Genome-Wide Analyses Reveal Gene Influence on HIV Disease Progression and HIV-1C Acquisition in Southern Africa

Wen Xie,^{1,2} Denis Agniel,³ Andrey Shevchenko,⁴ Sergey V. Malov,^{4,5} Anton Svitin,⁴ Nikolay Cherkasov,⁴ Marianna K. Baum,⁶ Adriana Campa,⁶ Simani Gaseitsiwe,^{1,2} Hermann Bussmann,^{1,2} Joseph Makhema,^{1,2} Richard Marlink,^{1,2} Vladimir Novitsky,^{1,2} Tun-Hou Lee,¹ Tianxi Cai,⁷ Stephen J. O'Brien,^{4,8} and M. Essex^{1,2}

Abstract

Sub-Saharan Africans infected with HIV-1C make up the largest AIDS patient population in the world and exhibit large heterogeneity in disease progression before initiating antiretroviral therapy. To identify host variants associated with HIV disease progression, we performed genome-wide association studies on a total of 556 treatment-naïve HIV-infected individuals in Botswana. We characterized the pattern of HIV disease progression using a novel functional principal component analysis, which can better capture longitudinal CD4 and viral load (VL) trajectories. Two single-nucleotide polymorphisms (SNPs) near HCG22 (chr6, peak variant rs2535307, combined $p = 3.72 \times 10^{-7}$, minor allele as risky allele) and CCNG1 (chr5, peak variant kgp22385164, combined $p = 1.88 \times 10^{-6}$, minor allele as risky allele) were significantly associated with CD4 and VL dynamics. Inspection of SNPs in these gene regions in a third Botswana cohort (using GWATCH) also revealed a strong association of HCG22 with HIV-1C acquisition, suggesting that this region is associated with infection as well as disease progression. Our study uncovered two genetic regions that are significant and have specific effects on HIV-1C acquisition or progression in sub-Saharan Africans, and the result suggested new potential targets for AIDS prevention and treatment. In addition, our results also indicate the possibility of using genetic markers as HIV disease progression indicators in sub-Saharan Africans to prioritize fast progressors for antiretroviral treatment.

Keywords: restriction factors, natural hosts, HIV transmission, HIV clinical outcomes research, HIV/AIDS pathogenesis

Introduction

HIV-1C ACCOUNTS FOR over half of HIV-1 infections worldwide, several-fold more than any other subtype, including HIV-1B.¹ Some 69.1% people living with HIV resided in sub-Saharan Africa in 2012, where HIV-1 subtype-C dominates² and ~60% of infected people are women.³ Although several genes, including HLA-B*5701, other HLA alleles,⁴⁻⁹ ring finger protein 39 (RFN39), and zinc ribbon

domain-containing 1 (ZNRD1),⁴ have been found to be associated with HIV disease progression in European males infected by HIV-1B,^{4,5,10-14} little is known about sub-Saharan Africans' host genetic restriction to HIV-1C.

Unfortunately, there have been no genome-wide association study (GWAS) reports on HIV-1C-infected African populations. The relatively large genetic diversity of the southern African population,¹⁵ combined with the fact that treatment-naïve Botswana patients display significant heterogeneity in natural

¹Harvard T.H. Chan School of Public Health AIDS Initiative, Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

²Botswana Harvard AIDS Institute, Gaborone, Botswana.

³Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts.

⁴Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia.

⁵Department of Mathematics, St. Petersburg Electrotechnical University, St. Petersburg, Russia.

⁶Department of Dietetics and Nutrition, Robert Stempel College of Public Health & Social Work, Florida International University, Miami, Florida.

⁷Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

⁸Oceanographic Center, Nova Southeastern University, Ft. Lauderdale, Florida.

disease progression, raises the possibility that the host genetic makeup of individuals may influence their varying ability to maintain high CD4 counts and low viral load (VL).

Association methods using a single estimate of VL set point or early CD4 counts cannot fully describe the course of disease over time. Linear approximation of CD4 and VL trajectory fails to capture variations in disease progression and may leave out single-nucleotide polymorphisms (SNPs) that are associated with variation of trajectory.

In our study, we employed functional principal component analysis (fPCA)^{16–19} to concisely characterize the variability in the trajectories of CD4 and VL. The fPCA approach allows us to identify association with trajectories of progression while incorporating variation beyond the limitations of linear fit models.

We identified two significant host genetic restriction factors for CD4 counts and viral replication. Employing GWASs (2.5 M SNP arrays) in three independent Botswana cohorts, our study is the first to identify significant variants associated with HIV-1C infection and progression in a southern African population. This is also the first study to model HIV/AIDS disease progression trajectories using fPCA based on longitudinal observations combining CD4 counts and VL and the first to employ GWATCH analysis of associated gene regions in gene discovery replication.

Methods

Cohort and patient eligibility

BHP010 (Botsogo and Botsogo Extension Study) was a natural history, observational, prospective cohort study recruited from clinics in Gaborone, which comprised HIV-1C-infected individuals with starting CD4 cell counts above 400 cells per μ l and not yet qualified for the Botswana HAART treatment program. Anyone below 18 years, those with active

AIDS-defining illness requiring the initiation of HAART, the presence of an AIDS-related malignancy, or pregnant or breast-feeding women who had been previously exposed to HAART were excluded. Participants returned to the clinic at 3-month intervals with an additional visit 1 month after enrollment and were followed up for up to 255 weeks (Supplementary Fig. S1A; Supplementary Data are available online at www.liebertpub.com/aid). The mean sampling time for CD4 was 91 weeks. Subjects were censored for the end of the study (216/456) or enrollment on HAART treatment (121/456). Patients still alive and not qualified on HAART may then be enrolled in other studies (28/456). Participants who could not be contacted after they had missed three consecutive visits were considered lost to follow-up (91/456). Of the 456 patients enrolled, 370 were women. The age distribution of BHP010 is shown in Supplementary Figure S1B. In 2008, a subset of 330 participants of the Botsogo study were consented and enrolled in the host genetics study. Specimens with sufficient DNA were available for 294 enrolled individuals (Table 1). In this subset, the age and gender distribution are similar to the whole cohort, consistent with a nonbiased sampling.

BHP011 (the Dikotlana Study) was a randomized, multi-factorial, double-blind placebo-controlled trial to determine the efficacy of micronutrient supplementation (supplementation of multivitamin, selenium, or both) in improving immune function and preventing early mortality of HIV-1C infection.²⁰ It comprised 876 treatment-naïve patients with CD4 higher than 350 cells/ μ l. Six hundred thirty-eight of 876 patients were women. The age distribution of BHP011 is shown in Supplementary Figure S1B. DNA samples were collected at around 2 years postenrollment and 364 samples were available (Table 1). In this subset, the age and gender distribution were similar to the whole cohort, consistent with a nonbiased sampling (Supplementary Table S1B; Supplementary Fig. S1E).

TABLE 1. SAMPLE AND SINGLE-NUCLEOTIDE POLYMORPHISM FILTERING

(A) Sample filtering	Not included ^a	Included ^a
Total subjects		1,332 (456, 876)
Blood sample collected	580 (162, 418)	752 (294, 458)
DNA sample available	94 (0, 94)	658 (294, 364)
Qualify for genotyping	66 (36, 30)	592 (258, 334)
Genotyped	4 (3, 1)	588 (255, 333)
Call rate pass 0.99	7 (5, 2)	581 (250, 331)
Gender matched	6 (1, 5)	575 (249, 326)
>2 clinical observations	19 (10, 9)	556 (239, 317)
Total used		556 (239, 317)
^a Numbers in parentheses indicate result of BHP010 (left) and BHP011 (right).		
(B) SNP filtering	Dropped	Included
Total n, SNPs on Illumina Omni2.5 BeadChip HD and LCG		2,454,132
Call rate pass 0.99 in either cohort	17,799	2,436,333
MAF above cutoff in either cohort	859,517	1,576,816
Pass HWE $\alpha = 10^{-4}$	43,022	1,533,794
Autosomal	39,574	1,494,220
Consistent phenotype direction in primary and secondary cohorts for both CD4 and VL	641,331	852,889
Total used (excluding X and Y chromosomes)		852,889

HWE, Hardy–Weinberg equilibrium; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; VL, viral load.

Patients were followed up for a maximum of 169 weeks postenrollment (Supplementary Fig. S1A). They returned to clinics every 3 months to get CD4 samples and every 6 months to get VL samples. The mean CD4 sampling time was 56 weeks postenrollment.

Age distribution was similar for the two cohorts (Supplementary Fig. S1B) despite BHP010 having a slightly higher proportion of women (Table 2). EIGENSTRAT population stratification analysis²¹ demonstrated that the genetic variants of BHP010 and BHP011 are not structured (Supplementary Fig. S2B). The four treatment arms of BHP011 are also genetically similar (Supplementary Fig. S1C). A total of 588 patients were genotyped with an Illumina Omni 2.5 genotyping array (Table 1A).

For the replication study on HIV susceptibility, two replication cohorts, including 4,462 European Americans and 1,460 African Americans subjected to published GWASs,^{8,14,22} were examined using GWATCH in the HCG22 (chr6) and CCNG1 (chr5) gene regions. We also included 809 new Botswana subjects (59 males, 750 females) in the replication study. Among the 809 Botswana subjects, 437 were HIV positive and 372 were HIV negative. These individuals originated from two primary populations that have been described in detail. One was a population of pregnant women (436) involved in a maternal transmission study^{23,24}; the other (373) was a population enrolled to identify acute or early infections.^{25–29} In the GWATCH study, there were 1,829,336 SNPs included, and after using the same criteria for SNP filtering, 1,446,325 SNPs were finally tested.

Genotyping process and quality control

Genotyping was performed using the Illumina[®] Infinium[®] HumanOmni2.5M BeadChip technology. Due to discontinuation of the manufacturer's product, two kinds of chips were used: Illumina HD BeadChip and Illumina LCG BeadChip. The numbers of patients genotyped by each kind of Illumina BeadChip are shown in Supplementary Table S2.

Samples with DNA concentration higher than 38 ng/μl were qualified for genotyping. As shown in Table 1, of the

qualified 592 samples, 4 samples did not get genotyped due to discontinuation of the manufacturer's product, so we genotyped 588 patients. Among them, seven patients' call rate did not pass 0.99. We also checked the specification of gender to verify that patients with gender recorded in the clinical records as male were XY genotypes (≥1%) and those with gender recorded as female were XX genotypes. Six patients who did not match the criteria were excluded. We also only included patients who had at least two clinical laboratory observations for both CD4 and VL; 19 patients were excluded by this criterion. After the initial quality control (QC), 556 patients were selected for further analysis, 239 from BHP010 and 317 from BHP011. Detailed characteristics of the patients before QC and after each step of initial QC are shown in Supplementary Table S1. The gender ratio, median baseline VL, median baseline CD4, average baseline VL, average baseline CD4, and average age after each step of QC were all similar to the initial cohort for both BHP010 and BHP011, indicating unbiased sampling.

To avoid artifacts associated with either chemistry or clustering errors, we performed quality control on the SNPs included in the association study as follows:

(1) Infinium BeadStudio Raw Data clustering:

Due to incompatibility of LCG and HD chip data processing through Illumina Genome Studio, BHP010 on HD, BHP010 on LCG, and BHP011 on LCG were processed separately. We used the reclustering file based on the previous unpublished Malawi genotyping data we generated. We applied quality control as previously reported.⁴ After the initial reclustering, any sample that had very low intensity or a call rate lower than 99% was deleted. All SNPs on Y chromosomes were individually evaluated to make sure they were only called in male samples. All SNPs on X chromosomes were also individually evaluated to make sure no male was heterozygous. The autosomal SNPs fulfilling any one of the five criteria were screened within BeadStudio and re-clustered manually: call frequencies of less than 0.62; mean of the normalized *r*-values for the AB genotypes of less than 0.28; mean of the normalized theta angles for the AB

TABLE 2. INFORMATION ON COHORTS

	BHP010 ^a	BHP011 ^b	Combined
Participants	456 (294) ^c	878 (363)	1332 (658)
Participants in intervention arm	No	Four arms: MVs alone: 219 (93) Selenium alone: 220 (88) MVs + Selenium: 220 (94) Placebo: 219 (88)	
Percentage of women	81.0% (84.7%)	72.9% (71.8%)	75.7% (77.6%)
Median baseline CD4 (cells/mm³)	468.5 (484.5)	422.2 (424.8)	439.1 (456.0)
25 th percentile	381.0 (388.0)	335.0 (334.5)	349.2 (365.0)
75 th percentile	593.5 (605.5)	549.3 (532.1)	564.4 (589.0)
Median baseline VL (copies/ml)	12.9K (10.1K)	14.0K (14.5K)	13.6K (12.1K)
25 th percentile	2.5K (1.8K)	3.1K (3.4K)	2.8K (2.6K)
75 th percentile	46.8K (31.5K)	62.0K (62.0K)	55.6K (46.2K)

^aBHP010: natural history observational cohort.

^bBHP011: randomized MV treatment cohort.

^cNumbers in parentheses indicate numbers of available DNA specimens from patients who were consented and enrolled in the host-genetic study, or values of measured parameters in these patients.

MV, multivitamin.

genotype smaller than 0.2 or larger than 0.8; Het excess value between -1.0 and -0.3 or between 0.2 and 1.0 ; and a minor allele frequency (MAF) less than 10%. If an SNP failed the reclustering, the SNP was zeroed manually. This process led to deletion of 3,739, 6,727, and 6,263 SNPs in BHP010 genotyped by Illumina HD BeadChip, BHP010 genotyped by Illumina LCG BeadChip, and BHP011 genotyped by Illumina LCG BeadChip, respectively.

(2) MAF check, genotyping missingness, and Hardy–Weinberg Equilibrium (HWE):

After reclustering, we further excluded SNPs that had MAF lower than 5% in both BHP010 and BHP011. The rationale behind this quality control is that at least 28 people should bear the SNP in the cohort. We excluded SNPs with call rates smaller than 0.99 after reclustering. We further tested the HWE with $\alpha = 0.0001$ and only included the SNPs that passed the test. The detailed result of the QC is shown in Table 1B.

EIGENSTRAT method to perform population stratification

We applied the EIGENSTRAT as described before.⁴ Basically, we selected EIGENSTRAT axes to analyze the population ancestry. We selected SNPs based on the QC described above.

fPCA steps

We performed the fPCA^{17,18,30} to summarize trajectories of logCD4 and logVL in BHP010 and BHP011 separately. For each cohort and measure, we decomposed each person's trajectory into a population mean function (shared among all patients) and a linear combination of eigenfunction, which are the functions that explain the most variability in the observed longitudinal data. They can be thought of as equivalents of standard principal components for longitudinal data. Because the eigenfunctions are constructed such that the first eigenfunction explains the most variability in the data, the second eigenfunction explains the second most variability, and so on, only three eigenfunctions are necessary to explain more than 99% of the variation in the underlying trajectories of logCD4 or logVL. The weights or scores, ξ_{ki} , associated with eigenfunctions succinctly summarize the patient-specific trajectories.

The key advantage of fPCA is its flexibility in capturing the potential complex trajectory of disease progression using a very low dimensional functional space. We do not require restrictive assumptions about the manner of disease progression, but rather let the data determine which trajectories are important. Standard methods such as the area under the curve (AUC) or those based on linear trajectories would have limited power to detect SNPs that are associated with more complex trajectories. For example, if patients with a particular mutation tend to have more variations in CD4 over time, neither AUC nor a linear model can detect such a variant.

The population means $\mu^{(CD4)}$ and $\mu^{(VL)}$ are estimated through nonparametric smoothing, and the eigenfunctions are obtained through eigen-decomposition of the smoothed covariance matrices of logCD4 and log VL.^{31,32} We used the best linear unbiased predictor^{18,33} method to estimate the fPCA scores $\xi_{ki}^{(CD4)}$ and $\xi_{ki}^{(VL)}$. These fPCA scores (3 for logCD4, 3 for logVL) can then be used as longitudinal phenotypes that describe disease progression for each patient.

We performed fPCA³⁴ separately on BHP010 and BHP011 due to the fact that the length of follow-up differed substantially between the two cohorts.

Linear regression model to characterize CD4 and VL trajectories

We also fit a more familiar linear mixed-effects model^{35–37} to bolster the evidence from the FPCA model. The linear mixed-effects model corresponds to a simplified version of the fPCA model, where linear models are assumed for the trajectories.

The extremity of slope can separate patients with high or low CD4/VL changes over time, while the extremity of intercept can separate patients with high or low CD4/VL average level (Supplementary Fig. S3).

Testing

To characterize the strength of association between an SNP, G_i , and the trajectories of logCD4 and logVL, we obtained variance component score statistics, similar to those used previously.^{38,39} The mathematical details of the test statistic can be found in the Supplementary Data.

Under the null hypothesis, the test statistic is distributed asymptotically as a mixture of χ^2 s similar to the null distribution of the commonly used SKAT association test.³⁹ We obtain p values by estimating its null distribution and comparing it with the observed test statistic. The main analysis calculates a single p value for the overall association between an SNP and both CD4 and VL. For each of the SNPs, we calculate both the p value from each individual cohort and the combined p value based on a Fisher combination of the two p values from the two cohorts. We calculate the p values separately in each cohort because the fPCA scores may not be directly comparable across cohorts with different follow-up times. By analyzing the two cohorts separately, we will detect SNPs that are consistently important despite the differences between the cohorts. As references, we also report p values specific to CD4 and specific to VL.

We also computed permutation p values for a small set of interesting SNPs and find that the results are consistent with our analyses by approximating the null distribution with the mixture of χ^2 .

We tested the direction of the effect and eliminated SNPs that had opposite direction of effect in BHP010 and BHP011 (Supplementary Fig. S4; Supplementary Tables S3 and S4). We only considered SNPs with the same association direction in both cohorts, eliminating 641,331 discordant SNPs and leaving 852,889 SNPs (Table 1). Using the dominant model increases power, but SNPs with the lowest p values remain the same as with the regular model (Table 3 and Supplementary Table S5). We excluded X and Y chromosomes in our association analysis to avoid potential bias introduced due to gender imbalance of our cohorts. We reported the result of GWASs on autosomal SNPs before and after adjusting for sex and age (Fig. 1; Supplementary Fig. S5; Supplementary Tables S6 and Table S7). We also replicated the study with the 1,533,794 SNPs, including ones on sex chromosomes without adjusting for age and sex. We found an additional interesting SNP lying in the intron of PCDH11X (peak variant kgp22829958, combined p value = 4.24×10^{-5}) (Supplementary Table S8). After adjusting for age and sex, we got additional three new SNPs with low p value (Supplementary

TABLE 3. SINGLE-NUCLEOTIDE POLYMORPHISM ASSOCIATIONS WITH VIRAL LOAD AND CD4 DEPLETION IN BHP011 AND BHP010 COHORTS

		<i>SNPs</i>			
SNP	rs2535307	kgp22385164	rs6921557		
CHR	6p21.33	5q34	6p21.33		
Coordinates	31053741	162859382	31049022		
Gene	HCG22	CCNG1	HCG22		
Allele common/minor	T/C	T/G	T/C		
MAF	0.48	0.08	0.48		
Direction of association	Minor allele as risky allele	Minor allele as risky allele	Minor allele as risky allele		
<i>Cohort</i>	<i>Measure</i>	<i>Stat model</i>	<i>p Value</i>		
BHP011	VL and CD4	fPCA	8.30E-04	1.78E-05	4.01E-04
BHP011	VL and CD4	Linear	1.57E-05	9.25E-05	3.28E-06
BHP010	VL and CD4	fPCA	2.39E-05	6.20E-03	6.83E-05
BHP010	VL and CD4	Linear	2.59E-05	6.44E-03	1.26E-04
Combined	VL and CD4	fPCA	3.72E-07	1.88E-06	5.04E-07
Combined	VL and CD4	Linear	9.24E-09	9.14E-06	5.37E-09
BHP011	CD4	fPCA	7.52E-04	9.50E-03	3.63E-04
BHP010	CD4	fPCA	2.24E-02	4.70E-01	3.29E-02
Combined	CD4	fPCA	2.02E-04	2.86E-02	1.47E-04
BHP011	VL	fPCA	1.22E-03	1.41E-05	6.07E-04
BHP010	VL	fPCA	2.26E-05	6.11E-03	3.19E-05
Combined	VL	fPCA	5.08E-07	1.49E-06	3.63E-07

fPCA, functional principal component analysis.

Table S9). The top SNPs before adjustment were unaffected by adjustment (Supplementary Table S10 and Supplementary Fig. S3C). As both BHP010 and BHP011 were randomized trials, any association between SNPs and disease progression should not be dependent on treatment arm.

Power

Conditional on the covariance structure observed in the fPCA scores and their approximate normality, we ran simulations to determine what effect size we would have power to detect. We base our simulations on the observed standard deviations of fPCA scores in each cohort. For example, the standard deviation of the first fPCA score for log VL in the BHP010 cohort is 1.3, the standard deviation of the first score for log CD4 in that same cohort is 0.3, and that of the second score for log VL is 0.5. These values are similar in BHP011.

In 500 simulations, an SNP with an MAF of 0.1 that was truly associated with a change in all fPCA scores of 75% of a standard deviation would be detected at the level of 3×10^{-6} about 83% of the time. The scale of that change would include a shift in log CD4 of about 0.24 and in log VL of about 0.96 in BHP010, as well as smaller changes to the trajectories due to changes in the other scores.

The power increases with MAF as well as with effect size. Detecting that same effect size for an SNP with MAF 0.2 is much easier as all p values are below 3×10^{-7} . Detecting a smaller effect size with MAF of 0.2 at the level 3×10^{-6} occurs about 56% of the time.

GWATCH analysis

The GWATCH analysis was performed as described.²² Twenty-four tests were applied to test the association between genotypes and phenotypes (Supplementary Table S11).

Results

We performed GWASs on two independently recruited cohorts of HIV-1C-infected individuals from Botswana (Table 2; Supplementary Fig. S1) not subjected to anti-retroviral therapies. After stringent filtering (Table 1), 556 patients were used for association analysis. Disease progression patterns (Supplementary Fig. S2A) can be well approximated by a weighted sum linear combination of three signature fPCA trajectories (maintaining, declining, and varying, respectively) for both CD4 and VL (Supplementary Fig. S2B), which capture about 99% of the variability in the underlying CD4 or VL trajectories. A patient's longitudinal CD4/VL pattern is succinctly summarized by his/her fPCA scores, and extreme values of these fPCA phenotypes can separate patients' patterns of maintaining high versus low CD4/VL (ξ_1) or stable versus declining CD4/VL (ξ_2) (Supplementary Fig. S2C). Our method using fPCA can precisely reconstitute patient-level CD4 and VL trajectories (Supplementary Fig. S2A) with higher pseudo- R^2 when compared with the linear model (Supplementary Fig. S2D). We used the combined VL and CD4 fPCA scores as our phenotype of interest, allowing us to use information about disease progression from both VL and CD4, without making any assumption about their interrelationship.

We conducted a GWAS on BHP010 and BHP011 (556 patients) with 852,889 SNPs with consistent direction in the two cohorts (Table 1), controlling for a false discovery rate (FDR) of 10%. We employ the standard Benjamini-Hochberg procedure to perform the FDR control, which is generally less stringent than the standard family-wise error rate control based on Bonferroni correction.⁴⁰ We further excluded SNPs with low MAF (<0.05) in either BHP011 or BHP010, and SNPs that have large p values in either cohort ($p_{\text{BHP010}} > 0.01$, $p_{\text{BHP011}} > 0.0005$). There are two regions that reached significance with a p value

FIG. 1. GWAS results for 852,889 SNPs with consistent phenotype direction using combined VL and CD4 fPCA scores as phenotype under a dominant model. (A) Linkage disequilibrium around rs2535307 identified by fPCA model. rs6921557 (chr6, pos 31049022, MAF 47%, $p_{lim} = 5.39E-09$), top SNP identified by linear model (Table 3 and Supplementary Fig. S6), is 4.7 kb upstream of rs2535307 and in high LD with it. (B) Linkage disequilibrium around kgp22385164 identified by fPCA model. (C) The Manhattan plot before adjusting for age and sex shown in combined p value of BHP010 and BHP011. Six SNPs in independent regions reached genome-wide significance level. Two SNPs (in *purple circle*) are significant in primary cohort of BHP010 and remain significant after replication in BHP010. For the other four SNPs (above significance level, but not circled), although their combined p value reached significance level, they are significant in only one cohort, but not the other. (D) Genes involved in the top 100 SNPs (smallest genome-wide $p_{combined}$) identified using fPCA model (*red*) or linear mixed model (*blue*). fPCA, functional principal component analysis; GWAS, genome-wide association study; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; VL, viral load.

FIG. 1. (Continued)

$<3.28 \times 10^{-6}$ (Table 3; Fig. 1C). The most significant gene is HCG22 (peak variant rs2535307, combined $p=3.72 \times 10^{-7}$). Another significant gene is CCNG1 (peak variant kgp22385164, combined $p=1.88 \times 10^{-6}$). Both these two SNPs have a stronger association with VL than with CD4 (Table 3). HCG22 still remains significant after adjusting for sex and age (Supplementary Table S6; Supplementary Fig. S5). The CCNG1 region does not remain significant after adjusting for sex and age at a lower threshold for rejection of 0.05.

By examining the average CD4 and VL trajectories among subgroups of patients with and without minor alleles of these two significant SNPs, we observe that patients with minor allele rs2535307 or kgp22385164 have lower CD4 and higher VL (Fig. 2), suggesting both minor alleles are risk alleles (Supplementary Fig. S4). The result can be mostly explained by the separation in the first fPCA score between patients with and without the minor allele, followed by separation in the second score (Supplementary Figs. S6 and S7).

We compared the above association analysis with an analysis using a traditional linear mixed model to quantify patient-specific CD4 and VL trajectories using individual intercept and slope (Table 3). Without adjusting for age and sex, the linear mixed model results in 52 genome-wide significance SNPs (combined $p < 4.24 \times 10^{-6}$ at FDR level of 10%), each within the HCG22 region (Supplementary Table S12B, C; Supplementary Fig. S8). The linear mixed model also differentiates patients with and without the minor alleles based on their intercepts and slopes (Supplementary Fig. S9). It is worth noting that fPCA is able to pick up SNPs the linear model fails to pick up, and those SNPs are likely to be associated with nonlinear modes of variation in disease progression. Although SNPs with the lowest p values in the linear mixed model largely overlap with SNPs detected using the fPCA model (Fig. 1D), for most of the top genes shown in Supplementary Tables S12 and S7 (except for HCG22), the linear mixed model has less power, as reflected by larger p values for the same SNPs compared with the fPCA model. Hence, while we were able to detect SNPs associated with linear trajectories using the linear mixed model, we generally gained power by using the fPCA model, which enables efficient detection of SNPs associated with nonlinear trajectories.

We further examined whether the SNPs in the CCNG1 or HCG22 regions are also associated with HIV susceptibility using an independent GWAS of 809 HIV-1C-exposed study participants from the same Botswana population genotyped with the Illumina 2.5M genotyping array. We searched for associations between HIV infection and SNPs implicated in the CCNG1 (chr5) and HCG22 (chr6) gene regions, as well as 120 adjacent SNPs around both genes, applying 24 distinct SNP association tests (Supplementary Table S11). For display (Fig. 3), we used GWATCH, a software platform designed to automate, display, and replicate GWAS analyses within chromosomal regions.²² We also examined the SNPs around CCNG1 and HCG22 using GWAS results from American HIV/AIDS cohorts (Supplementary Fig. S10). No significant associations were observed for either gene region (CCNG1 or HCG22) in the three American cohorts, nor was there any significant association with HIV-1C acquisition within the CCNG1 (chr5) region with the Botswana HIV-1C infection cohort (Supplementary Fig. S10).

For the HCG22 gene region, a strong pattern of association was apparent in both the progression and infection cohorts from Botswana (Figs. 2 and 3). SNPs associated with VL/CD4

FIG. 2. Average trajectories of VL and CD4 are highly corrected with the top significant SNP using dominant model. Average trajectory of patients with 0 minor alleles (black) or at least 1 minor allele (gray) of significant loci of (A) rs2535307 (chr6, pos31053741, MAF 48%, $p_{\text{FPCA}}=3.72\text{E-}07$); (B) kgp22385164 (chr5, pos162859382, MAF=8%, $p_{\text{FPCA}}=9.14\text{E-}06$). Primary cohort of BHP011 and replication cohort of BHP010 are shown separately. Having at least one copy of minor allele of rs2535307 and kgp22385164 is associated with lower CD4 and higher VL, reflected by lower value of eigenfunction 1 of CD4 and higher eigenfunction 1 of VL compared with patients with zero minor alleles in both BHP010 and BHP011 (Supplementary Fig. S6).

spanned a 56 kb region on chromosome 6 around the HCG22 gene (Fig. 3A), while an adjacent 18 kb region—3 kb downstream, including 14 associated SNPs—showed a highly significant signal with HIV-1C infection, representing one of the top association signals across the Botswana HIV-1C infection GWASs (Fig. 3B–D). Because our search in the Botswana HIV acquisition cohort was hypothesis driven by the CCNG1 and HCG22-discovered gene regions, the obtained p values (10^{-4} – 10^{-5} ; Region B) remained highly significant after statistical correction.

The HLA region of chromosome 6, which includes the HCG22 gene, is well known to retain a strong backbone of linkage disequilibrium among adjacent SNPs. SNPs within the CD4/VL-associated region of HCG22 (Region A) and within the HIV-1C acquisition region of HCG22 (Region B) showed near-complete linkage disequilibrium within each group ($D' = 0.92$ – 1.0 , $r^2 = 0.78$ – 1.0 , for Region A; $D' = 0.97$ – 1.0 , $r^2 = 0.91$ – 1.0 , for Region B), as well as relatively strong LD ($D' = 0.3$ – 0.39 , $r^2 = 0.34$ – 0.39) between SNPs of the two regions (Supplementary Table S13). The separate, strong, adjacent, near-complete intracluster LD segments would suggest that the HCG22 region contains two separate operative SNPs, one which regulates disease progression (Region A) and another which affects HIV-1C acquisition (Region B). The parallel associations with different HIV/AIDS stages, infection, and HIV disease progression can be explained by different causal determinants from the same gene.

Discussion

In this study, we applied both fPCA and the traditional linear mixed model to summarize HIV/AIDS disease trajectory with longitudinally observed CD4 and VL of HIV-infected adults. We showed that in HIV-1C-infected treatment-naïve patients in Botswana, HCG22 is significantly associated with HIV acquisition as well as HIV/AIDS disease progression, while CCNG1 is only associated with progression, but not HIV-1C acquisition (Table 3; Fig. 3; Supplementary Fig. S10). None of the SNPs we identified have been reported to be associated with HIV/AIDS before. Using fPCA in a progression study, we were able to gain power and identify SNPs that may not be identified using the traditional linear mixed model alone.

Our peak SNP near HCG22 is in high LD with the genes within 200 kb (Fig. 1D), including PSORS1C1, CDSN, CCHCR1, TCF19, POU5F1, PSORS1C3, MUC22, and MUC21. The gene cluster of HCG22 has been reported to be associated with total white blood cell counts,⁴¹ susceptibility to follicular lymphoma,⁴² and multiple myeloma,⁴³ suggesting its role in immune regulation. Similarly to other previously identified genes (such as HLA-B/HLA-C and CCR5Δ32) that were shown to be important in both HIV acquisition¹² and HIV VL set point, HCG22 is associated with both acquisition and progression, possibly by modulating immune response. Despite its localization on chromosome 6, we did not find any significant

FIG. 3. GWATCH snapshots of HCG22 gene region association for AIDS progression and for HIV infection. GWATCH is a web-based dynamic real-time genome browser designed to discover, view, and assess hits from GWASs and WGS association studies (<http://gen-watch.org/>). The platform displays a 3-dimensional viewer above and moving along human chromosomes with baseline X-axis listing all association tests performed in the GWAS. Adjacent SNPs genotyped along each chromosome are the indices of the Y-axis. The rising blocks illustrate the resulting log of p value obtained (blocks are shown only for $p < 10^{-3}$ to avoid clutter of noise) for each SNP-test combination with color showing direction of association (red for susceptible vs. green for protective) and color intensity indicating the hazard/odds ratio (unified under the term QAS for quantitative association statistic) of each association. The combination of nonindependent proxy SNPs (which track the operative SNP by LD) and the nonindependent SNP association tests produce clusters of p value peaks in associated regions. Nonassociated regions are generally flat with random false signals appearing infrequently. (A) Region A: GWATCH 3D snapshot hovering view of AIDS progression (VL and CD4; 6 association test; 81 SNPs) association in the region of HCG22 (56 kb on chromosome 6: 31023148–31079643), illustrating a strong LD backbone of association in the HCG22 region. (B) Region B: GWATCH snapshot view of associations of the HIV infection cohort GWASs (24 association tests; Supplementary Table S4) for the HCG22 area. This 18 kb (31,002,013–31,020,152) is 3 kb upstream from HCG22 region A. (C) GWATCH view of associations for AIDS progression (region A) adjacent to region B, which associates with HIV infection. (D) GWATCH view of associations for HIV infection (region B) adjacent to region A, which associates with AIDS progression. Since our search was hypothesis driven concerning the CCNG1 and HCG22-discovered gene regions, the obtained p values remained highly significant after statistical correction.

associations or significant linkage disequilibrium tracking of our implicated SNP region for HCG22 across the HLA region. Exome sequencing around HCG22 and HLA regions will be needed to completely rule out the LD of these two regions.

CCNG1 is a cyclin linked to the p53 and p21 pathways that affect cell cycle. Its closest family member, CCNG2, was shown to promote HIV replication by shRNA screen in Jurkat cells.⁴⁴ In our study, CCNG1 is associated with HIV disease progression, but not acquisition. It is possible that CCNG1 is more linked to the pathway to control HIV intracellular replication and thus is not able to modulate HIV infection. Previously, many other SNPs were shown to be associated with disease progression, but were not successfully replicated in acquisition phenotypes, such as ZNRD1,⁴ PARD3,¹⁴ and SOX5⁴⁵ (Supplementary Table S14). It is very likely that SNPs that control HIV acquisition are also linked to low disease progression, while SNPs only associated with progression are more likely to play a role in HIV intracellular replication such as transcription control. Our association study further raised the possibility of an important role of multiple cross-talks between HIV proteins and cyclins to regulate HIV replication as well as cell cycle *in vivo* reflected by both VL and CD4 counts.

We also detected several SNPs with combined *p* values below a significant level that consistently have small *p* values in both cohorts (combined *p* value of BHP010 and BHP011 $<4.4 \times 10^{-5}$, with *p* value <0.05 in both cohorts) (Supplementary Table S8). Many genes in the list have been reported to directly regulate HIV accessory proteins to regulate HIV replication (such as PRKAA2,^{46–51} ARHGEF12,⁵² ALK,⁵³ TRIM26,^{54,55} and PCDH11X^{31,56}) or their close family members have been shown to play a role in HIV pathogenesis (such as ARL8B^{57,58} and TRIM16^{31,51,59}). In addition, we found an SNP, kgp7815029, near gene ZBTB7C (Supplementary Fig. S11), which had a small *p* value in BHP011 ($p=9.90E-07$; Supplementary Table S12A), but failed to reach significance in BHP010 ($p=0.064$; Supplementary Table S12A). This minor allele is associated with higher CD4 and lower VL (Supplementary Fig. S12).

Our study is able to suggest new SNPs that have not been reported to be associated with AIDS disease phenotypes before. This may be due to two reasons: first, we applied a new statistical method (fPCA), which does not assume linearity of the trajectory of CD4/VL. Second, the SNPs reported before were mostly shown by GWASs conducted on Caucasian men infected with HIV-1B, while we conducted our study in HIV-1C-infected southern Africans, ~60% of whom were women. Even though we replicated the previous study using their reported linear mixed method⁴ on our cohorts, the *p* values of SNPs suggested by previous GWASs that were strongly associated with VL or disease progression (rs9264942, rs9261174, and many other SNPs identified to be associated with disease progression or CD4/VL levels)^{4,60} were not among the 100,000 smallest *p* values in our study (Supplementary Table S14). Therefore, it is possible that the host genome uses different mechanisms to control HIV-1 subtype C infection, and the larger variation within the southern African population¹⁵ enables us to discover new SNPs.

However, this study has some limitations. First, our result is limited by the duration of follow-up of the cohorts and sample size; thus, to confirm the effect we have seen in this GWAS, further studies are needed. Second, seroconversion

times were not known for subjects in these cohorts. Third, because this study population was not followed to development of clinical outcomes, a specific association with AIDS disease could not be made. In conclusion, while conducting the GWAS, we found SNPs that are significantly associated with HIV disease progression, but we have not proven these are the causal SNPs. Furthermore, exome sequencing around the two regions that we uncovered and more *in vitro* experiments are needed to link our genes with HIV disease.

From the standpoint of public health, treatment prioritization is crucial for improving quality of life and averting premature death in adults and children living with HIV. According to the *Botswana National HIV & AIDS Treatment Guidelines*, patients with CD4 counts lower than 350 cells/ μ l or clinical illness were eligible for HAART treatment, patients with CD4 higher than 350 cells/ μ l and lower than 500 cells/ μ l were eligible for periodic follow-up every 3 months, and patients with CD4 higher than 500 cells/ μ l were followed every 6 months.^{61,62} However, in our cohort, we observed fast progressors who took less than half a year for CD4 to decline from above 500 cells/ μ l to less than 200 cells/ μ l, as well as slow progressors who maintained CD4 around 500 cells/ μ l for at least 4 years. For fast progressors, even short delays (less than 1 month) in initiating treatment are associated with increased mortality. Thus, eligibility of initiating HAART based on arbitrary CD4 levels is not an optimal approach. The association we established between genetic polymorphisms and disease progression may allow prediction of CD4 trajectory using genotypes as an indicator for HAART eligibility or more frequent CD4 monitoring. However, more prospective analysis is needed to validate the approach.

In summary, by identifying significant associations between variants near HCG22 and CCNG1 with HIV acquisition or progression, we suggest that HLA-mediated HIV restriction plays an important role in viral control in African genetic backgrounds, and we revealed that the cyclins may be an important element in viral replication restriction. Our results may provide important insights for the use of genetic markers to prioritize sub-Saharan HIV-1C-infected Africans, who are more susceptible to fast disease progression, for HAART or more frequent blood tests. In addition, our results also further our understanding of host restriction of viral replication, which may suggest avenues for development of drugs and vaccines.

Acknowledgments

The Botsogo and Botsogo extension studies (BHP010) were supported by the National Institute of Allergy and Infectious Diseases (R01AI052032) and by the African Comprehensive HIV/AIDS Partnership (ACHAP). M.K.B. and A.C., as well as the sample collection for the Dikotlana study (BHP011), were supported by the National Institute on Drug Abuse (R01DA016551). This study was supported by the National Institute of Allergy and Infectious Diseases (RC4 AI092715). W.X. was supported by the AIDS International Training and Research Program through the NIH Fogarty International Center (D43 TW000004). S.J.O.B., A.Sh., N.C., A.Sv., and S.V.M. were supported, in part, by the Russian Ministry of Science (Mega-Grant no. 11.G34.31.0068) and St. Petersburg State University (Genome Russia Grant no. 1.52.1647.2016).

Author Disclosure Statement

No competing financial interests exist.

References

- Geretti AM: HIV-1 subtypes: Epidemiology and significance for HIV management. *Curr Opin Infect Dis* 2006; 19:1–7.
- UNAIDS: Global Report: UNAIDS Report on the Global AIDS Epidemic 2013. UNAIDS, Geneva, 2013.
- AIDSinfo: UNAIDS. <http://aidsinfo.unaids.org/> (2015). Accessed October 14, 2015.
- Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB: A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317:944–947.
- van Manen D, Kootstra NA, Boeser-Nunnink B, Handulle MA, van't Wout AB, Schuitemaker H: Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS* 2009;23:19–28.
- Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M: 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.
- Dalmaso C, Carpentier W, Meyer L, Rouzioux C, Goujard C, Chaix ML, Lambotte O, Avettand-Fenoel V, Le Clerc S, de Senneville LD, Deveau C, Boufassa F, Debre P, Del-fraissy JF, Broet P, Theodorou I: ANRS Genome Wide Association 01: Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: The ANRS Genome Wide Association 01 study. *PLoS One* 2008;3:e3907.
- O'Brien SJ, Hendrickson SL: Host genomic influences on HIV/AIDS. *Genome Biol* 2013;14:201.
- An P, Winkler CA: Host genes associated with HIV/AIDS: Advances in gene discovery. *Trends Genet* 2010;26:119–131.
- Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP, Castagna A, Cozzi-Lepri A, De Luca A, Easterbrook P, Gunthard HF, Mallal S, Mussini C, Dalmau J, Martinez-Picado J, Miro JM, Obel N, Wolinsky SM, Martinson JJ, Detels R, Margolick JB, Jacobson LP, Descombes P, Antonarakis SE, Beckmann JS, O'Brien SJ, Letvin NL, McMichael AJ, Haynes BF, Carrington M, Feng S, Telenti A, Goldstein DB: NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI): Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* 2009;5:e1000791.
- van Manen D, Delaneau O, Kootstra NA, Boeser-Nunnink BD, Limou S, Bol SM, Burger JA, Zwinderman AH, Moerland PD, van 't Slot R, Zagury JF, van 't Wout AB, Schuitemaker H: Genome-wide association scan in HIV-1-infected individuals identifying variants influencing disease course. *PLoS One* 2011;6:e22208.
- McLaren PJ, Coulounges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, Cossarizza A, Dalmau J, Deeks SG, Delaneau O, De Luca A, Goedert JJ, Haas D, Herbeck JT, Kathiresan S, Kirk GD, Lambotte O, Luo M, Mallal S, van Manen D, Martinez-Picado J, Meyer L, Miro JM, Mullins JI, Obel N, O'Brien SJ, Pereyra F, Plummer FA, Poli G, Qi Y, Rucart P, Sandhu MS, Shea PR, Schuitemaker H, Theodorou I, Vannberg F, Veldink J, Walker BD, Weintrob A, Winkler CA, Wolinsky S, Telenti A, Goldstein DB, de Bakker PI, Zagury JF, Fellay J: Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS Pathog* 2013;9:e1003515.
- Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, Haas DW, Martinez-Picado J, Dalmau J, Lopez-Galindez C, Casado C, Rauch A, Gunthard HF, Bernasconi E, Vernazza P, Klimkait T, Yerly S, O'Brien SJ, Listgarten J, Pfeifer N, Lippert C, Fusi N, Kutalik Z, Allen TM, Muller V, Harrigan PR, Heckerman D, Telenti A, Fellay J: A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* 2013;2:e01123.
- Troyer JL, Nelson GW, Lautenberger JA, Chinn L, McIntosh C, Johnson RC, Sezgin E, Kessing B, Malasky M, Hendrickson SL, Li G, Pontius J, Tang M, An P, Winkler CA, Limou S, Le Clerc S, Delaneau O, Zagury JF, Schuitemaker H, van Manen D, Bream JH, Gomperts ED, Buchbinder S, Goedert JJ, Kirk GD, O'Brien SJ: Genome-wide association study implicates PARD3B-based AIDS restriction. *J Infect Dis* 2011;203:1491–1502.
- Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, Alkan C, Kidd JM, Sun Y, Drautz DI, Bouffard P, Muzny DM, Reid JG, Nazareth LV, Wang Q, Burhans R, Riemer C, Wittekindt NE, Moorjani P, Tindall EA, Danko CG, Teo WS, Buboltz AM, Zhang Z, Ma Q, Oosthuysen A, Steenkamp AW, Oosthuisen H, Venter P, Gajewski J, Zhang Y, Pugh BF, Makova KD, Nekrutenko A, Mardis ER, Patterson N, Pringle TH, Chiaromonte F, Mullikin JC, Eichler EE, Hardison RC, Gibbs RA, Harkins TT, Hayes VM: Complete Khoisan and Bantu genomes from southern Africa. *Nature* 2010;463:943–947.
- Yao F: Functional principle component analysis for longitudinal and survival data. *Stat Sin* 2007;17:965–983.
- James GM, Hastie TJ, Sugar CA: Principal component models for sparse functional data. *Biometrika* 2000;87:587–602.
- Yao F, Müller H-G, Wang J-L: Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 2005;100:577–590.
- Wu S, Müller H-G: Response-adaptive regression for longitudinal data. *Biometrics* 2011;67:852–860.
- Baum MK, Campa A, Lai S, Sales Martinez S, Tsalaile L, Burns P, Farahani M, Li Y, van Widenfelt E, Page JB, Bussmann H, Fawzi WW, Moyo S, Makhema J, Thior I, Essex M, Marlink R: Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naïve, HIV-infected adults in Botswana: A randomized clinical trial. *JAMA* 2013;310:2154–2163.
- 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.
- Svitin A, Malov S, Cherkasov N, Geerts P, Rotkevich M, Dobrynin P, Shevchenko A, Guan L, Troyer J, Hendrickson S, Dilks H, Oleksyk T, Donfield S, Gomperts E, Jabs D, Sezgin E, Van Natta M, Harrigan P, Brumme Z, O'Brien S: GWATCH: A web platform for automated gene association discovery analysis. *Gigascience* 2014;3:18.
- Shapiro RL, Thior I, Gilbert PB, Lockman S, Wester C, Smeaton LM, Stevens L, Heymann SJ, Ndung'u T, Gaseitsiwe S, Novitsky V, Makhema J, Lagakos S, Essex

- M: Maternal single-dose nevirapine versus placebo as part of an antiretroviral strategy to prevent mother-to-child HIV transmission in Botswana. *AIDS* 2006;20:1281–1288.
24. Thior I, Lockman S, Smeaton LM, Shapiro RL, Wester C, Heymann SJ, Gilbert PB, Stevens L, Peter T, Kim S, van Widenfelt E, Moffat C, Ndase P, Arimi P, Kebaabetswe P, Mazonde P, Makhema J, McIntosh K, Novitsky V, Lee TH, Marlink R, Lagakos S, Essex M; Mashi Study Team: Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child HIV transmission in Botswana: A randomized trial: The Mashi Study. *JAMA* 2006;296:794–805.
 25. Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhan R, Nkwe D, Margolin L, Musonda R, Moyo S, Woldegabriel E, van Widenfelt E, Makhema J, Essex M: Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. *Virology* 2009;383:47–59.
 26. Novitsky V, Wang R, Kebaabetswe L, Greenwald J, Rossenkhan R, Moyo S, Musonda R, Woldegabriel E, Lagakos S, Essex M: Better control of early viral replication is associated with slower rate of elicited antiviral antibodies in the detuned enzyme immunoassay during primary HIV-1C infection. *J Acquir Immune Defic Syndr* 2009;52:265–272.
 27. Novitsky V, Wang R, Margolin L, Baca J, Kebaabetswe L, Rossenkhan R, Bonney C, Herzig M, Nkwe D, Moyo S, Musonda R, Woldegabriel E, van Widenfelt E, Makhema J, Lagakos S, Essex M: Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. *PLoS One* 2009;4:e7727.
 28. Novitsky V, Woldegabriel E, Kebaabetswe L, Rossenkhan R, Mlotshwa B, Bonney C, Finucane M, Musonda R, Moyo S, Wester C, van Widenfelt E, Makhema J, Lagakos S, Essex M: Viral load and CD4+ T-cell dynamics in primary HIV-1 subtype C infection. *J Acquir Immune Defic Syndr* 2009;50:65–76.
 29. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhan R, Ketunuti M, Makhema J, Seage GR, 3rd, Essex M: Identification of primary HIV-1C infection in Botswana. *AIDS Care* 2008;20:806–811.
 30. James GM: Generalized linear models with functional predictors. *J R Stat Soc Series B Stat Methodol* 2002;64:411–432.
 31. Rice JA, Silverman BW: Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Series B Stat Methodol* 1991;53:233–243.
 32. Hall P, Müller H-G, Wang J-L: Properties of principal component methods for functional and longitudinal data analysis. *Ann Stat* 2006;34:1493–1517.
 33. Robinson GK: That BLUP is a good thing: The estimation of random effects. *Stat Sci* 1991;6:15–32.
 34. Holte S, Randolph T, Ding J, Tien J, McClelland R, Baeten J, Overbaugh J: Efficient use of longitudinal CD4 counts and viral load measures in survival analysis. *Stat Med* 2012;31:2086–2097.
 35. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES: Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 2010;42:355–360.
 36. Laird NM, Ware JH: Random-effects models for longitudinal data. *Biometrics* 1982;38:963–974.
 37. Smith EN, Chen W, Kahonen M, Kettunen J, Lehtimäki T, Peltonen L, Raitakari OT, Salem RM, Schork NJ, Shaw M, Srinivasan SR, Topol EJ, Viikari JS, Berenson GS, Murray SS: Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet* 2010;6:e1001094.
 38. Lin X: Variance component testing in generalised linear models with random effects. *Biometrika* 1997;84:309–326.
 39. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011;89:82–93.
 40. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57:289–300.
 41. Nalls MA, Couper DJ, Tanaka T, van Rooij FJ, Chen MH, Smith AV, Toniolo D, Zakai NA, Yang Q, Greinacher A, Wood AR, Garcia M, Gasparini P, Liu Y, Lumley T, Folsom AR, Reiner AP, Gieger C, Lagou V, Felix JF, Volzke H, Gouskova NA, Biffi A, Doring A, Volker U, Chong S, Wiggins KL, Rendon A, Dehghan A, Moore M, Taylor K, Wilson JG, Lettre G, Hofman A, Bis JC, Pirastu N, Fox CS, Meisinger C, Sambrook J, Arepalli S, Nauck M, Prokisch H, Stephens J, Glazer NL, Cupples LA, Okada Y, Takahashi A, Kamatani Y, Matsuda K, Tsunoda T, Tanaka T, Kubo M, Nakamura Y, Yamamoto K, Kamatani N, Stumvoll M, Tonjes A, Prokopenko I, Illig T, Patel KV, Garner SF, Kuhnle B, Mangino M, Oostra BA, Thein SL, Coresh J, Wichmann HE, Menzel S, Lin J, Pistis G, Uitterlinden AG, Spector TD, Teumer A, Eiriksdottir G, Gudnason V, Bandinelli S, Frayling TM, Chakravarti A, van Duijn CM, Melzer D, Ouwehand WH, Levy D, Boerwinkle E, Singleton AB, Hernandez DG, Longo DL, Soranzo N, Witteman JC, Psaty BM, Ferrucci L, Harris TB, O'Donnell CJ, Ganesh SK: Multiple loci are associated with white blood cell phenotypes. *PLoS Genet* 2011;7:e1002113.
 42. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, Iyadurai K, Becker N, Brooks-Wilson A, Curry JD, Spinelli JJ, Holly EA, Riby J, Zhang L, Nieters A, Smith MT, Brown KM: Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat Genet* 2009;41:873–875.
 43. Chubb D, Weinhold N, Broderick P, Chen B, Johnson DC, Forsti A, Vijayakrishnan J, Migliorini G, Dobbins SE, Holroyd A, Hose D, Walker BA, Davies FE, Gregory WA, Jackson GH, Irving JA, Pratt G, Fegan C, Fenton JA, Neben K, Hoffmann P, Nothen MM, Muhleisen TW, Eisele L, Ross FM, Straka C, Einsele H, Langer C, Dorner E, Allan JM, Jauch A, Morgan GJ, Hemminki K, Houlston RS, Goldschmidt H: Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet* 2013;45:1221–1225.
 44. Yeung ML, Houzet L, Yedavalli VS, Jeang KT: A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J Biol Chem* 2009;284:19463–19473.
 45. Le Clerc S, Limou S, Coulonges C, Carpentier W, Dina C, Taing L, Delaneau O, Labib T, Sladek R, ANRS Genomic Group, Deveau C, Guillemain H, Ratsimandresy R, Montes M, Spadoni JL, Therwath A, Schachter F, Matsuda F, Gut I, Lelievre JD, Levy Y, Froguel P, Delfraissy JF, Hercberg S, Zagury JF: Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for

- AIDS (ANRS Genomewide Association Study 03). *J Infect Dis* 2009;200:1194–1201.
46. Zhang HS, Ruan Z, Sang WW: HDAC1/NFkappaB pathway is involved in curcumin inhibiting of Tat-mediated long terminal repeat transactivation. *J Cell Physiol* 2011;226:3385–3391.
47. Zhang HS, Wu MR: SIRT1 regulates Tat-induced HIV-1 transactivation through activating AMP-activated protein kinase. *Virus Res* 2009;146:51–57.
48. Shirakawa K, Takaori-Kondo A, Yokoyama M, Izumi T, Matsui M, Io K, Sato T, Sato H, Uchiyama T: Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat Struct Mol Biol* 2008;15:1184–1191.
49. Francis AC, Di Primio C, Allouch A, Cereseto A: Role of phosphorylation in the nuclear biology of HIV-1. *Curr Med Chem* 2011;18:2904–2912.
50. Barnitz RA, Wan F, Tripuraneni V, Bolton DL, Lenardo MJ: Protein kinase A phosphorylation activates Vpr-induced cell cycle arrest during human immunodeficiency virus type 1 infection. *J Virol* 2010;84:6410–6424.
51. Nisole S, Stoye JP, Saib A: TRIM family proteins: Retroviral restriction and antiviral defence. *Nat Rev Microbiol* 2005;3:799–808.
52. Jager S, Cimermanic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, Hernandez H, Jang GM, Roth SL, Akiva E, Marlett J, Stephens M, D'Orso I, Fernandes J, Fahey M, Mahon C, O'Donoghue AJ, Todorovic A, Morris JH, Maltby DA, Alber T, Cagney G, Bushman FD, Young JA, Chanda SK, Sundquist WI, Kortemme T, Hernandez RD, Craik CS, Burlingame A, Sali A, Frankel AD, Krogan NJ: Global landscape of HIV-human protein complexes. *Nature* 2012;481:365–370.
53. Bergonzini V, Calistri A, Salata C, Del Vecchio C, Sartori E, Parolin C, Palu G: Nef and cell signaling transduction: A possible involvement in the pathogenesis of human immunodeficiency virus-associated dementia. *J Neurovirol* 2009;15:238–248.
54. Abdel-Mohsen M, Raposo RA, Deng X, Li M, Liegler T, Sinclair E, Salama MS, Ghanem Hel D, Hoh R, Wong JK, David M, Nixon DF, Deeks SG, Pillai SK: Expression profile of host restriction factors in HIV-1 elite controllers. *Retrovirology* 2013;10:106.
55. Raposo RA, Abdel-Mohsen M, Bilska M, Montefiori DC, Nixon DF, Pillai SK: Effects of cellular activation on anti-HIV-1 restriction factor expression profile in primary cells. *J Virol* 2013;87:11924–11929.
56. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ: Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008;319:921–926.
57. Wonderlich ER, Leonard JA, Kulpa DA, Leopold KE, Norman JM, Collins KL: ADP ribosylation factor 1 activity is required to recruit AP-1 to the major histocompatibility complex class I (MHC-I) cytoplasmic tail and disrupt MHC-I trafficking in HIV-1-infected primary T cells. *J Virol* 2011;85:12216–12226.
58. Blagoveshchenskaya AD, Thomas L, Feliciangeli SF, Hung CH, Thomas G: HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway. *Cell* 2002;111:853–866.
59. Ozato K, Shin DM, Chang TH, Morse HC, 3rd: TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol* 2008;8:849–860.
60. van Manen D, van 't Wout AB, Schuitemaker H: Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology* 2012;9:70.
61. Botswana Ministry of Health: 2012 Botswana National HIV & AIDS Treatment Guidelines. Ministry of Health, Gaborone, Botswana, 2012.
62. World Health Organization: Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach. World Health Organization, Geneva, June 2013. ISBN: 978 92 4 150572 7.

Address correspondence to:

M. Essex

Harvard T.H. Chan School of Public Health AIDS Initiative

Department of Immunology and Infectious Diseases

Harvard T.H. Chan School of Public Health

651 Huntington Avenue FXB 402

Boston, MA 02115

E-mail: messex@hsph.harvard.edu