

SUPPLEMENTAL DIGITAL CONTENT

WHHS Study Design. Participants are seen for in-person visits every six months during which time trained medical interviewers administer an extensive questionnaire, a clinical examination is performed, and biological samples are collected. Laboratory tests are conducted on the women's blood samples to measure HIV viral load (copies/ml) and white blood cell counts, including lymphocyte subsets. Detailed information on participants' current treatment regimen and the antiretroviral medications taken in the six months prior to the current follow-up visit is obtained. Comprehensive information on medication use for other conditions is also obtained and recorded at each follow-up visit. Commencing at visit 9 (1998), data were collected on self-reported adherence to antiretroviral medication (100%, 95-99%, 75-95%, <75% over the past six months) and whether any side effects were experienced.

Race and ethnicity were collected at entry into the cohort. Participants indicated which ethnic group they identify with from the following choices: non-Hispanic White, Hispanic White, 'other Hispanic', non-Hispanic African-American, Hispanic African-American, Asian/Pacific Islander, Native American/Alaskan and 'other'. For purposes of this study, we collapsed Hispanic White and 'other Hispanic' into one group and non-Hispanic and Hispanic African-Americans into one group as a means to graphically display the genetic ancestry substructure results.

Selection of Ancestry Informative Markers (AIMS). Briefly, Smith et al. used a greedy algorithm to select 3 011 markers dispersed throughout the genome that were maximally informative (Shannon information content > 0.035 out of a maximum of 0.709) for high-density admixture mapping [30]. These markers conformed to Hardy-Weinberg

equilibrium in West African and European founder populations, were spaced 50 kb from each other, were not in linkage disequilibrium with one another or any known human disease loci, and were similar in frequency for intracontinental populations [30]. These criteria are widely accepted for the selection of markers for admixture mapping [18,37,39]. Smith and colleagues then identified four lists of 100 markers (n=345 unique markers) that were optimal for distinguishing between West African, European, East Asian, and Native American populations [30]. The average difference in allele frequencies (F_{δ}) for these 345 unique markers were 78% between West African/European, 85% between West African/Native American, 56% between European/Native American, and 57% for European/East Asian [30].

This panel of 345 markers was reduced to 185 markers that had a high probability for successful genotyping on the Illumina Goldengate[®] platform and that were also genotyped as part of The International HapMap Project [35].

Genotyping and Quality Control (QC). Raw genotype data were edited using Illumina BeadStudio software to cluster the three genotypes (AA, AB, BB) at each locus. The systematic QC procedure includes multiple elimination steps; the first round eliminated SNPs that failed (call rate=0%), then DNA samples with missing data for more than 20% of the remaining SNPs were eliminated. SNP assay performance was then assessed, where SNPs with a call rate <90% are eliminated. Lastly, samples that had more than 10% missing data after the previous SNP exclusions were eliminated. Hardy-Weinberg equilibrium (HWE) was tested within each self-identified ethnic group; SNPs with p-values < 0.05 in any one population were visually reviewed for presence of three distinct clusters for accurate genotype determination.

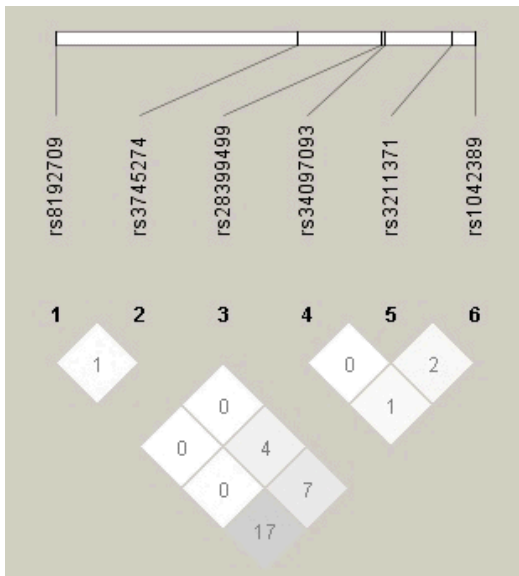
The ancestry informative markers panel (n=185) was genotyped as a subset of a larger panel (n=384). The QC eliminated 24 of 384 SNPs (6%) for which three distinct genotypes were not observed. Next, 76 samples (3%) were eliminated that had missing data for more than 20% of the remaining 360 SNPs. Assay performance for the remaining 360 SNPs was then assessed and seven additional SNPs (2%) were eliminated because they had a call rate < 90%. Lastly, 110 samples (5%) were excluded because they had more than 10% missing data after the elimination of the 31 SNPs. Three SNPs were eliminated due to deviation from HWE. Duplicates for 53 participants were included and concordance was 99.9%. Overall, 34 of the 384 SNPs did not meet quality control criteria, which included 17 AIMs. In summary, a total of 119 participants did not meet QC criteria and were excluded from the genetic ancestry analysis using the remaining 168 AIMs.

The CYP2B6 SNPs (n=11) were genotyped as a subset of a larger panel (n=384). The QC eliminated 11 of 384 SNPs (3%) for which three distinct genotypes were not observed. Next, 51 samples (4%) were eliminated that had missing data for more than 20% of the remaining 373 SNPs. Six additional SNPs (2%) were eliminated because they had a call rate < 90%. Lastly, 13 samples (5%) were excluded because they had more than 10% missing data after the elimination of the 17 SNPs. Duplicates for 78 participants were included and concordance was 99.9%. In total, six CYP2B6 polymorphisms (rs3211371, rs8192709, rs34097093, rs1042389, rs3745274, and rs28399499) passed the QC criterion of successful genotyping on 89.5% of the samples.

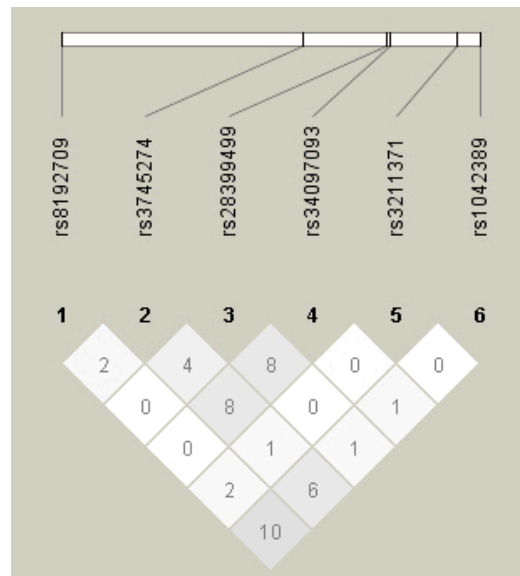
Of the 100 eligible subjects who initiated a NNRTI-based HAART regimen, a total of nine (9%) subjects did not pass the QC criterion of <10% missing data for both the AIMs and CYP2B6 data and were excluded from the analysis.

Estimation of genetic ancestry components. Reference population genotype data were obtained from 11 known populations included in HapMap2 and HapMap3 [35]; European ancestry individuals from Utah (CEU), Yorubans from Nigeria (YRI), Chinese Han from Beijing (CHB), Japanese from Tokyo (JPT), Chinese Han from Denver, Colorado (CHD), Mexicans from Los Angeles (MEX), African Americans from the southwest, U.S. (ASW), Gujarati Indians from Houston, Texas (GIH), Toscani from Tuscany, Italy (TSI), Maasai people from southern Kenya (MKK), and Luhya people from Kenya (LWK). The MCMC scheme was run on two merged datasets: 1) WIHS (n=2 318) and HapMap2 (CEU, YRI, CHB, JPT, n=270) participants for 168 AIMs, and 2) WIHS (n=2 318) and HapMap3 (all populations, n=1 002) participants for the 105 AIMs included in the HapMap3 dataset.

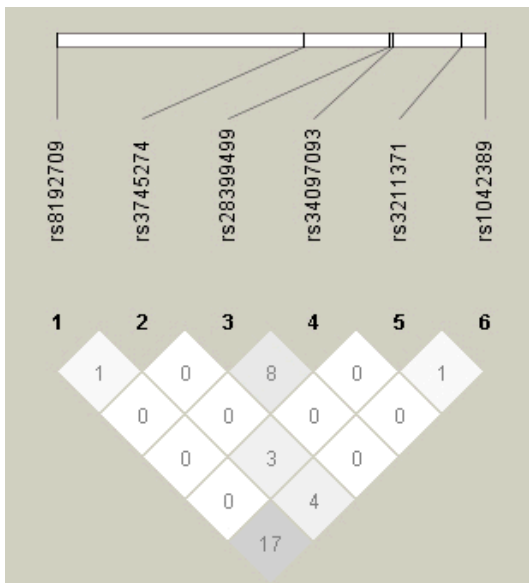
Supplementary Figure 1. Linkage disequilibrium blocks (r^2) for rs3211371, rs1042389, rs8192709, rs3745274, rs28399499 for White, African American, Hispanic, and Other self-reported race/ethnicity groups using all WIHS women who were genotyped for the 384 SNP panel which included these CYPB26 SNPs (n=1,188).



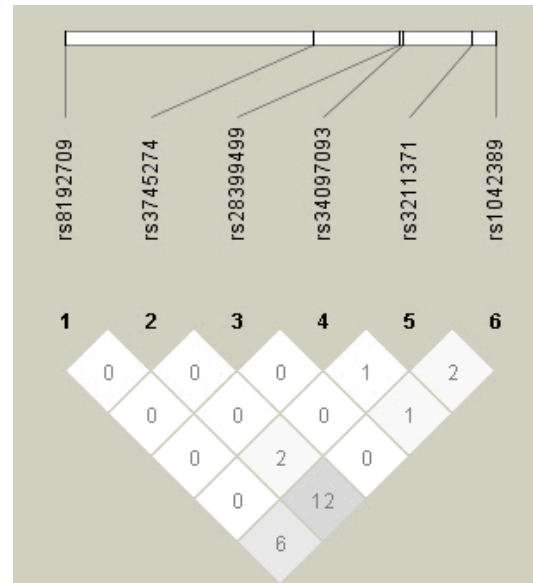
White



African American



Hispanic



Other