016	0.011	0.023	0.019
Power	0.003	0	0.597	0.016	0.551	0.014
	0	0.012	0.011	0.791	0.024	0.702

Uncertainty is introduced based on the model described in Table II.

TABLE IV

Analysis results on APOB

	AWDS <sub>AFR</sub>	AWDS <sub>EUR</sub>	WDS	WDS <sub>GPC</sub>	WDS <sub>LPC</sub>
31 less common SNPs (MAF < 0.05)	0.062	0.131	0.053	0.066	0.075

*P*-values are based on 10,000 bootstrap samples.