A peer-reviewed version of this preprint was published in PeerJ on 4 November 2014.

View the peer-reviewed version (peerj.com/articles/660), which is the preferred citable publication unless you specifically need to cite this preprint.

Zhang X, Mu W, Liu C, Zhang W. 2014. Ancestry-informative markers for African Americans based on the Affymetrix Pan-African genotyping array. PeerJ 2:e660 https://doi.org/10.7717/peerj.660

Ancestry-informative markers for African Americans based on the Affymetrix Pan-African genotyping array

Genetic admixture has been utilized as a tool for identifying loci associated with complex traits and diseases in recently admixed populations such as African Americans. In particular, admixture mapping is an efficient approach to identifying genetic basis for those complex diseases with substantial racial or ethnic disparities. Though current advances in admixture mapping algorithms may utilize the entire panel of SNPs, providing ancestry-informative markers (AIMs) that can differentiate parental populations and estimate ancestry proportions in an admixed population may particularly benefit admixture mapping in studies of limited samples, help identify unsuitable individuals (e.g., through genotyping the most informative ancestry markers) before starting large genome-wide association studies (GWAS), or guide larger scale targeted deep re-sequencing for determining specific disease-causing variants. Defining panels of AIMs based on commercial, high-throughput genotyping platforms will facilitate the utilization of these platforms for simultaneous admixture mapping of complex traits and diseases, in addition to conventional GWAS. Here, we describe AIMs detected based on the Shannon Information Content (SIC) or F_{st} for African Americans with genome-wide coverage that were selected from ~2.3 million single nucleotide polymorphisms (SNPs) covered by the Affymetrix Axiom Pan-African array, a newly developed genotyping platform optimized for individuals of African ancestry.

Ancestry-Informative Markers for African Americans Based on the Affymetrix Pan-African Genotyping Array

Xu Zhang[1,2], Wenbo Mu[3], Cong Liu[3], and Wei Zhang[1,3]

[1]The Affiliated Hospital of Medical School, Ningbo University, Ningbo, Zhejiang Province, China
[2]Section of Hematology/Oncology, Department of Medicine; [3]Department of Pediatrics, University of Illinois, Chicago, Illinois, USA

***Correspondence to:**
Wei Zhang, Ph.D., 840 S. Wood St., 1200 CSB (M/C 856), Chicago, IL 60612, USA; Tel: +1 (312) 413-2024; E-mail: weizhang.chicago@gmail.com

Running Title: AIMs for African Americans

Abstract

Genetic admixture has been utilized as a tool for identifying loci associated with complex traits and diseases in recently admixed populations such as African Americans. In particular, admixture mapping is an efficient approach to identifying genetic basis for those complex diseases with substantial racial or ethnic disparities. Though current advances in admixture mapping algorithms may utilize the entire panel of SNPs, providing ancestry-informative markers (AIMs) that can differentiate parental populations and estimate ancestry proportions in an admixed population may particularly benefit admixture mapping in studies of limited samples, help identify unsuitable individuals (e.g., through genotyping the most informative ancestry markers) before starting large genome-wide association studies (GWAS), or guide larger scale targeted deep re-sequencing for determining specific disease-causing variants. Defining panels of AIMs based on commercial, high-throughput genotyping platforms will facilitate the utilization of these platforms for simultaneous admixture mapping of complex traits and diseases, in addition to conventional GWAS. Here, we describe AIMs detected based on the Shannon Information Content (SIC) or F_{st} for African Americans with genome-wide coverage that were selected from ~2.3 million single nucleotide polymorphisms (SNPs) covered by the Affymetrix Axiom Pan-African array, a newly developed genotyping platform optimized for individuals of African ancestry.

Introduction

High throughput genotyping arrays have facilitated genome-wide association studies (GWAS) on complex traits (Hindorff et al. 2009) including risks for common, complex diseases

and drug response. In contrast to a conventional GWAS in a homogeneous parental populations (e.g., Caucasians), admixture mapping or mapping by admixture linkage disequilibrium (MALD) has begun to be demonstrated as a powerful tool for identifying disease-causing genetic variants in recently admixed populations, such as African Americans that have both West African and European American ancestry (McKeigue 2005). For example, recent admixture mapping studies have identified loci associated with disease risks such as prostate cancer (Ricks-Santi et al. 2012), lung cancer (Schwartz et al. 2011), and traits like blood pressure/obesity (Shetty et al. 2012) in African Americans. Admixture mapping assumes that near a disease causing genetic variant there will be enhanced ancestry from the population that has greater risk of getting the disease (Patterson et al. 2004). Therefore, by calculating the proportion of ancestry along the genome, one could use that information to identify disease causing loci in an admixed population with low resolution. Subsequent fine mapping restricted to the identified genomic regions may greatly increase the power of the study.

It has been demonstrated that 1,500–2,500 ancestry-informative markers (AIMs) with genome-wide coverage would be sufficient (Winkler et al. 2010) to identify the ancestral chromosome segments for recently admixed populations. To leverage on the power of admixture mapping in African American for identifying disease causing genetic variants that may explain health disparities between populations, panels of AIMs have been proposed for commercially-available high throughput genotyping arrays including the Affymetrix SNP 6.0 and Illumina 1M (Chen et al. 2010; Tandon et al. 2011). These genotyping arrays however are likely biased to genetic variations detected from Caucasian samples. The Affymetrix Pan-African array, which interrogates approximately 2.3 million SNPs, was designed for a much greater coverage of genetic variations in African individuals. A panel of AIMs based on the Pan-African array may enhance the distinguishing of parental populations as well as improve genome coverage. Recent

advances in statistical genetics have begun to allow admixture mapping utilizing the entire panel of genotyped SNPs (Baran et al. 2012; Churchhouse & Marchini 2013; Maples et al. 2013), however, we reasoned that providing a panel of AIMs may particularly benefit studies of a limited sample size, help identify unsuitable individuals by genotyping the most informative markers before starting a large GWAS, or guide larger scale targeted re-sequencing projects to pinpoint causal variants. We describe here AIMs identified for the Affymetrix Pan-African array based on Shannon Information Content (SIC) or F_{st} for African Americans using the 1000 Genomes Project (Abecasis et al. 2010) data as references for parental populations.

Materials and Methods

SNPs covered on the Pan-African array

The Affymetrix Axiom Genome-Wide Pan AFR Genotyping platform (Pan-African array) (Affymetrix, Inc., Santa Clara, California) covers ~2.3 million SNPs optimized for individuals of African ancestry. The Pan-African array was designed to offer ≥90% coverage of SNPs on the Yoruba genome with minor allele frequency (MAF) greater than 2%. Annotations for the Pan-African array can be accessed at the Affymetrix website (http://www.affymetrix.com/). As a platform optimized for individuals of African individuals, the Pan-African array has been extensively validated in African populations from the HapMap Project (Altshuler et al. 2010), including the Luhya from western Kenya (LWK), Maasai from eastern Kenya (MWK), Yoruba from Ibadan, Nigeria (YRI), and the African Ancestry in the Southwest USA (ASW) (Lu et al. 2011). This platform offers high genomic coverage (>85%) in admixed populations with West African ancestry, thus particularly suitable for genome-wide scans in African American individuals (admixture of African and European populations).

Obtaining allele frequency and genetic map distances on parental populations

Genotypes for 2176716 SNPs covered by the Pan-African array were extracted from the 1000 Genomes Project (Abecasis et al. 2010) Phase I data for the 85 CEU (Caucasian residents from Utah, USA) and 88 YRI unrelated samples, representing the two major parental populations for African Americans (Western Africans and Europeans). Genome-wide genetic map distances of SNPs for genome assembly GRCh37 (Frazer et al. 2007) were downloaded from the website (http://bochet.gcc.biostat.washington.edu/beagle/genetic_maps).

Selection of ancestry-informative markers

We aimed to pick the SNPs that were expected to provide the highest mutual information content to ancestry or fixation index (i.e., F_{st}, a measure of population differentiation due to genetic structure) in the genome using an iterative procedure, conditional on the observed allele frequencies in the 1000 Genomes Project CEU and YRI samples.

(a) *Calculation of mutual information content*: Allele frequencies for the CEU and YRI samples were used to calculate the Shannon Information Content (SIC) for each SNP using a formula from previous studies (Smith et al. 2004; Tandon et al. 2011),

$$\mathrm{SIC} = -\sum_{i=0}^{1}\left(a_{i0}+a_{i1}\right)\ln\left(a_{i0}+a_{i1}\right) - \sum_{j=0}^{1}\left(a_{0j}+a_{1j}\right)\ln\left(a_{0j}+a_{1j}\right) + \sum_{i=0}^{1}\sum_{j=0}^{1} a_{\mathrm{ij}}\ln\left(a_{\mathrm{ij}}\right)$$

, where $a_{00} = (1-m)\times p^{\mathrm{YRI}}$, $a_{01} = \left(m\times p^{\mathrm{CEU}}\right)$, $a_{10} = (1-m)\times\left(1-p^{\mathrm{YRI}}\right)$, and $a_{11} = m\times\left(1-p^{\mathrm{CEU}}\right)$. Here, p^{CEU} and p^{YRI} are the allele frequencies in the CEU (European) and YRI (African) samples, and m is the proportion of European ancestry in African Americans, which was set to 0.20 following the same assumption of 20% European ancestry (Tandon et al. 2011). Notably, SNP selection was found not very sensitive to the choice of m (Smith et al. 2004). In addition, the F_{st} was also computed for each of the 2176716 SNPs between the two parental populations based on Wright's approximate formula (Wright 1950),

$$F_{\mathrm{ST}} = \left(H_T - H_S\right)/H_T$$

, where H_T represents expected heterozygosity per locus of the total population and H_S represents expected heterozygosity of a subpopulation.

(b) *Selection of AIMs*: We aimed to detect AIMs that are not packed around certain genomic regions due to linkage disequilibrium (LD), thus being more representative of the genome. Since LD declines gradually with increased genetic distance (Shifman et al. 2003), we assume each SNP is not in LD with distant SNPs more than 0.25 cM (~250 kb) away, similar to what was used in previous publications (Tandon et al. 2011). We selected AIMs using an iterative procedure for each chromosome: 1) SNPs were ranked based on SIC; 2) SNP with the highest SIC was selected as a candidate AIM; 3) Any SNPs within 0.25 cM or within 250 kb of the selected SNP were excluded; 4) Steps 2 and 3 were repeated until no more SNPs left. To avoid densely packed markers, no more than 8 candidate AIMs were selected within any 4 cM region. This procedure ensured a good coverage of AIMs across the entire genome. The quality of the detected candidate AIMs was examined using the build-in data quality checking procedure of ANCESTRYMAP 2.0 (Patterson et al. 2004) f or extracting top "bad" markers, for which allele counts for the ancestral (African and European) genotypes appeared to be grossly inconsistent with counts on the 56 unrelated ASW samples from 1000 Genome Project (Abecasis et al. 2010). After applying the ANCESTRYMAP quality checks, we obtained the final panel of AIMs. We also repeated the same selection procedure using F_{st} to identify a companion panel of AIMs. Supplemental Tables 1 and 2 contain detailed information on the final AIMs.

Evaluation of the detected AIMs for the Pan-African array

The informativeness of the AIMs was evaluated at each SNP using the ANCESTRYMAP-generated *rpower* value, which is a measure of uncertainty in ancestry inference at a given locus. Specifically, *rpower* is the expected value of the squared correlation between inferred and true ancestry (Patterson et al. 2004). In addition, proportion of variance explained (PVE) by the first principal component (PC) using the detected AIMs on the CEU, YRI, and ASW samples was

compared with PVE's from previously published AIMs (based on Affymetrix SNP 6.0 and Illumina 1M arrays) (Tandon et al. 2011) as well as 1000 random sets of SNPs.

Results and Discussion

Given that the Pan-African array was population-optimized, this platform is expected to offer higher coverage of genetic variation for individuals of African ancestry than previous platforms mostly designed based on Caucasians. Genotyping using the Affymetrix Pan-African array will provide opportunities for performing admixture mapping in African Americans to detect genetic variants associated with those traits that exhibit disparities between parental populations, for instance certain cancers (Schwartz et al. 2011). The primary result from this study was a panel of SNPs based on the Pan-African array. We acknowledge that with recent advances in statistical genetics, admixture mapping in African Americans may not rely on a limited number of AIMs any more (Baran et al. 2012; Churchhouse & Marchini 2013; Maples et al. 2013). We propose that some applications for our detected AIMs could include: 1) to facilitate admixture mapping in limited samples; 2) to help identify problematic individuals through genotyping some top-ranking AIMs before starting a large GWAS; 3) to guide targeted re-sequencing projects that may not have genome-wide genotypic data.

Using an iterative selection algorithm, a total of 6011 candidate AIMs were detected based on SIC, which can measure the uncertainty in genome-wide ancestry or ancestry at a given locus (Tandon et al. 2011). We further examined the quality of these candidates using the build-in checking procedure of ANCESTRYMAP (Patterson et al. 2004) and identified a final set of AIMs with 5995 SNPs based on SIC. We also repeated the same analysis using F_{st} to identify a companion panel of 6012 after ANCESTRYMAP checking from 6034detected candidate SNPs.

The selected AIMs with rs numbers, genomic positions, reference alleles, alternative alleles, and allele counts in the CEU or YRI samples are shown in supplemental materials. Overall, AIMs based on SIC and F_{st} performed consistently with each other. The average *rpower* (i.e., average ancestry information) of the AIMs based on SIC or F_{st} was 0.85 (**Figure 1A**), compared to ~0.81 for previous AIMs detected for Affymetrix SNP 6.0 and Illumina 1M arrays (Tandon et al. 2011). The average proportion of European ancestry in ASW was estimated to be 0.25 and 0.24 and the average generations of admixture was estimated to be 5.4 and 5.5 using the AIMs based on SIC and F_{st}, respectively, consistent with previous estimation (Tandon et al. 2011).

The availability of dense genetic variation data from the HapMap Project (HapMap 2003; HapMap 2005) allows a genome-wide analysis of population differentiation. In particular, the CEU (European) and YRI (African) samples represented the two major parental populations of African Americans. Our major criteria of identifying AIMs were designed 1) to enrich SNPs with higher information content (or F_{st}) between the CEU and YRI samples; and 2) to have a comprehensive genomic coverage. The genome-wide iterative scan for AIMs based on a genetic distance bin in a size of 0.25 cM, guaranteed a comprehensive coverage of the entire human genome, as well as limit the possibility that the identified AIMs are in strong LD in a particular genomic region, as described in previous publications (Chen et al. 2010; Tandon et al. 2011). The final AIMs are those SNPs with the highest SIC (or F_{st}) separated by at least the distance of 0.25 cM (~250 kb) between the two parental populations. the detected AIMs were able to recapture the most prominent population structures by being tested on the combined HapMap CEU, YRI, and ASW samples (**Figure 1B**). A simulation analysis demonstrated that the detected AIMs based on the Pan-African array explained substantially higher proportion of variance by the first PCs in the same population than random sets of SNPs in the human genome (**Figure 1C**). Though our analysis showed that the AIMs detected based on SIC and F_{st} performed consistently, given some

potential problems of F_{st}, in particular its dependency on within-population diversity (Sherwin 2010), we generally recommend the use of the final panel of AIMs detected based on SIC.

The assumption of no LD based on 0.25 cM (~250 kb) could be stringent and cause loss of some informative SNPs, given that the average distance of LD decay between SNP pairs is around 20-30 kb across diverse populations, with generally shorter distance in African Americans (Shifman et al. 2003). Nevertheless, this cutoff was chosen to balance between minimizing the possibility of LD and the comprehensive genomic coverage of AIMs (Tandon et al. 2011).

In summary, the Affymetrix Pan-African array provides a population-optimized genotyping platform for GWAS in individuals of African ancestry. The genotypic data profiled by this platform also offers opportunities for admixture mapping in African Americans, a recently admixed population, for certain complex traits and disease susceptibilities with disparities between parental populations. The AIMs we described in this study represent the most informative sets of unlinked markers that can be an important resource to facilitate such applications based on this new tool.

References

Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, and McVean GA. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061-1073.

Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, and McEwen JE. 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467:52-58.

Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, Rodriguez-Cintron W, Chapela R, Ford JG, Avila PC, Rodriguez-Santana J, Burchard EG, and Halperin E. 2012. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics 28:1359-1367.

Chen G, Shriner D, Zhou J, Doumatey A, Huang H, Gerry NP, Herbert A, Christman MF, Chen Y, Dunston GM, Faruque MU, Rotimi CN, and Adeyemo A. 2010. Development of admixture mapping panels for African Americans from commercial high-density SNP arrays. BMC Genomics 11:417.

Churchhouse C, and Marchini J. 2013. Multiway admixture deconvolution using phased or unphased ancestral panels. Genet Epidemiol 37:1-12.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, and Stewart J. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-861.

HapMap. 2003. The International HapMap Project. Nature 426:789-796.

HapMap. 2005. A haplotype map of the human genome. Nature 437:1299-1320.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, and Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106:9362-9367.

Lu Y, Patterson N, Zhan Y, Mallick S, and Reich D. 2011. Technical design document for a SNP array that is optimized for population genetics.ftp://ftp.cephb.fr/hgdp_supp10/.
Maples BK, Gravel S, Kenny EE, and Bustamante CD. 2013. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet 93:278-288.
McKeigue PM. 2005. Prospects for admixture mapping of complex traits. Am J Hum Genet 76:1-7.
Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, and Reich D. 2004. Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74:979-1000.
Ricks-Santi LJ, Apprey V, Mason T, Wilson B, Abbas M, Hernandez W, Hooker S, Doura M, Bonney G, Dunston G, Kittles R, and Ahaghotu C. 2012. Identification of genetic risk associated with prostate cancer using ancestry informative markers. Prostate Cancer Prostatic Dis 15:359-364.
Schwartz AG, Wenzlaff AS, Bock CH, Ruterbusch JJ, Chen W, Cote ML, Artis AS, Van Dyke AL, Land SJ, Harris CC, Pine SR, Spitz MR, Amos CI, Levin AM, and McKeigue PM. 2011. Admixture mapping of lung cancer in 1812 African-Americans. Carcinogenesis 32:312-317.
Sherwin W. 2010. Entropy and information approaches to genetic diversity and its expression: genomic geography. Entropy 12:1765-1798.
Shetty PB, Tang H, Tayo BO, Morrison AC, Hanis CL, Rao DC, Young JH, Fox ER, Boerwinkle E, Cooper RS, Risch NJ, and Zhu X. 2012. Variants in CXADR and F2RL1 are associated with blood pressure and obesity in African-Americans in regions identified through admixture mapping. J Hypertens 30:1970-1976.
Shifman S, Kuypers J, Kokoris M, Yakir B, and Darvasi A. 2003. Linkage disequilibrium patterns of the human genome across populations. Hum Mol Genet 12:771-776.
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, and Reich D. 2004. A high-density admixture map for disease gene discovery in african americans. Am J Hum Genet 74:1001-1013.
Tandon A, Patterson N, and Reich D. 2011. Ancestry informative marker panels for African Americans based on subsets of commercially available SNP arrays. Genet Epidemiol 35:80-83.
Winkler CA, Nelson GW, and Smith MW. 2010. Admixture mapping comes of age. Annu Rev Genomics Hum Genet 11:65-89.
Wright S. 1950. Genetical structure of populations. Nature 166:247-249.

Figure 1

Evaluation anlaysis of ancestry-informative markers.

(A) The *rpower* distributions for AIMs selected based on SIC and F_{st}. The average *rpower* is 0.85 (sd= 0.06) for both lists. (B) Principal components analysis on the 1000 Genomes Project CEU, YRI and ASW panels (n=85 88, 56 unrelated samples, respectively) using the AIMs detected based on SIC. (C) Comparison of the proportion of variance explained (PVE) by the first PCs derived from the CEU, YRI, and ASW samples. The histogram shows the distribution from 1000 randomly-sampled sets of SNPs according to the number of AIMs (based on SIC) on each chromosome. Circles denote real PVE observations for each panel of AIMs: AIMs selected by SIC (5885 SNPs) and F_{st} (6012 SNPs) from Pan-African array, AIMs selected from Affymetrix SNP 6.0 (4290 SNPs), and Illumina 1M (4285 SNPs), respectively.