MAJOR ARTICLE

Host Determinants of HIV-1 Control in African Americans

Kimberly Pelak,[1] David B. Goldstein,[1] Nicole M. Walley,[1] Jacques Fellay,[1] Dongliang Ge,[1] Kevin V. Shianna,[1] Curtis Gumbs,[1] Xiaojiang Gao,[2] Jessica M. Maia,[1] Kenneth D. Cronin,[1] Shehnaz K. Hussain,[5] Mary Carrington,[2] Nelson L. Michael,[3] Amy C. Weintrob,[4] and the Infectious Disease Clinical Research Program HIV Working Group, on behalf of the National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology (CHAVI)

[1]Center for Human Genome Variation, Duke University Medical School, Durham, North Carolina; [2]Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, National Cancer Institute at Frederick, Frederick, [3]Division of Retrovirology, Walter Reed Army Institute of Research, Rockville, and [4]Infectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, Maryland; [5]Department of Epidemiology, School of Public Health, University of California, Los Angeles

(See the editorial commentary by Motsinger-Reif, on pages 1118–1120.)

We performed a whole-genome association study of human immunodeficiency virus type 1 (HIV-1) set point among a cohort of African Americans ($n = 515$), and an intronic single-nucleotide polymorphism (SNP) in the *HLA-B* gene showed one of the strongest associations. We use a subset of patients to demonstrate that this SNP reflects the effect of the HLA-B*5703 allele, which shows a genome-wide statistically significant association with viral load set point ($P = 5.6 \times 10^{-10}$). These analyses therefore confirm a member of the HLA-B*57 group of alleles as the most important common variant that influences viral load variation in African Americans, which is consistent with what has been observed for individuals of European ancestry, among whom the most important common variant is HLA-B*5701.

A recent genome-wide association study performed among individuals of European ancestry identified 2 polymorphisms associated with human immunodeficiency virus type 1 (HIV-1) load at set point and a third set of polymorphisms associated with a simple measure of disease progression [1]. One variant that was found to be associated with set point (rs2395029) encodes a nonsynonymous change in the *HCP5* gene and is also a tag for HLA-B*5701, which has been shown to be associated with improved early outcomes after exposure to HIV-1 [2, 3]. The other variant that is associated with set point (rs9264942) is 35 kilobases upstream of the *HLA-C* locus and appears to be tagging a causative variant or variants. A third variant (rs9261174) associated with disease progression is located near the *ZNRD1* gene in the major histocompatibility complex (MHC) region, although functional work on this gene has not yet identified a causal variant. Together, these variants are able to explain ~14% of the observed variation in outcome after HIV-1 exposure.

A follow-up study investigated the impact of these same single-nucleotide polymorphisms (SNPs) on an HIV-1–positive African American cohort ($n = 121$) [4]. As was seen in individuals of European ancestry, the *HLA-C*–associated variant (rs9264942) was again found to be associated with viral load, with the C "high-expression" allele leading to lower viral load. Although no association was observed with the G allele of rs2395029, this allele is rare in African Americans; in

Received 20 July 2009; accepted 29 October 2009; electronically published 5 March 2010.

Potential conflicts of interest: none reported.

Financial support: Infectious Disease Clinical Research Program (Department of Defense program executed through the Uniformed Services University of the Health Sciences and funded by the National Institute of Allergy and Infectious Diseases [NIAID], National Institutes of Health [NIH], under interagency agreement Y1-AI-5072); NIH (genetics training grant 5 T32 GM007754-29 to K.P.); NIAID Center for HIV/AIDS Vaccine Immunology (grant AI067854); National Cancer Institute (NCI), NIH (contract HHSN261200800001E); Center for Cancer Research, NCI, NIH. The Multicenter AIDS Cohort Study is funded by NIAID with supplemental funding from NCI and the National Heart, Lung, and Blood Institute (grants UO1-AI-35042, 5-M01-RR-00052 [GCRC], UO1-AI-35043, UO1-AI-37984, UO1-AI-35039, UO1-AI-35040, UO1-AI-37613, and UO1-AI-35041).

Reprints or correspondence: Dr Amy C. Weintrob, Staff Physician, Walter Reed Army Medical Center, Bldg 2, Ward 63, Room 6312, Washington, DC 20307 (amy.weintrob@us.army.mil).

The Journal of Infectious Diseases 2010; 201:1141–1149
© 2010 by the Infectious Diseases Society of America. All rights reserved.
0022-1899/2010/20108-0006$15.00
DOI: 10.1086/651382

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

Table 1. Baseline Characteristics of Participants in the Multicenter AIDS Cohort Study (MACS) and Department of Defense (DoD) Cohorts ($n = 515$)

Variable	MACS Cohort	DoD Cohort
No. of participants	118	397
Male sex, %	100	94
Mean age at seroconversion (range), years	32.0 (20–55)	27.5 (18–55)
Mean viral load set point (range), $\log_{10}$ copies/mL	4.00 (1.91–6.01)	4.12 (1.91–5.97)
Mean year of seroconversion (range)	1991 (1984–2005)[a]	1996 (1986–2003)
Minor allele frequency for rs2523608, %	39.4[b]	35.4
HLA-B*5703 carriers, %	7.5[b]	8.3
Samples measured with Illumina Bead Chip, %		
1M	53	77
1M-Duo	37	23
550K	10	0

[a] Serocoversion date was available for only 52 of the 118 MACS participants.
[b] $P > .05$ comparing MACS with DoD cohort.

people of European ancestry, the allele is in linkage disequilibrium with HLA-B*5701, which is also rare in people of African descent. However, an analysis of the *HLA-B* alleles present in the region showed an association between HLA-B*57 (comprised predominantly of HLA-B*5703) and favorable virologic outcome.

Although the study described above [4] and others [5, 6] have assessed the impact of variants in African Americans that were first identified in patients of European ancestry, to our knowledge there has not yet been any genome-wide investigation of the most important common variants that influence viral load in patients of primarily African ancestry. Here, we present the first genome-wide association study of determinants of HIV-1 control performed among a non-European population. Using a cohort of African American individuals ($n = 515$), we sought to evaluate the associations previously reported and to discover novel or population-specific genetic variants that are associated with HIV-1 control.

METHODS

Samples. This study included HIV-1–infected African American adult participants enrolled in either the United States military Department of Defense Human Immunodeficiency Virus Natural History Study (DoD HIV NHS) or the Multicenter AIDS Cohort Study (MACS). This study was approved by local institutional review boards, and each participant provided written, informed consent.

The DoD HIV NHS (http://www.idcrp.org/hiv-natural-history-study.html) is an ongoing, prospective, continuous-enrollment cohort study of consenting military personnel and beneficiaries with HIV infection and includes participants from the Army, Navy/Marines, and Air Force and their dependents. Since 1985, routine HIV testing (by enzyme-linked immunosorbent assay and confirmatory Western blot analysis) has been used to exclude HIV-infected persons from enlisting for military service or from overseas deployment. Periodic testing among active duty members occurs every 1–5 years, resulting in a defined seroconversion window for incident HIV infection. Participants with HIV infection are referred to military medical centers, where they receive evaluation and ongoing care and are invited to enroll as participants in the DoD HIV NHS.

Those who consent to enroll in the DoD HIV NHS are seen every 6 months by an HIV specialist as part of the study, in addition to receiving routine clinical care. Data are collected on demographic characteristics, markers of HIV disease progression, medication use, and clinical events with medical record confirmation. Cells, plasma, and serum are collected at each visit and stored in a central repository.

Information was extracted from the database on HIV-infected African American individuals with ≤4 years between their last negative and first positive HIV test results, at least 5 million cells stored in the repository, and either 1 viral load measurement taken 3–12 months after seroconversion ($n = 140$) or 2 viral load measurements taken within 3 months to 3 years after seroconversion ($n = 347$). Ethnicity was self-identified. The seroconversion date was estimated as the midpoint between the last negative and first positive HIV test results.

The MACS (http://www.statepi.jhsph.edu/macs/macs.html) is an ongoing prospective study of the natural and treated histories of HIV-1 infection among men who have sex with men that is conducted by sites located in Baltimore, Chicago, Pittsburgh, and Los Angeles. A total of 6973 men have been enrolled; 3427 participants were HIV-seronegative at study entry and were tested for seroconversion semiannually by means of enzyme-linked immunosorbent assay, with positive test results confirmed by Western blot analysis. Of the seroincident par-

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

Table 2. Variants Most Strongly Associated with Set Point in African Americans ($n = 515$)

SNP	Rank	Chromosome	*P*	Closest gene	Type	Frequency
rs454422	1	20	1.49×10^{-6}	*MCM8*	Intronic	0.258
rs2523608	2	6	2.29×10^{-6}	*HLA-B*	Intronic	0.366
rs6948404	3	7	3.41×10^{-6}	*AOAH*	Intronic	0.074
rs558718	4	19	3.71×10^{-6}	*EVI5L*	Intronic	0.110
rs1357339	5	11	4.58×10^{-6}	NA	Intergenic	0.035
rs1413191	6	13	4.61×10^{-6}	*GPC5*	Intronic	0.181
rs236104	7	20	6.72×10^{-6}	*MCM8*	Intronic	0.252
rs7998089	8	13	7.41×10^{-6}	*GPC5*	Intronic	0.191
rs2593321	9	3	7.70×10^{-6}	AC023798.16[a]	Intergenic	0.231
rs6492611	10	13	7.97×10^{-6}	*GPC5*	Intronic	0.208
rs4872511	11	8	9.25×10^{-6}	*PPP3CC*	Downstream	0.011
rs2280890	12	8	9.25×10^{-6}	*SORBS3*	Upstream	0.011
rs2789066	13	6	9.41×10^{-6}	RP11-100A16.1[a]	Upstream	0.127
rs430374	14	18	1.01×10^{-5}	*ST8SIA5*	Intergenic	0.196
rs9910853	15	17	1.07×10^{-5}	*ZNF652*	Intronic	0.085
rs762372	16	21	1.10×10^{-5}	NA	Intergenic	0.470
rs236106	17	20	1.23×10^{-5}	*MCM8*	Intronic	0.252
rs1348478	18	5	1.26×10^{-5}	*PRR16*	Intergenic	0.337
rs12103812	19	17	1.34×10^{-5}	*ZNF652*	3′ UTR	0.084
rs8014482	20	14	1.44×10^{-5}	AL355773.4-1[a]	Intergenic	0.256

NOTE. Age, sex, cohort, and 1 statistically significant EIGENSTRAT axis were used as covariates. NA, not applicable; SNP, single-nucleotide polymorphism; UTR, untranslated region.

[a] Listed as transcript in Ensembl (July 2009).

ticipants, African Americans with available DNA and viral load data from before treatment initiation were selected for inclusion in the current study.

Other cohorts referenced in this analysis include HIV-1–infected adult participants of European ancestry in a study performed by the European Center for HIV/AIDS Vaccine Immunology (Euro-CHAVI) and MACS participants who were included in a previous whole-genome association study ($n = 2362$) [7]. The Euro-CHAVI cohort represents a consortium of 8 European cohorts and 1 Australian cohort of patients who agreed to participate in the Host Genetic Core initiative of the Center for HIV/AIDS Vaccine Immunology (CHAVI).

Genotyping. All samples were genotyped using Illumina HumanHap 1M ($n = 368$), HumanHap 1M-Duo ($n = 135$), or Illumina HumanHap 550K ($n = 12$) Bead Chips. All samples were brought into a single BeadStudio file using the standard Illumina cluster file. For quality control purposes, any sample that had very low intensity or a very low call rate with Illumina cluster (<99%) was deleted. All SNPs that had a call frequency of <99% were put into a filter and reclustered (excluding those on the X chromosome). The reclustering step created SNP calling errors, but the following procedures were used to prevent the errant calls from being released in the final report: (1) the SNPs with a cluster separation value of <0.3 were deleted, and (2) any SNPs with a Het Excess value (an indicator of the quantity of excess heterozygote calls relative to expectations) between −1.0 and −0.1 or between 0.1 and 1.0 were deleted. The filter was then released, and any SNP with a call frequency of <99% was deleted. These procedures resulted in a success rate of genotyping calls ranging from 99.20% to 99.999%, and 1,212,217 SNPs were included in the analysis. Ten samples were excluded because of insufficient call rate.

Specification of sex check, cryptic relatedness check, low minor allele frequency check, Hardy-Weinberg equilibrium check, and a recheck of the genotyping quality were all performed as described elsewhere [1], and no samples were omitted at these steps. A total of 129,723 SNPs were dropped because of a low minor allele frequency. We also required that all SNPs included in the study were successfully genotyped in at least 50% of the samples; hence, 202,676 SNPs were dropped at this point, many of which were those not genotyped on all of the chips.

We used the Illumina 1M and 1M-Duo Bead Chip data as input into the PennCNV program [8], which allowed us to look at deletions (0 or 1 copy) compared with wild types (2 copies) and duplications (3 or 4 copies) compared with wild types (2 copies). Because of the complications of hemizygosity in males and X chromosome inactivation in females, this analysis was restricted to autosomes. In addition, to ensure that we worked with high-confidence copy number variations (CNVs), we excluded any CNV for which the difference in the log like-

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

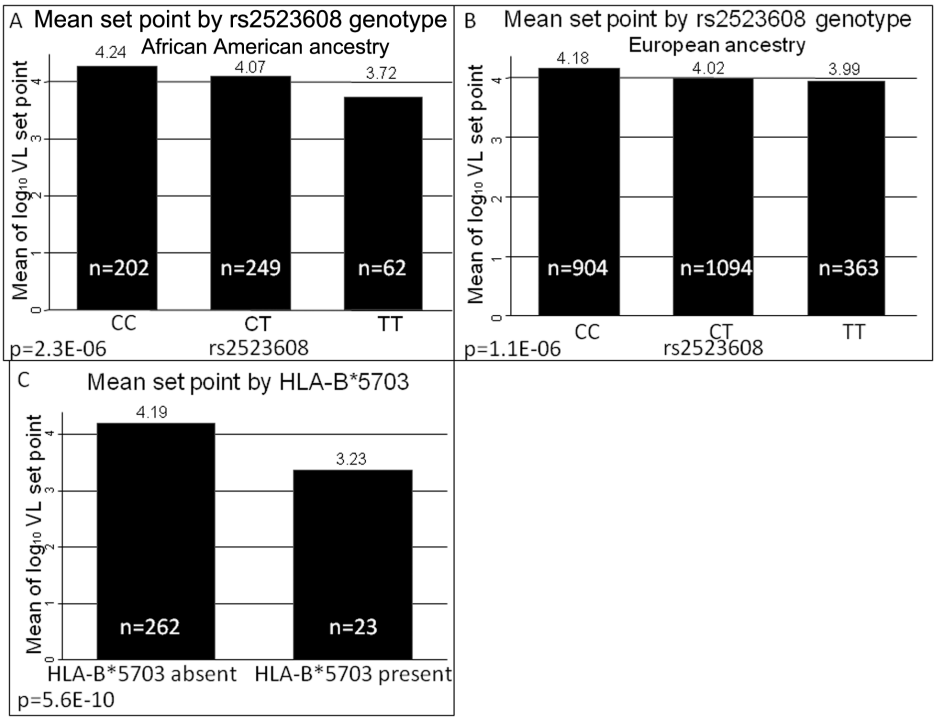

Figure 1. Distribution of mean human immunodeficiency virus type 1 (HIV-1) set point according to patient genotype. The direction of the effect of the rs2523608 genotype is consistent in African American patients (*A*) and European patients (*B*). Patients with HLA-B*5703 have a lower HIV-1 set point (*C*). VL, viral load.

lihood between the most likely copy number state and the less likely copy number state was <10 (generated using the "conf" function in PennCNV). We limited our analysis to CNVs that occurred in at least 3 people (minor allele frequency, >0.003). Further quality control thresholds in PennCNV that were used are detailed in Ge et al [9]. In this analysis, 497 participants were included.

HLA-B allotypes were assigned by DNA sequencing, beginning with the amplification of genomic DNA using primers that flank exons 2 and 3. Polymerase chain reaction products were cleaned using Ampure (Beckman Coulter). The cleaned products were cycle sequenced on an ABI 9700. The cycle sequenced products were cleaned using CleanSEQ (Beckman Coulter) and then run on an ABI Prism 3730. Sequence analysis was performed using Assign (Conexio Genomics).

The EIGENSTRAT method [10] was used to control for population stratification. Assessment of population structure in 616 African Americans by use of the EIGENSTRAT method resulted in 73 statistically significant axes of stratification after the removal of 35 population outliers. The first axis made a larger contribution to the proportion of variation (0.6%) explained than the contribution made by the second axis (0.2%) and reflected the degree of African versus European ancestry in individuals. We therefore used only the first axis as a covariate in our association analyses, to control for population stratification.

Phenotype. Set point viral load was defined as (1) the mean viral load for samples with 2 or more viral load values that were collected at least 30 days apart, within 3 months to 3 years after seroconversion, and were within a 1 log range, similar to what was done in the previous study of determinants of set point in participants of European ancestry [1], or (2) the first available viral load within 3–12 months after seroconversion, as long as there was a corresponding $CD4^+$ T cell count of >350 cells/μL. All viral loads were measured prior to the initiation of antiretroviral therapy. Viral loads from the 2 definitions were highly correlated ($r^2 = 0.74$), and the first definition was preferentially used when sufficient data were available.

Statistical analysis. Viral load at set point was used as a quantitative trait in a linear regression using additive allelic effects. Because previous studies have found that sex and age may be associated with viral load, these factors were used as covariates in the model [11, 12]. The first EIGENSTRAT axis was used to control for population stratification, and the cohort was also included in the model because of baseline differences between the 2 groups. Individual regressions for each SNP were performed using PLINK software (version 1.06) [13, 14]. The Bonferroni correction was used to control for multiple comparisons. Associations for which $P < 5 \times 10^{-8}$ were considered to have genome-wide statistical significance.

All *HLA-B* allotypes were tested for association with viral

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

Table 3. Major Histocompatibility Complex (MHC) Variants and Functional Variants That Are Associated with Viral Load Set Point in African Americans ($n = 515$)

SNP	Overall rank	*P*	Gene	Type	Frequency
Top 10 SNPs in the MHC region					
rs2523608	2	2.29×10^{-6}	*HLA-B*	Intronic	0.366
rs34548063	23	1.57×10^{-5}	*STK19*	Stop gained	0.028
rs2523933	29	2.03×10^{-5}	*HLA-W*	Intergenic	0.205
rs2844538	53	5.07×10^{-5}	*ZDHHC20P2*	Downstream	0.492
rs2596503	81	8.15×10^{-5}	*HLA-B*	Downstream	0.162
rs9266689	96	9.86×10^{-5}	*ZDHHC20P2*	In noncoding gene	0.485
rs4151650	99	1×10^{-4}	*CFB, C2*	Synonymous coding	0.030
rs1736936	99	1×10^{-4}	*HLA-G*	Upstream	0.492
rs9378200	150	2×10^{-4}	*UQCRHP*	Intergenic	0.013
rs4713213	150	2×10^{-4}	*OR5V1, OR12D3*	Intronic	0.331
Top 10 functional SNPs					
rs34548063	23	1.57×10^{-5}	*STK19*	Stop gained	0.028
rs1034405	91	9.65×10^{-5}	*C3orf18*	Nonsynonymous coding	0.332
rs4838865	99	1×10^{-4}	*TUBGCP6*	Nonsynonymous coding	0.048
rs6542522	99	1×10^{-4}	*C2orf76*	Nonsynonymous coding, splice site	0.196
rs2280801	332	4×10^{-4}	*BAT2*	Nonsynonymous coding	0.012
rs13146272	332	4×10^{-4}	*CYP4V2*	Nonsynonymous coding	0.394
rs2273549	332	0.0004	*TCP11L1*	Nonsynonymous coding	0.092
rs2278329	332	0.0004	*OSMR*	Nonsynonymous coding	0.042
rs10423723	435	0.0005	AC020907.3[a]	Nonsynonymous coding	0.625
rs7258700	435	0.0005	AC020907.3[a]	Nonsynonymous coding	0.625

NOTE. Age, sex, cohort, and 1 statistically significant EIGENSTRAT axis were used as covariates. SNP, single-nucleotide polymorphism.

[a] Listed as transcript in Ensembl (July 2009).

load set point in a linear regression model and were also evaluated to determine whether they were responsible for associations observed for SNPs in the genome-wide SNP association analyses.

RESULTS

From the DoD HIV NHS cohort, 487 participants met the inclusion criteria and 471 were successfully genotyped. From the MACS cohort, 158 participants met the inclusion criteria and 145 were successfully genotyped. Thirty-five participants were dropped from the analysis because of the EIGENSTRAT correction for ancestry, and 66 participants were not included because their viral load results did not meet the definition of set point described in Methods, which left 515 participants in the final data set. Table 1 shows the baseline characteristics of participants in the DoD and MACS cohorts, as well as the Bead Chips that were used for genotyping. There were more women in the DoD cohort ($P = .006$), and participants in this cohort were, on average, younger at seroconversion ($P < .001$); however, there was no difference between the cohorts in mean set point viral load ($P = .13$).

No single SNP had a genome-wide statistically significant association ($P < 5 \times 10^{-8}$) with viral load at set point. Table 2 lists the top 20 genome-wide associations with set point, and Table 3 lists the top 10 associations from the MHC region and the top 10 functional SNPs that are associated with set point. A functional SNP was defined as a SNP that would cause the gain or loss of a stop codon, would cause a nonsynonymous coding change, or occurred in a splice site.

The most statistically significant SNP in the MHC region for association with viral load set point in this African American cohort was rs2523608, located in the *HLA-B* gene ($P = 2.3 \times 10^{-6}$) (Figure 1*A* and Table 3). We found that the same SNP was also statistically significantly associated with HIV-1 set point in a large sample of individuals of European ancestry ($P = 1.1 \times 10^{-6}$, corrected for age and sex) (Figure 1*B*). This association remains nominally statistically significant ($P = .0083$) after accounting for variants in the MHC region that were previously shown to be associated with HIV-1 outcomes (rs2395029, rs9264942, and rs9261174) and 12 statistically significant EIGENSTRAT axes to control for population stratification in this cohort.

The rs2523608 variant is located in intron 5 (according to Ensembl transcript ENST00000376228), which is >100 base pairs from the nearest exon. Analysis of the *HLA-B* allotypes from 285 genotyped study participants showed that this association was due to the association between rs2523608 and HLA-B*5703 ($D' = 1$; $r^2 = 0.075$). The degree of linkage dis-

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

Table 4. Top Associations between *HLA-B* Allotypes and Human Immunodeficiency Virus Type 1 Set Point ($n = 285$)

Allotype	Rank	*P*	Frequency
HLA-B*5703	1	5.6×10^{-10}	0.040
HLA-B*3910	2	0.00032	0.006
HLA-B*1517	3	0.00040	0.006
HLA-B*4501	4	0.00084	0.062
HLA-B*1302	5	0.020	0.005
HLA-B*580101	6	0.021	0.016
HLA-B*5802	7	0.022	0.040
HLA-B*4201	8	0.022	0.040
HLA-B*140201	9	0.024	0.020
HLA-B*1801	10	0.025	0.016

NOTE. Age, sex, cohort, and 1 statistically significant EIGENSTRAT axis were used as covariates.

equilibrium can be quantified using the D' statistic [15]. This statistic compares the ancestral recombination patterns between 2 variants by standardizing allele frequencies. A value of $D' = 1$ indicates that 1 variant always appears on the background of the other. On the other hand, the r^2 statistic is sensitive to allele frequency differences and assesses the degree to which the 2 variants appear together [16, 17]. When considered alone, HLA-B*5703 had by far the strongest association with viral load set point of any *HLA-B* allotype, showing genome-wide statistical significance ($P = 5.6 \times 10^{-10}$, with age, sex, cohort, and 1 EIGENSTRAT axis as covariates) (Table 4 and Figure 1*C*). Moreover, when HLA-B*5703 was included as a covariate, it was able to account for the effect of the rs2523608 genotype. This analysis shows that HLA-B*5703 is the most important common variant in influencing viral load in African Americans, explaining ~10% of the variation in viral load set point in this data set, with an allele frequency of ~4.0%.

CNV analysis. There were 8724 SNPs that showed evidence of a duplication and 16,778 SNPs that showed evidence of a deletion. The CNV calls for each SNP were then run as genotypes in a regression using an additive genetic model, testing for association with HIV-1 set point. Sex, cohort, and the first EIGENSTRAT axis were used as covariates. After a Bonferroni correction was applied (6×10^{-6} for duplications and 3×10^{-6} for deletions), no SNPs reached genome-wide statistical significance for either deletions or duplications. Furthermore, when the association results from these 2 models were compared, there was no SNP associated with both deletions and duplications for which $P < .05$.

Association with previously implicated variants. We also analyzed genetic variants that had previously been shown to have an effect on HIV-1 set point. First, we tested the association with rs2395029, a nonsynonymous SNP in the *HCP5* gene that is a tag for the functional allele HLA-B*5701, and found this SNP to show a weak association with viral load set point ($P = .030$) (Table 5). This SNP has a very low minor allele frequency in African Americans (minor allele frequency, 0.008) because of its virtual absence in West African populations. Therefore, the power to detect an association in this cohort is only 81% at $P < .05$, assuming that the effect size in our cohort is comparable to that seen for individuals of European descent [1].

We then tested the association with rs9264942, a C→T polymorphism that is 35 kilobases upstream of the *HLA-C* gene. This SNP itself is not causal, but it is a tagging SNP for an unknown causal variant or variants. We observed a weak association between rs9264942 and set point in African Americans ($P = .018$) (Table 5).

We also tested the associations between viral load at set point and rs9261174, a SNP located near the *ZNRD1* gene in the MHC region, and *CCR5-Δ32*, a 32–base pair deletion in the *CCR5* gene (rs333) that is rare in non-European populations. In our African American cohort, neither rs9261174 ($P = .352$) nor *CCR5-Δ32* ($P = .484$) showed an association with viral load at set point (Table 5).

DISCUSSION

To our knowledge, this was the first genome-wide association study on HIV-1 outcomes to be performed among an African American cohort, the majority of whom were infected with HIV-1 subtype B. We have shown that the intronic SNP rs2523608, the top associated SNP in the MHC region, is tagging HLA-B*5703. The value of D' for the association between rs2523608 and HLA-B*5703 is 1, and a regression model shows that the HLA-B*5703 genotype is able to account for the effect of this intronic SNP. HLA-B*5703 is strongly associated with viral load at set point and reached whole-genome statistical significance in the subset of samples for which *HLA-B* allotype data were available ($P = 5.6 \times 10^{-10}$, in a model that also included age, sex, cohort, and the first EIGENSTRAT axis; $n = 285$).

HLA-B*5701 is an important mechanism of HIV-1 control in the European population. It had an allele frequency of ~6.1% in a European population, but it was not observed in a Yoruban population [18]. Its close relative HLA-B*5703 was absent in a European population, but HLA-B*5703 had an allele frequency of ~5.8% in a Yoruban population [18]. Here, we have shown that African American individuals who have HLA-B*5703 also show improved viral control, of a magnitude similar to that afforded by HLA-B*5701 in people of European descent. Thus, our results indicate that the general mechanism of genetic control of HIV-1 in African Americans is similar to that in Europeans: HLA-B*5701 accounts for ~6% of the observed variation in viral load set point in Europeans [7], and HLA-B*5703 accounts for ~10% of the observed variation in viral load set point in African Americans. There was also a

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

Table 5. Association Outcomes for Genetic Variants That Have Been Previously Reported to Be Associated with Human Immunodeficiency Virus Type 1 Set Point or Disease Progression

			African Americans			European Americans		
SNP	Gene	Type	No. of patients	P^a	MAF	No. of patients	P^b	MAF
rs2395029	*HCP5*	Nonsynonymous	513	0.030	0.008	2362	4.5×10^{-35}	0.048
rs9264942	*HLA-C*	Upstream	515	0.018	0.286	2362	5.9×10^{-32}	0.412
rs9261174	*ZNRD1*	Intergenic	511	0.352	0.245	2362	1.1×10^{-4}	0.141
rs333 (Δ32)	*CCR5*	Genic deletion	502	0.484	0.017	2333	1.7×10^{-10}	0.099

NOTE. MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

[a] *P* values are corrected for age, sex, cohort, and 1 statistically significant EIGENSTRAT axis.

[b] *P* values are corrected for age, sex, and 12 statistically significant EIGENSTRAT axes.

small contribution to HIV-1 control by HLA-B*5701 (frequency, 0.3%) in our African American data set, because of admixture.

We also found that HLA-B*3910 and HLA-B*1517, in addition to HLA-B*5703, may be playing a lesser role in viral control, although additional studies would be needed to confirm these observations. This pattern was similar to that observed in people of European descent, among whom HLA-B*5701 is the largest determinant of HIV-1 control but other *HLA-B* alleles (HLA-B*27, HLA-B*35, and others) also play a role [7]. The effects of the alleles that we have identified are further supported by an analysis of the MHC region that was conducted in southern African populations infected with HIV-1 subtype C, in which HLA-B*5703 was found to be the HLA class I allele most strongly associated with a decreased viral load set point [3, 19], with a weaker contribution by HLA-B*39 [19].

We found a reduced or absent association with viral load set point when we explicitly checked variants that had been shown to be associated with set point in a cohort of patients of European descent. Similar outcomes were seen in Shrestha et al [4], who saw a reduced association between rs9264942 and set point and no association between rs2395029 and set point. Our sample size was >4 times larger than that in Shrestha et al [4] and revealed only weak associations between both variants and set point. So although both rs9264942 and rs2395029 are definitively associated with viral load set point in European populations, neither study was able to replicate these associations in an African American cohort.

It is worth noting that it is only the HLA-B*5701 association in the previous study [1] for which the causal site is thought to have been identified. In the cases of rs9264942 and rs9261174, it is likely that the associated variants are not

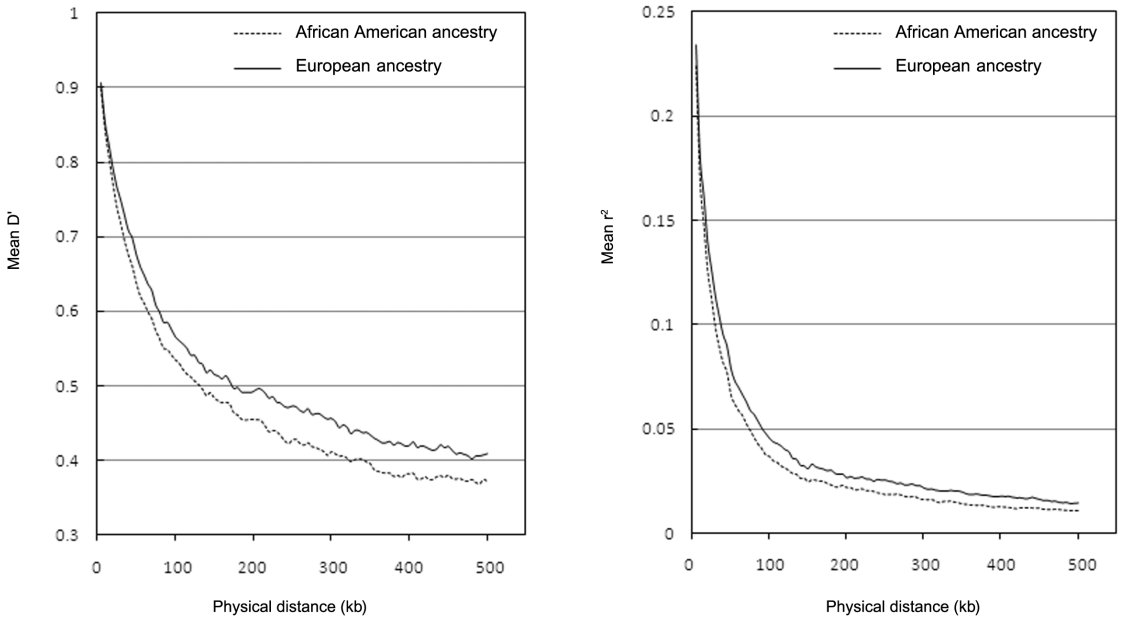

Figure 2. Comparison of the mean linkage disequilibrium in the major histocompatibility complex region between persons of European and African American ancestry. The mean values of D' and r^2 are both smaller in the African American genome than in the European genome. Kb, kilobase.

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

themselves causal but rather markers of an as yet unidentified causal site or sites. It may therefore not be a coincidence that the HLA-B*57 association is the only one to show an effect in African Americans that is similar to that shown in individuals of European ancestry. For the other 2 associations, the different and generally lower linkage disequilibrium in this region in African Americans (Figure 2) could mean that the causal sites are no longer being tagged by these variants.

A potential limitation of this study is that HIV-infected persons with rapid disease progression may have been excluded. Rapid progressors may not have had many research visits and therefore may not have had enough cells in the repository to be included in this study. In addition, because of their rapid disease progression, they may not have had available viral load measurements that satisfied the definition of set point. Approximately 10% of the participants in this study progressed to a $CD4^+$ cell count of <200 cells/mm^3 within 2 years after seroconversion; therefore, there were at least some rapid progressors included in this study.

Earlier studies have likewise suggested that HLA-B*5703 may also be involved in HIV-1 control in African Americans, but our study shows that it is indeed the most statistically significant common genetic factor affecting early viral control in this population. By using a genome-wide scan to implicate another allele of HLA-B*57 in HIV-1 control, in an ancestral background that was entirely different from that where this association had previously been observed, we provide further support for the important role played by HLA-B*57 in HIV-1 control and the decreased fitness level of the viral mutants that are selected for by HLA-B*57. Given the increased burden of disease in African and African American populations and the paucity of common variants that clearly influence HIV-1 control, it is important to continue to investigate rare variants that may function specifically in these populations.

MEMBERS OF THE INFECTIOUS DISEASE CLINICAL RESEARCH PROGRAM HIV WORKING GROUP

M. Polis, J. Powers, and E. Tramont (National Institute of Allergy and Infectious Diseases, Bethesda, MD); J. Maguire (Naval Medical Center, Portsmouth, VA; M. Bavaro, N. Crum-Cianflone, and H. Chun (Naval Medical Center, San Diego, CA); C. Decker, A. Ganesan, and T. Whitman (National Naval Medical Center, Bethesda, MD); W. Bradley, V. Marconi, S. Merritt, and J. Okulicz (San Antonio Military Medical Center, San Antonio, TX); A. Johnson (Tripler Army Medical Center, Honolulu, HI); B. Agan (Uniformed Services University of the Health Sciences, Bethesda, MD); C. Eggleston, L. Jagodzinski, R. O'Connell, and S. Peel (Walter Reed Army Institute of Research, Rockville, MD); C. Hawkes, G. Wortmann, and M. Zapor (Walter Reed Army Medical Center, Washington, DC); and L. Eberly and A. Lifson (University of Minnesota).

MULTICENTER AIDS COHORT STUDY CENTERS

The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD (J. Margolick); Howard Brown Health Center and Northwestern University Medical School, Chicago, IL (J. Phair); University of California, Los Angeles, CA (R. Detels); University of Pittsburgh, Pittsburgh, PA (C. Rinaldo); and Data Analysis Center, Baltimore, MD (L. Jacobson).

Acknowledgments

The content of this publication is the sole responsibility of the authors and does not necessarily reflect the views or policies of the National Institutes of Health or the Department of Health and Human Services, or of the Department of Defense or the Department of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the United States government.

We thank all of the patients in the Department of Defense HIV Natural History Study, Multicenter AIDS Cohort Study, and European Center for HIV/AIDS Vaccine Immunology (Euro-CHAVI) cohorts.

References

1. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. Science **2007**; 317: 944–947.
2. Migueles SA, Sabbaghian MS, Shupert WL, et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci U S A **2000**; 97:2709–2714.
3. Kiepiela P, Leslie AJ, Honeyborne I, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature **2004**; 432:769–775.
4. Shrestha S, Aissani B, Song W, Wilson CM, Kaslow RA, Tang J. Host genetics and HIV-1 viral load set-point in African-Americans. AIDS **2009**; 23:673–677.
5. Shrestha S, Strathdee SA, Galai N, et al. Behavioral risk exposure and host genetics of susceptibility to HIV-1 infection. J Infect Dis **2006**; 193: 16–26.
6. Lama J, Planelles V. Host factors influencing susceptibility to HIV infection and AIDS progression. Retrovirology **2007**; 4:52.
7. Fellay J, Ge D, Shianna KV, et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet **2009**; 5:e1000791.
8. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res **2007**; 17:1665–1674.
9. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature **2009**; 461:399–401.
10. 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.
11. Napravnik S, Poole C, Thomas JC, Eron JJ. Gender difference in HIV RNA levels: a meta-analysis of published studies. J Acquir Immune Defic Syndr **2002**; 31:11–19.
12. Quinn TC, Wawer MJ, Sewankwambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. N Engl J Med **2000**; 342:921–929.
13. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011

genome association and population-based linkage analysis. Am J Hum Genet **2007**; 81:559–575.
14. Purcell S. PLINK: whole genome association analysis toolset. Harvard University Web site. http://pngu.mgh.harvard.edu/purcell/plink/. Updated 10 October 2009. Accessed 15 June 2009.
15. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **1964**; 49:49–67.
16. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet **1999**; 22:139–144.
17. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet **2001**; 69:1–14.
18. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet **2006**; 38:1166–1172.
19. Tang J, Tang S, Lobashevsky E, et al. Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. J Virol **2002**; 76:8276–8284.

Downloaded from jid.oxfordjournals.org at UniversitÃ© & EPFL Lausanne on April 11, 2011