OPEN

Variants in Host Viral Replication Cycle Genes Are Associated With Heterosexual HIV-1 Acquisition in Africans

Abigail W. Bigham, PhD, Romel D. Mackelprang, PhD,† Connie Celum, MD,†‡§ Guy De Bruyn, MD,‖ Kristin Beima-Sofie, PhD,‡ Grace John-Stewart, MPH, MD, PhD,†‡§ Allan Ronald, MD,¶ Nelly R. Mugo, MD, MPH, MMed,†# Kati Buckingham, BS,** Michael J. Bamshad, MD,**†† James I. Mullins, PhD,¶‡‡ M. J. McElrath, MD, PhD,¶§§ and Jairam R. Lingappa, MD, PhD†§***

Objective: We evaluated genetic variants in 51 candidate genes encoding proteins that interact with HIV-1 during the virus life cycle for association with HIV-1 outcomes in an African cohort.

Methods: Using a nested case–control study within a cohort of heterosexual HIV-1–serodiscordant couples, we genotyped 475 haplotype-tagging single-nucleotide polymorphisms (tagSNPs) and 18 SNPs previously associated with HIV-1 transmission and/or progression (candidate SNPs) in 51 host genes. We used logistic and Cox proportional hazard regression with adjustment for sex, age, and population stratification to detect SNP associations with HIV-1 acquisition, plasma HIV-1 set point, and a composite measure of HIV-1 disease progression. Significant thresholds for tagSNP, but not candidate SNP, associations were subjected to Bonferroni correction for multiple testing.

Results: We evaluated 491 HIV-1–infected and 335 HIV-1–uninfected individuals for 493 SNPs, 459 of which passed quality control filters. Candidate SNP *PPIA* rs8177826 and tagSNP *SMARCB1* rs6003904 were significantly associated with HIV-1 acquisition risk (odds ratio = 0.14, $P = 0.03$, and odds ratio = 2.11, $P_{corr} = 0.01$, respectively). Furthermore, the TT genotype for *CCR5* rs1799988 was associated with a mean 0.2 $\log_{10}$ copies per milliliter lower plasma HIV-1 RNA set point ($P = 0.04$). We also identified significant associations with HIV-1 disease progression for variants in *FUT2* and *MBL2*.

Conclusions: Using a targeted gene approach, we identified variants in host genes whose protein products interact with HIV-1 during the virus replication cycle and were associated with HIV-1 outcomes in this African cohort.

Key Words: HIV, cell cycle genes, genotype–phenotype association, HIV genetics, Africa

(*J Acquir Immune Defic Syndr* 2014;66:127–134)

Received for publication January 3, 2014; accepted January 3, 2014.

From the *Department of Anthropology, University of Michigan, Ann Arbor, MI; Departments of †Global Health; ‡Epidemiology; §Medicine, University of Washington, Seattle, WA; ‖Perinatal HIV Research Unit, University of Witwatersrand, Johannesburg, South Africa; ¶Department of Medicine, University of Manitoba, Winnipeg, Canada; #Department of Obstetrics and Gynaecology, University of Nairobi, Nairobi, Kenya; Departments of **Pediatrics; ††Genome Sciences; ‡‡Laboratory Medicine, University of Washington, Seattle, WA; and §§Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

A.W.B. and R.D.M. performed the analysis, A.W.B., R.D.M., and J.R.L. wrote the manuscript. J.R.L. conceived of the experiments. M.J.B. and K. B. provided materials and reagents. K.B.S., C.C., G.D.B., G.J.S., M.J.M., J.I.M., N.R.M., and A.R. assisted with sample collection.

Presented at the Keystone Symposia: HIV Evolution, Genomics and Pathogenesis, March 20-25, 2011, British Columbia, Canada.

J.R.L. is a scientific board member of Prosetta AntiViral Inc. G.J.S. receives royalties from UpToDate. M.J.B. receives royalties from textbook publication.

Supported by the Bill and Melinda Gates Foundation (grants 26469 and 41185), National Institutes of Health (NIH)/*National Institute of Allergy and Infectious Diseases* (R21 grant AI073115), and the University of Washington Center for AIDS Research, an NIH funded program (P30 AI027757). A.W.B. was supported by a training fellowship from the NIH/National Human Genome Research Institute. R.D.M. was supported by the University of Washington STD/AIDS Research Training Program (T32AI007140) from the NIH, US Public Health Service.

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License, where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially. http://creativecommons.org/licenses/by-nc-nd/3.0.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Correspondence to: Abigail W. Bigham, PhD, Department of Anthropology, University of Michigan, 101 West Hall, 1085 S University Avenue, Ann Arbor, MI 48109 (e-mail: awbigham@umich.edu).

Copyright © 2014 by Lippincott Williams & Wilkins

INTRODUCTION

A variety of host proteins interact with HIV-1 during the replication cycle of the virus, such as the *CCR5* chemokine receptor (CCR5) and the chemokine (*C-C motif*) ligand 3 (CCL3). Variation in genes encoding these and other host proteins may influence the efficiency of HIV-1 replication. Specifically, variants in host genes mediating HIV-1 cell entry, nuclear localization and integration, transcription, capsid formation, intracellular trafficking/virus budding, and viral restriction may contribute to defining the risk of HIV-1 acquisition, plasma HIV-1 RNA levels, and the pace of HIV-1 disease progression. This was first evident in studies of a 32 base pair deletion in the cellular *CCR5* chemokine receptor (*CCR5Δ32*) that serves as a coreceptor for HIV-1 cell entry. *CCR5Δ32* was shown to be strongly associated with resistance to HIV-1 infection and delayed disease progression.[1] However, it is absent or extremely rare in Africans, where 65% of global HIV-1 disease burden is concentrated.[2] This observation highlights the need for independent discovery and confirmation of host genetic variants influencing HIV-1 acquisition and pathogenesis in global populations.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Such efforts offer the potential to identify biological factors and pathways to target in vaccine or drug development and may also help stratify vaccine trial participants by HIV-1 susceptibility risk.

In the context of recent HIV-1 prevention trials in Eastern and Southern Africa, we have prospectively followed cohorts of HIV-1–serodiscordant couples (one partner HIV-1 infected and the other HIV-1 uninfected). These trials have provided us with a unique opportunity to evaluate host genetic associations with HIV-1 outcomes in African populations while allowing us to control for epidemiologic factors known to alter the risk of HIV-1 acquisition (eg, plasma HIV-1 level of the HIV-1–infected partner, frequency of unprotected sex, and circumcision status of male HIV-1–uninfected partners). Such factors can be used to classify individuals according to HIV-1 exposure, thus reducing misclassification.[3]

Recently, we performed a genome-wide association study (GWAS) to evaluate common variants for association with HIV-1 acquisition and set point viral load. This analysis found no variants associated with GWAS statistical significance.[4] However, given the large number of common variants tested, this analysis was only powered to identify factors with large relative risks (RR) [eg, RR $>$ 3.2 for variants with minor allele frequency (MAF) of 5%]. Furthermore, the low level of linkage disequilibrium (LD) present in African populations may have reduced the capacity of standard GWAS chips to broadly capture genome-wide disease associations.

Recognizing these limitations, we performed a targeted gene evaluation of single-nucleotide polymorphisms (SNPs) in 51 genes previously demonstrated to interact with HIV-1 during the virus life, thus selecting a set of variants with a higher a priori likelihood of being associated with HIV-1 outcomes and which have been understudied for susceptibility to or disease progression/reactivation from HIV. Here we report the association of these candidate and haplotype-tagging SNPs (tagSNPs) with HIV-1 outcomes in the same population previously evaluated through a genome-wide analysis.

METHODS

Study Subjects

Study participants originated from 2 cohorts of HIV-1–serodiscordant couples in East and Southern Africa (the Couples Observational study and The Partners in Prevention HSV/HIV Transmission Study), as previously described.[4,5] Briefly, these studies collected prospective clinical, epidemiologic, and laboratory data and specimens at quarterly visits from both partners in heterosexual couples who were HIV-1 serodiscordant at the time of enrollment. Laboratory confirmation of HIV-1 transmission linkage within a partnership was performed through HIV-1 *env* and *gag* sequencing of plasma virus from both partners.[6] For this analysis, we used a nested case–control study design to match 126 couples with HIV-1 transmission to 250 couples without transmission based on estimated levels of HIV-1 exposure (Fig. 1), for a total enrollment of 376 couples. HIV-1 exposure scores represent the risk of infection at enrollment and were derived from data on HIV-1 plasma RNA levels of the infected partner, unprotected sex, age, sex, and male circumcision.[4] We further augmented the set of uninfected controls with 94 HIV-1–exposed seronegative (HESN) participants with high levels of HIV-1 exposure to improve our power to detect protective alleles.[3] HIV-1–infected partners were defined as with either seroincident HIV-1 infections who seroconverted during study follow-up or seroprevalent HIV-1 infections who were HIV-1 infected before study enrollment and most likely in chronic phase HIV-1 infection. Overall, this study included 846 subjects comprised of 502 HIV-1–infected cases (375 seroprevalent and 127 seroincident) and 344 HIV-1–uninfected controls (250 matched to cases and 94 unmatched) enrolled at 10 study sites. The study sites included 4 from Southern Africa: (1) Gaborone, Botswana; (2) Cape Town, South Africa; (3) Johannesburg, South Africa; and (4) Orange Farm, South Africa, and 6 from East Africa: (1) Thika, Kenya; (2) Eldoret, Kenya; (3) Kisumu, Kenya; (4) Nairobi, Kenya; (5) Moshi, Tanzania; and (6) Kampala, Uganda. The epidemiology of HIV-1 subtypes for these 10 study sites is consistent with the global epidemiology, wherein subtype C predominates in Southern Africa and subtypes A and D dominate in East Africa. In this cohort, neither subtype was associated with increased risk of HIV-1 transmission nor did plasma RNA viral load significantly differ between C and non-C subtypes.[7,8] All participants provided written informed consent, and the study was approved by the institutional review boards at the University of Washington, the University of Michigan, and all local study sites and affiliated institutions.

DNA Isolation and Genotyping

DNA was isolated from archived venous blood using Qiagen's (Valencia, CA) Puregene DNA purification system according to the manufacturer's instructions. Candidate genes (n = 51) were selected from 6 categories of proteins (Table 1). Each gene encodes a protein that was either predicted or demonstrated empirically to interact with HIV-1 during its life cycle, including entry into the host cell,[1] capsid formation,[9,10] integration into the host genome,[11] intracellular restriction,[12] intracellular trafficking,[13–16] and genes identified through small interfering RNA knockdown.[17] Haplotypes were inferred for each of the 51 candidate genes using the Haplotype Mapping Project (HapMap) Yoruba from Ibadan, Nigeria. SNPs that tagged each haplotype with a frequency $>$5% in this population were selected to be genotyped. In addition, previously identified SNPs with known association with HIV-1 resistance/susceptibility (candidate SNPs) were also selected for genotyping.

In total, 638 SNPs were genotyped using an Illumina Custom Oligo Pooled Assay that included 475 tagSNPs, 18 candidate SNPs, and 144 ancestry informative markers (AIMs) (see **Table S1, Supplemental Digital Content**, http://links.lww.com/QAI/A506). AIMs were included to adjust for population stratification and therefore chosen based on their ability to distinguish among East Asian, European, West African, and East African populations. Details, including allele frequencies in the parental populations, DNA sequences, and the genomic positions of SNPs, are provided in Table S2 (see **Supplemental Digital Content**, http://links.lww.com/QAI/A506). Principal components (PCs) were estimated from the 144 AIMs using EIGENSTRAT.[18]

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

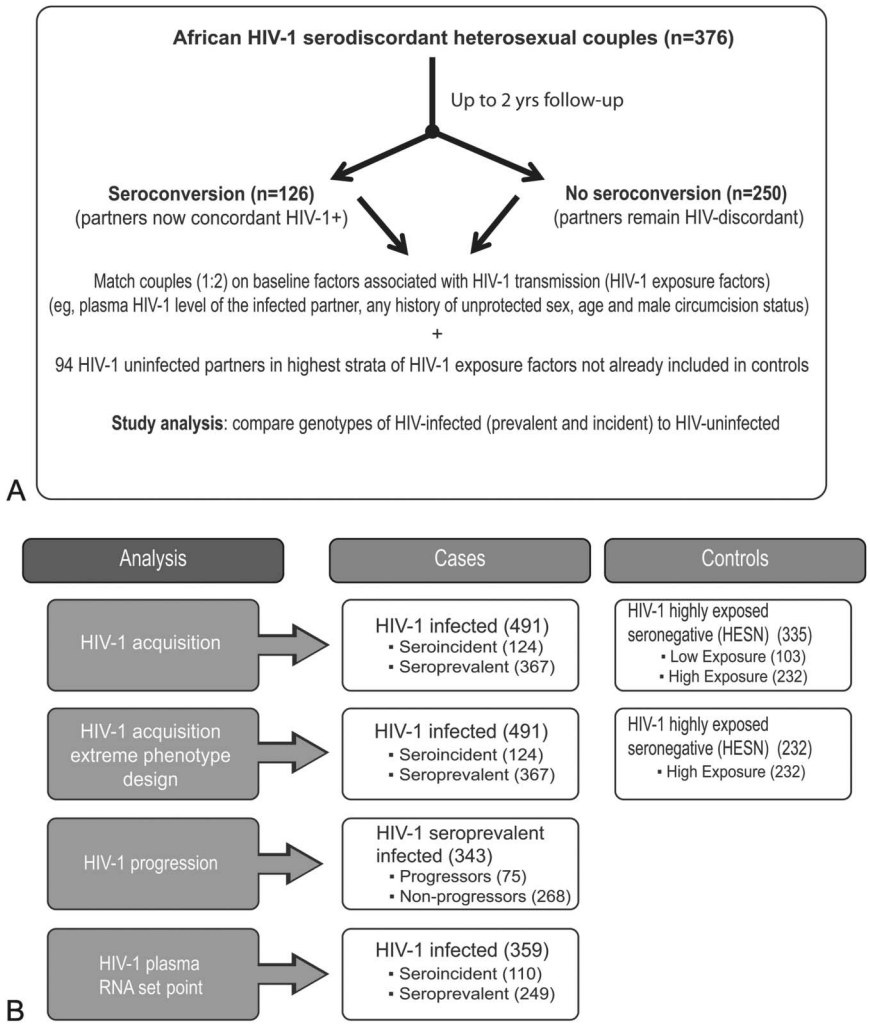

FIGURE 1. Selection of study population. HIV-1–uninfected partners were chosen to have a distribution of HIV-1 exposure risk scores that could have resulted in transmission (A). Four analyses were performed looking at SNP associations with (1) HIV-1 acquisition; (2) HIV-1 acquisition using an extreme phenotype design, wherein only high-exposure highly exposed seronegative (HESN) controls were included; (3) HIV-1 disease progression; and (4) HIV-1 plasma RNA set point. Sample sizes are shown in parentheses (B).

Quality control (QC) was applied to the 493 candidate and tagSNPs and included filtering monomorphic SNPs (n = 6), those with ≥10% missing data among cases and controls (n = 31), and those violating Hardy–Weinberg equilibrium ($P < 0.001$, n = 6) in controls. No SNPs were excluded based on MAF. Because these categories were not mutually exclusive, a total of 39 SNPs were excluded leaving 459 SNPs, including 442 tagSNPs and 17 candidate SNPs for association analysis (see **Table S1, Supplemental Digital Content**, http://links.lww.com/QAI/A506).

We also excluded individuals from our analysis who failed QC metrics. Fourteen individuals (8 cases and 6 controls) failed a sex check that compared reported sex with genotypic sex based on X chromosome heterozygosity, which is likely because of sample processing errors. Six individuals (3 cases and 3 controls) displayed genotypic missingness ≥10%. These 20 individuals were excluded from further analysis. Thus, 826 study participants (491 HIV-1–seropositive cases and 335 seronegative control individuals) were included in the analysis.

TABLE 1. Candidate Genes (n = 51) Were Selected From 6 Categories of Proteins Interacting With HIV-1 During Replication

Category	No. Genes	SNPs*	QC-Passed SNPs	SNPs Included on Lingappa et al,[4] GWAS
Capsid formation	1	4	4	2
Cell entry	10	115	111	66
Intracellular restriction	7	47	39	27
Intracellular trafficking	4	40	39	19
Integration	3	62	58	25
Other (eg, small interfering RNA knockdown)	26	225	208	132
Total	51	493	459	271

*SNPs include tagSNPs and candidate SNPs. For tagSNPs, only one SNP per haplotype bin was selected to be included on the custom Oligo Pooled Assay.

Statistical Analyses

SNP associations with HIV-1 acquisition risk were performed using standard logistic regression analysis in PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). Odds ratios (ORs) were calculated for the minor

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

allele, defined based on the allele frequencies in this study population. Sex, age, and population stratification using the first 5 PCs were included as covariates in the model. For tagSNP associations, Wald *P* values were subjected to Bonferroni correction for multiple (n = 442 for autosomal and X chromosome tests and n = 434 for autosomal tests) tests (with a significant threshold of $P < 0.0001$, $\alpha = 0.05$, or $P_{corr} < 0.05$). Allelic association was determined using 2×2 contingency table analysis in PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). ORs and *P* values for SNPs with cell values below 2 were calculated in R (version 3.0.1) using the median unbiased estimation, an exact method. *P* values for candidate SNPs were not corrected for multiple comparisons because these tests sought to validate previously observed associations. Although our primary analysis evaluated all cases and controls that passed QC, to maximize our ability to detect genetic associations for HIV-1 acquisition, we also used a secondary extreme phenotype design that compared all 491 seropositive cases (367 seroprevalent and 124 seroincident) with HIV-1–seronegative controls (high HESN) with a baseline exposure risk score ≥4 (n = 232), wherein values ranged from 0 (lowest exposure) to 7 (highest exposure).

In addition to testing for associations with HIV-1 acquisition, we also used linear regression to test SNPs for association with plasma HIV-1 RNA set point among prevalent and incident seropositive cases. For this analysis, plasma HIV-1 RNA set point was calculated as the average $\log_{10}$ plasma HIV-1 RNA level for each individual after excluding all plasma HIV-1 RNA levels after implementation of antiretroviral therapy (ART) or after CD4 counts dropped below 200, as previously described.[4] We also used a previously defined composite measure of HIV-1 disease progression[19] (first occurrence of death, CD4 count <200 cells/mm^3, or of ART) to test for SNP associations with disease progression using Cox proportional hazard regression and Kaplan–Meier curves. In addition to including sex, age, and population stratification (using the first 5 PCs) as covariates in these models, we also adjusted for the study drug arm (acyclovir versus placebo) from the Partners in Prevention HSV/HIV Transmission study because acyclovir was demonstrated to reduce plasma HIV-1 RNA[5] and disease progression.[19] We adjusted for these covariates in both the disease progression and plasma HIV-1 set point analyses.

RESULTS

Subject Characteristics

Of the 826 participants included in this analysis, 491 were HIV-1 infected and 335 were HESN controls. Among the HIV-1–infected cases, 367 were seroprevalent and 124 seroincident. Among the 335 HESN controls, 232 were classified as having high HIV-1 exposure (high HESN). Compared with controls with low HIV-1 exposure, high-exposure HESN had HIV-1–infected partners with higher plasma HIV-1 RNA levels [5.16 $\log_{10}$ copies/mL (c/mL) versus 4.30 $\log_{10}$ c/mL in low exposure, $P < 0.001$]. However, levels of unprotected sex were similar in the 2 groups (24% in low exposure versus 26% in high exposure, $P = 0.9$). Subject characteristics are provided in Table 2. Our decision to aggregate across study sites was based on our previous data showing that transmission risk and plasma HIV-1 RNA level are not modified by HIV-1 subtype (an important confounder for geographical study site) and our use of PCs to control for population stratification. Nevertheless, there was variation in MAFs among study sites (see **Table S20, Supplemental Digital Content**, http://links.lww.com/QAI/A506). The extent to which unappreciated geographical factors may have confounded our analysis is best evaluated through replication of our findings in additional cohorts.

SNP Associations With HIV-1 Acquisition

We tested 434 autosomal tagSNPs and 17 candidate SNPs for associations with HIV-1 acquisition risk using dominant, recessive, and additive models of inheritance (see **Tables S3–S5, Supplemental Digital Content**, http://links.lww.com/QAI/A506). After adjusting for multiple comparisons, no significant tagSNP associations with HIV-1 infection were observed when comparing

TABLE 2. Subject Characteristics

Variable	Cases Seroprevalent (n = 367)	Cases Seroincident (n = 124)	Controls, High Exposure (n = 232)	Controls, Low Exposure (n = 103)
East African, %	84 (81–88)	81 (75–88)	87 (82–91)	87 (81–94)
Age, yrs	33 ± 8.4	32 ± 9.7	33 ± 10	33 ± 11
Female, %	54 (49–60)	45 (36–54)	50 (44–56)	41 (31–50)
Male circumcision, %	31 (24–38)	43 (31–54)	33 (24–41)	41 (29–53)
HIV-1 RNA set point ($\log_{10}$ c/mL)	4.75 ± 0.72	4.31 ± 1.04	NA	NA
Any unprotected sex with partner	32 (27–36)	46 (37–55)	50 (43–56)	26 (18–35)
HIV-1 exposure score	NA	4.0 ± 1.7	5.4 ± 0.9	1.93 ± 1.1
Enrollment CD4$^+$ count (cells/mm^3)	479 ± 225	NA	NA	NA
Enrollment CD4 <200, %	2 (0.5–3)	NA	NA	NA
Death during follow-up, %	4 (2–5)	NA	NA	NA

Values are reported for study subjects included in the HIV-1 seroconversion case–control analysis. All individuals who failed the QC analysis have been removed. All characteristics are reported as the mean ± SD or 95% CI for proportions.

NA, not available.

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

HIV-1–infected cases with all HESN controls. When restricting our analysis to high-HESN controls, a single tagSNP *SMARCB1* (also known as *INI1*) rs6003904 was significantly associated with HIV-1 acquisition in a dominant model (OR = 2.11, 95% confidence interval (CI): 1.50 to 3.98, P_{corr} = 0.01) (Fig. 2, Table 3; see **Tables S8–S11, Supplemental Digital Content**, http://links.lww.com/QAI/A506). One candidate SNP, *PPIA* rs8177826, was associated with decreased risk of HIV-1 infection in an allelic model that included all HESN (OR = 0.15, 95% CI: 0.01 to 0.98, P = 0.04) (Table 3; see **Table S7, Supplemental Digital Content**, http://links.lww.com/QAI/A506). This SNP had low prevalence with an MAF = 0.004; the minor (G) allele was present in 0.1% of HIV-1–seropositive cases compared with 0.7% of controls. After restricting our analysis to high-HESN controls, this association was modestly strengthened (OR = 0.13, 95% CI: 0.01 to 0.94, P = 0.04) (Table 3; see **Table S12, Supplemental Digital Content**, http://links.lww.com/QAI/A506).

SNP Associations With Plasma HIV-1 RNA Set Point

The mean plasma HIV-1 RNA set point among all HIV-1–seropositive cases was 4.61 $\log_{10}$ c/mL. When stratified by stage of HIV-1 infection (ie, seroprevalent, chronically infected, or seroincident, recent infection), seroprevalent cases exhibited higher mean plasma HIV-1 RNA set point (4.75 $\log_{10}$ c/mL) compared with seroincident cases (4.31 $\log_{10}$ c/mL). Among all the HIV-1–seropositive cases, we identified a single candidate SNP association with plasma HIV-1 RNA set point for the *CCR5* 5′ UTR SNP rs1799988 (dominant: β = 0.20, 95% CI: 0.01 to 0.40, P = 0.04). Using a dominant model of inheritance, the CC and CT genotypes displayed a higher mean plasma HIV-1 set point (4.65 $\log_{10}$ c/mL) than the TT homozygote (4.46 $\log_{10}$ c/mL). However, there were no associations with plasma HIV-1 levels in the analyses stratified by seroprevalent or seroincident infection, perhaps because of the lower statistical power.

SNP Associations With HIV-1 Progression

Four candidate SNPs were significantly associated with HIV-1 progression to $CD4^+$ T-cell counts <200 cells per cubic millimeter, initiation of ART, or death among participants with baseline CD4 counts >200. SNP rs601338 in *FUT2* [hazard ratio (HR) = 1.49, 95% CI: 1.05 to 2.11, P = 0.02] was associated with increased progression, and SNP rs5030737 in *MBL2* (HR = 0.22, 95% CI: 0.05 to 0.92, P = 0.04) exhibited protective effects. Two *CCL3* SNPs, rs35511254 (HR = 0.03, 95% CI: 0.004 to 0.27, P = 0.001) and rs1130371 (HR = 0.04, 95% CI: 0.004 to 0.30, P = 0.002), were also associated with protection (see **Table S18, Supplemental Digital Content**, http://links.lww.com/QAI/A506). However, the MAF for both SNPs was low (MAF = 0.001), with only 2 out of 1648 alleles. These 2 SNPs were not in strong LD (r^2 = 0.25).

DISCUSSION

Sixty-five percent of the ~34 million people worldwide infected with HIV-1 live in sub-Saharan Africa.[2] Yet, our knowledge of genetic risk factors for HIV-1 outcomes in African populations is limited. Our results are an important first step toward extending genotype–phenotype association studies of HIV-1 phenotypes to African populations and provide support to the idea that host genetic variants modify HIV-1 acquisition risk and disease progression. Importantly, our analysis benefited from the HIV-1–serodiscordant couples' study design by incorporating epidemiological quantification of HIV-1 exposure to capture a population of HIV-1–uninfected individuals with similar HIV-1 exposures as HIV-1 seroconverters and to assess the interaction of this exposure with protection by specific SNPs.[3]

We identified *PPIA* rs8177826 as a variant with protective association with HIV-1 acquisition. Notably, we also found that the magnitude of this association increased when we restricted the analysis to comparing cases with control individuals with the highest levels of HIV-1 exposure. *PPIA* encodes the protein cyclophilin A, which has been shown to bind to HIV-1 capsid shortly after viral membrane fusion.[20] Cyclophilin A facilitates virus uncoating, thereby promoting replication and infectivity.[13] *PPIA* SNP rs8177826 has been previously associated with progression to AIDS in African Americans with 1 or 2 copies of the minor (G) allele associated with more rapid $CD4^+$ T-cell loss.[21] In contrast to the study of disease progression by An et al,[21] we did not detect an effect of this SNP on progression or viral levels. In addition, we also found this same variant to be associated with decreased risk of acquisition. It is possible to have divergent

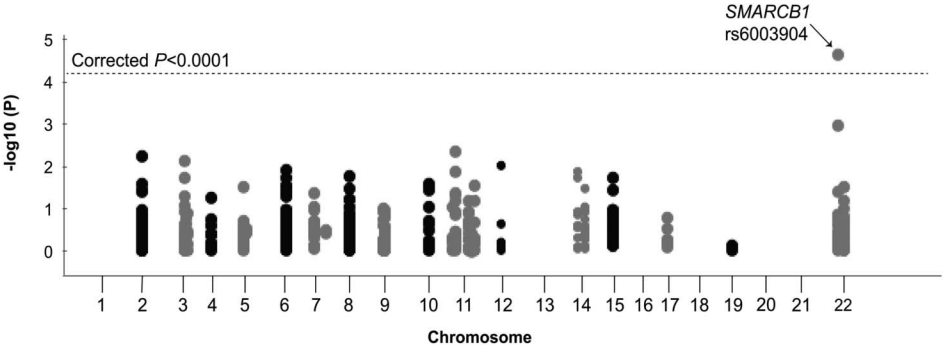

FIGURE 2. Manhattan plot of $-\log_{10}$ (P values) for the HIV-1 tagSNP acquisition analysis restricted to seronegative controls with an exposure risk score ≥4 using a dominant model of inheritance. SNP position is shown along the x axis. The dashed line indicates the Bonferroni-corrected P value threshold for 442 tests (P < 0.0001, α = 0.05). The *SMARCB1* SNP rs6003904 was significantly associated with HIV-1 transmission in this exposure-restricted cohort.

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

TABLE 3. Summary of Significant SNP Associations With HIV-1 Transmission, Plasma RNA Set Point, and HIV-1 Disease Progression

Disease Model*	Inheritance Model	SNP	Gene	Chr	Function	Minor Allele	MAF	OR/β†/HR	P‡	P_{corr}
HIV-1 transmission										
All exposure scores	Allelic	rs8177826	*PPIA*	7	5′ UTR	G	0.004	0.15	0.048	
Exposure scores ≥ 4	Allelic	rs8177826	*PPIA*	7	5′ UTR	G	0.004	0.13	0.042	
	Dominant	rs6003904	*SMARCB1*	22	Intron	G	0.491	2.11	—	0.01
Plasma RNA set point	Dominant	rs1799988	*CCR5*	3	5′ UTR	C	0.488	0.2	0.044	
Disease progression	NA	rs601338	*FUT2*	19	Nonsense	A	0.433	1.49	0.024	
	NA	rs5030737	*MBL2*	10	Missense	T	0.005	0.22	0.038	

ORs are reported for the minor allele.

*Four disease models were used: HIV-1 transmission models compared HIV-1–seropositive cases with HIV-1–seronegative controls either without regard to controls' exposure score (all exposure scores) or restricted to high-exposure controls (exposure score ≥ 4); the plasma HIV-1 RNA set point model evaluated SNP associations with plasma HIV-1 RNA set point for all seropositive cases; disease progression model assessed SNP associations with HIV-1 progression.

†ORs are reported for HIV-1 transmission and HIV-1 transmission exposure score ≥4. β Coefficients are reported for RNA set point. HRs are reported for disease progression. Dominant analyses were corrected for population stratification, sex, and age with the plasma HIV-1 set point model corrected for Partners in Prevention HSV/HIV study arm.

‡TagSNP *P* values were corrected for multiple tests using the Bonferroni correction. No correction was applied for candidate SNPs.

Chr, chromosome; NA, not available.

genetic associations with HIV-1 acquisition and progression as host mutations may simultaneously influence target cell susceptibility while enhancing effective immune responses to contain the infection. These differences in results may also be attributable to cohort differences, different measures of disease progression, or the relatively short follow-up duration (2 years) for our study cohort.

We also identified an association between *SMARCB1* rs6003904 and increased risk of HIV-1 acquisition. The SMARCB1 protein is thought to interact with HIV-1 integrase during the process of HIV-1 integration into the host genome.[22,23] Recently, in vitro studies have demonstrated that a truncated 111-amino acid mutant fragment of *SMARCB1* also interferes with HIV-1 trafficking, inhibiting the early stages of HIV-1 assembly before particle formation and budding.[24] However, rs6003904 is an intronic SNP whose haplotype bin does not encompass this 111-amino acid mutant fragment of *SMARCB1*. Thus, a mechanistic role for this SNP in regulating HIV-1 replication remains undefined.

Our analysis identified a single SNP (rs1799988) in the *CCR5* 5′ UTR associated with plasma HIV-1 set point. A *CCR5* open reading frame deletion variant, *CCR5Δ32*, provides almost complete resistance to HIV-1 infection among homozygous carriers from European populations[1] but has not been found to date in sub-Saharan African populations. The rs1799988 promoter region variant has been identified previously in both European American and African American cohorts, showing associations with set point and disease progression between both.[25,26] Additionally, in vitro work has shown that this polymorphism influences both *CCR5* expression and the magnitude of HIV-1 propagation.[27] Furthermore, recent studies have reported associations between *CCR5* and *CCR5/CCR2* haplotypes with control of HIV-1 in Zambians.[28] Our results support these previous findings and may extend to a wider range of African populations insofar as participants with this variant were identified from Eastern and Southern African sites in our analysis, with the C allele frequency among Eastern and Southern Africans, 51% and 40%, respectively.

Finally, we also identified 4 candidate SNPs having significant associations with our measure of HIV-1 disease progression, *FUT2* rs601338, *MBL2* rs5030737, and *CCL3* rs35511254 and rs1130371. However, the MAF of the 2 *CCL3* SNPs (MAF = 0.001) is too low to consider this statistical association as meaningful. The *FUT1* and *MBL2* variants have been shown previously to be associated with HIV-1 disease progression in European ancestry cohorts.[26,29] Again, our results underscore that these associations may be more broadly relevant to sub-Saharan Africans. All these genes have known functions related to innate immune response (*MBL2*) or biosynthesis of innate response factors (*FUT2* for complement factor H and blood group antigens). Both variants are predicted to result in protein sequence alterations: *FUT2* rs601338 introduces a stop codon at position 154/343 of the protein, whereas *MBL2* rs5030737 is a missense mutation. A previous study in a cohort of European ancestry found that individuals homozygous for the *FUT2* rs601338 nonsense mutation were associated with reduced levels of blood group antigens and reduced HIV-1 disease progression.[29] Similarly, mannose-binding lectin (MBL) binds envelope glycoproteins and has been implicated in HIV-1 pathogenesis, including a study of West Africans showing *MBL2* variants associated with susceptibility to infection.[30,31] The rs5030737 variant associated with disease progression in our cohort introduces an arginine to cysteine amino acid change in exon 1. This missense mutation has been shown to reduce the amount of functional MBL and has been associated with increased risk of infection among European Americans and Brazilians.[26,32,33]

In line with the previous genome-wide and targeted SNP HIV-1 association studies, our study design combined all HIV-1–positive cases into a single case group, thus maximizing the power of our cohort.[4,26,34,35] However, given the variety of factors that potentially modify the risk of heterosexual transmission, it is important to consider differences in exposure risk between the seroincident and the seroprevalent infections that may have affected our results. To look into this potential caveat, we explored associations between HIV-1

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

seroconversion and SNP genotype in an analysis restricted to HIV-1–seroincident infections. The point estimates for the variants presented in the larger analysis (seroincident and seroprevalent infections combined) were similar to the point estimates for the seroincident-restricted analysis (see **Tables S21 and S22, Supplemental Digital Content**, http://links.lww.com/QAI/A506). However, owing to the restricted sample size (124 QC-passed HIV-1–seroincident infections), no associations within this seroincident subset reached statistical significance. Taken together, this analysis suggests that combining seroincident and seroprevalent into a single HIV-1 seroconverter group increased the power of our analysis without modifying the relevant associations.

Our targeted evaluation of host genes associated with HIV-1 replication identified variants significantly associated with HIV-1 acquisition and plasma HIV-1 set point despite our previous GWAS in the same cohort that failed to identify significant associations with these traits. Several SNPs with significant statistical associations with HIV-1 outcomes in this study were not included on the GWAS platform (*PPIA* rs8177826) or were not tested for an association with disease progression in the GWAS (*FUT2* rs601338 and *MBL2* rs5030737). A single SNP, *CCR5* SNP rs1799988, was included on the GWAS SNP platform but failed QC and was excluded from the analysis. A second SNP, *SMARCB1* SNP rs6003904, was significantly associated with HIV-1 disease outcomes and found to be not associated with HIV-1 outcomes in the GWAS. However, the GWAS association was in the same direction as the association identified here ($\beta = 2.86$ versus $\beta = 2.11$, respectively), but it did not reach genome-wide significance. There are 3 major reasons that could explain these discrepant results. First, the GWAS was powered to detect variants with strong effect sizes (eg, RR $>$ 3.2 for a variant with a 5% MAF) and therefore had more limited power to detect variants with a lower risk association, such as *SMARCB1*. Second, the GWAS analysis was performed using genotyping data that were primarily supported by common variants identified in populations of European ancestry. The targeted gene approach described here used tagSNPs and candidate SNPs present in Africans, thus giving them a higher a priori relevance to this cohort. Third, because the GWAS analysis evaluated a limited number of common variants across the genome, it depended on LD between genotyped and nongenotyped variants to broadly capture gene–disease associations.[36] LD is low in many African populations[37]; thus, the GWAS approach has a significant risk of missing important gene–disease associations in these populations.[38] Despite the limitations of the GWAS to detect associations with HIV-1 outcomes given the size of this cohort, the associations reported will require confirmation through replication in other cohorts. Furthermore, although tagSNP variants were initially selected to have MAFs greater than 5%, some had rare prevalence in this population making replication of these findings of even greater importance. Many of these considerations also apply to GWASs that have specifically failed to identify variants influencing HIV-1 acquisition in European[35] and African[39] cohorts. These studies of HIV-1 acquisition also differed from our analysis insofar as they did not explicitly quantify HIV-1 exposure among the population of HIV-1–exposed uninfected individuals studied.

In summary, our study supports the view that variants in genes that interact with HIV-1 during viral replication may modify HIV-1 outcomes in sub-Saharan Africans. If replicated, such findings may identify cellular pathways that can be exploited for future preventive and therapeutic interventions particularly relevant to African populations.

REFERENCES

1. Smith MW, Dean M, Carrington M, et al. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science.* 1997;277:959–965.
2. UN Joint Program on HIV/AIDS. *Global Report: UNAIDS Report on the Global AIDS Epidemic: 2010*. Available at: http://www.refworld.org/docid/4cfca9c62.html. Accessed August 1, 2013.
3. Mackelprang RD, Baeten JM, Donnell D, et al. Quantifying ongoing HIV-1 exposure in HIV-1-serodiscordant couples to identify individuals with potential host resistance to HIV-1. *J Infect Dis.* 2012;206:1299–1308.
4. Lingappa JR, Petrovski S, Kahle E, et al. Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. *PLoS One*. 2011;6:e28632.
5. Celum C, Wald A, Lingappa JR, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. *N Engl J Med.* 2010;362:427–439.
6. Campbell MS, Mullins JI, Hughes JP, et al. Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial. *PLoS One*. 2011;6:e16986.
7. Kahle EM, Campbell M, Lingappa J, et al. HIV-1 subtype C is not associated with higher risk of heterosexual transmission: a multinational study among African HIV-1 serodiscordant couples. Presented at: AIDS 2012; July 22-27, 2012; Washington, DC.
8. Campbell MS, Kahle EM, Celum C, et al. Plasma viral loads during early HIV-1 infection are similar in subtype C- and non-subtype C-infected African seroconverters. *J Infect Dis.* 2013;207:1166–1170.
9. Lingappa JR, Dooher JE, Newman MA, Kiser PK, Klein KC. Basic residues in the nucleocapsid domain of Gag are required for interaction of HIV-1 gag with ABCE1 (HP68), a cellular protein important for HIV-1 capsid assembly. *J Biol Chem.* 2006;281:3773–3784.
10. Zimmerman C, Klein KC, Kiser PK, et al. Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature*. 2002;415:88–92.
11. Jayappa KD, Ao Z, Yao X. The HIV-1 passage from cytoplasm to nucleus: the process involving a complex exchange between the components of HIV-1 and cellular machinery to access nucleus and successful integration. *Int J Biochem Mol Biol.* 2012;3:70–85.
12. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond B Biol Sci.* 2009;364:675–687.
13. Franke EK, Yuan HE, Luban J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature*. 1994;372:359–362.
14. Berthoux L, Sebastian S, Sokolskaja E, et al. Cyclophilin A is required for TRIM5{alpha}-mediated resistance to HIV-1 in Old World monkey cells. *Proc Natl Acad Sci U S A*. 2005;102:14849–14853.
15. Blot V, Perugi F, Gay B, et al. Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. *J Cell Sci.* 2004;117(pt 11):2357–2367.
16. Strack B, Calistri A, Craig S, et al. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell*. 2003;114:689–699.
17. Bushman FD, Malani N, Fernandes J, et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog*. 2009;5:e1000437.
18. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–909.

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

19. Lingappa JR, Baeten JM, Wald A, et al. Daily acyclovir for HIV-1 disease progression in people dually infected with HIV-1 and herpes simplex virus type 2: a randomised placebo-controlled trial. *Lancet*. 2010;375:824–833.
20. Luban J, Bossolt KL, Franke EK, et al. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell*. 1993;73:1067–1078.
21. An P, Wang LH, Hutcheson-Dilks H, et al. Regulatory polymorphisms in the cyclophilin A gene, PPIA, accelerate progression to AIDS. *PLoS Pathog*. 2007;3:e88.
22. Brown P. Integration. In: Coffin JM, Hughes SH, Varmus HE, eds. *Retroviruses*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997:161–203.
23. Yung E, Sorin M, Pal A, et al. Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. *Nat Med*. 2001;7: 920–926.
24. Cano J, Kalpana GV. Inhibition of early stages of HIV-1 assembly by INI1/hSNF5 transdominant negative mutant S6. *J Virol*. 2011;85: 2254–2265.
25. Martin MP, Dean M, Smith MW, et al. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science*. 1998;282: 1907–1911.
26. Fellay J, Ge D, Shianna KV, et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet*. 2009;5:e1000791.
27. Salkowitz JR, Bruse SE, Meyerson H, et al. CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. *Clin Immunol*. 2003;108:234–240.
28. Malhotra R, Hu L, Song W, et al. Association of chemokine receptor gene (CCR2-CCR5) haplotypes with acquisition and control of HIV-1 infection in Zambians. *Retrovirology*. 2011;8:22.
29. Kindberg E, Hejdeman B, Bratt G, et al. A nonsense mutation (428G→A) in the fucosyltransferase FUT2 gene affects the progression of HIV-1 infection. *AIDS*. 2006;20:685–689.
30. Worthley DL, Bardy PG, Mullighan CG. Mannose-binding lectin: biology and clinical implications. *Intern Med J*. 2005;35:548–555.
31. Mombo LE, Lu CY, Ossari S, et al. Mannose-binding lectin alleles in sub-Saharan Africans and relation with susceptibility to infections. *Genes Immun*. 2003;4:362–367.
32. Garred P, Madsen HO, Balslev U, et al. Susceptibility to HIV infection and progression of AIDS in relation to variant alleles of mannose-binding lectin. *Lancet*. 1997;349:236–240.
33. da Silva GK, Guimaraes R, Mattevi VS, et al. The role of mannose-binding lectin gene polymorphisms in susceptibility to HIV-1 infection in Southern Brazilian patients. *AIDS*. 2011;25:411–418.
34. 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.
35. McLaren PJ, Coulonges C, Ripke S, et al. Association Study of Common Genetic Variants and HIV-1 Acquisition in 6,300 Infected Cases and 7,200 Controls. *PLoS Pathog*. 2013;9:e1003515.
36. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6:95–108.
37. Conrad DF, Jakobsson M, Coop G, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet*. 2006;38:1251–1260.
38. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet*. 2010;11: 149–160.
39. Petrovski S, Fellay J, Shianna KV, et al. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS*. 2011;25:513–518.

APPENDIX

Partners in Prevention HSV/HIV Transmission study team (for genomic studies): University of Washington Coordinating Center and Central Laboratories, Seattle, WA: C.C. (principal investigator), Anna Wald (protocol cochair), J.R.L. (medical director), Jared M. Baeten, Mary S. Campbell, Lawrence Corey, Robert W. Coombs, James P. Hughes, Amalia Magaret, M.J.M., Rhoda Morrow, and J.I.M.

Study site principal investigators and study coordinators relevant to this analysis: Cape Town, South Africa (University of Cape Town): David Coetzee; Eldoret, Kenya (Moi University, Indiana University): Kenneth Fife, Edwin Were; Gaborone, Botswana (Botswana Harvard Partnership): Max Essex, Joseph Makhema; Kampala, Uganda (Infectious Disease Institute, Makerere University): Elly Katabira, A.R.; Kisumu, Kenya (Kenya Medical Research Institute, University of California San Francisco): Elizabeth Bukusi and Craig Cohen; Moshi, Tanzania (Kilimanjaro Christian Medical College, Harvard University): Saidi Kapiga and Rachel Manongi; Nairobi, Kenya (University of Nairobi, University of Washington): Carey Farquhar, G.J.-S., and James Kiarie; Orange Farm, South Africa (Reproductive Health Research Unit, University of the Witwatersrand): Sinead Delany-Moretlwe and Helen Rees; Soweto, South Africa (Perinatal HIV Research Unit, University of the Witwatersrand): G.D.B., Glenda Gray, and James McIntyre; Thika, Kenya (University of Nairobi, University of Washington): N.R.M.

© 2014 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.