Immunogenetic influences on acquisition of HIV-1 infection: consensus findings from two African cohorts point to an enhancer element in IL19 (1q32.2)

X Li¹, K Zhang², NM Pajewski³, I Brill⁴, HA Prentice⁴, S Shrestha⁴, W Kilembe⁵, E Karita⁶, S Allen⁶,⁷, E Hunter⁸, RA Kaslow⁴,¹⁰ and J Tang²

Numerous reports have suggested that immunogenetic factors may influence human immunodeficiency virus (HIV-1) acquisition, yet replicated findings that translate between study cohorts remain elusive. Our work aimed to test several hypotheses about genetic variants within the IL10-IL24 gene cluster that encodes interleukin (IL)-10, IL-19, IL-20 and IL-24. In aggregated data from 515 Rwandans and 762 Zambians with up to 12 years of follow-up, 190 single-nucleotide polymorphisms passed quality control procedures. When HIV-1-exposed seronegative subjects (n = 486) were compared with newly seroconverted individuals (n = 313) and seroreverting subjects (n = 478) who were already infected at enrollment, rs12407485 (G > A) in IL19 showed a robust association signal in adjusted logistic regression models (odds ratio = 0.64, P = 1.7 × 10⁻⁴ and q = 0.033). Sensitivity analyses demonstrated that (i) results from both cohorts and subgroups within each cohort were highly consistent; (ii) verification of HIV-1 infection status after enrollment was critical; and (iii) supporting evidence was readily obtained from Cox proportional hazards models. Data from public databases indicate that rs12407485 is part of an enhancer element for three transcription factors. Overall, these findings suggest that molecular features at the IL19 locus may modestly alter the establishment of HIV-1 infection.

INTRODUCTION

Sexual transmission of human immunodeficiency virus (HIV)-1 infection is an inefficient process¹ that is limited in part by a clear selection bias for the transmitted/founder viruses.²,³ Studies in the past two decades have revealed various host genetic factors as probable determinants of HIV-1 acquisition (seroconversion) after high-risk sexual exposure.⁴-⁶ However, with rare exceptions, these findings have proved to be difficult to confirm or replicate. Indeed, recent genome-wide association studies (GWAS) have repeatedly failed to generate consistent findings when HIV-1-exposed seronegative (HESN) subjects are compared with HIV-1-infected individuals in a typical case–control study design.⁷-¹¹ Notably, the only confirmed variant to confer resistance to HIV-1 seroconversion, a 32-bp deletion in the CCR5 open-reading frame (CCR5-Δ32), is not applicable to populations of African and Asian ancestries.⁴

Several investigations with partially consistent results have highlighted single-nucleotide polymorphisms (SNPs) within the interleukin (IL)-10 gene (IL10) on chromosome 1q32.1. In particular, −592A>C (rs1800872T>G), −819T>C (rs1800871A>G) and −1082A>G (rs1800896T>C) in the IL10 promoter have been implicated individually or collectively (as haplotypes) in several cohorts.¹²-¹⁹ As IL-10 is a potent anti-inflammatory cytokine that regulates viral clearance,²⁰ a biological role in HIV-1 acquisition is quite plausible. Fine mapping for IL10 and neighboring loci, including IL19, IL20 and IL24, has been attempted in an African American cohort,¹⁹ but causal variants have not been identified. Our work here aimed to close this gap through dense coverage of the IL10-IL24 gene cluster in two distinct African cohorts, under the assumption that fine mapping based on African populations with relatively short haplotype block²¹-²³ can resolve ambiguities and offer new insights.

RESULTS

Overall characteristics of two prospective cohorts available to this study

Within two cohorts of HIV-1 serodiscordant couples from Lusaka, Zambia (1995–2012) and Kigali, Rwanda (2005–2012),²⁴-²⁸ our work focused on the analyses of three subgroups, that is, HESN subjects who had sufficient follow-up (≥9 months) for repeated verification of HIV-1 transmission status, HIV-1 seroconverters (SCs) identified during quarterly follow-up visits, as well as HIV-1 seroreverting subjects (SPs) who were already infected at the time of enrollment. After several layers of quality control, a total of 515 Rwandans and 762 Zambians were available for this study (Table 1). Within each cohort, the three subgroups shared similar demographic features (age and sex), but HESNs had substantially longer follow-up than SCs (P < 0.001 in both cohorts; Table 1), suggesting that their categorization as HESNs was not an artifact of infrequent testing. On the other hand,
genital ulcer/inflammation (GUI) as a well-known risk for HIV-1 acquisition\textsuperscript{28–30} was more frequent in SCs than HESNs (\(P < 0.001\) in both cohorts). In subsequent analyses of host genetic variants, statistical models were adjusted for potential confounders whenever possible.

**Immunogenetic findings based on both cohorts (screening models)**

For the genomic region spanning \(\text{IL10, IL19, IL21 and IL24}\) (nucleotide positions 205\,007\,570 to 205\,144\,106, plus a 3-kb extra region on each side), genotyping using a fine-mapping assay (the ImmunoChip; Illumina, San Diego, CA, USA) yielded 189 candidate SNPs in Rwandans and 191 candidate SNPs in Zambians (Supplementary Tables S1 and S2 in Supplementary Materials), which more than tripled the SNP coverage provided by a gene-centric, custom bead array that helped with generating hypotheses.\textsuperscript{19} In joint logistic regression models adjusted for sex, age, cohort and genetic ancestry, QQ plot (Supplementary Figure S1), Manhattan regional plot (Figure 1) and other summary statistics (Supplementary Table S3) for the combined data set (515 Rwandans and 762 Zambians combined), \(P\)-values were corrected for 190 candidate SNPs and 1277 subjects revealed multiple SNPs of interest. When the \(P\)-values were corrected for 190 candidate SNPs (assuming dominant effects for minor alleles) or 95 independent tests (eliminating 95 tagging SNPs), only rs12407485, an intronic SNP within \(\text{IL19}\), retained statistical significance (\(P = 0.031\) or 0.016): the minor allele of rs12407485 was less prevalent in HIV-1-infected subjects (SCs+SPs) than in HESNs (odds ratio (OR) = 0.64, \(P = 1.7 \times 10^{-4}\), and false discovery rate (FDR) = 0.033; Supplementary Tables S3 and S4).

**Local haplotype structures**

The candidate SNPs used for association analyses did not form any extended haplotypes in either cohort (data not shown). In a 6-kb subregion around rs12407485, the haplotype structures were quite similar between Rwandans and Zambians, regardless of HIV-1 infection status (Figure 2). A single haplotype block containing four SNPs was 871-bp upstream from rs12407485, while the downstream SNPs did not form any blocks. In both cohorts, SNPs that could partially \((r^2 > 0.50)\) tag rs12407485 included rs12042745, rs10863861 and rs1878673 \((r^2 = 0.53–0.67); Figure 2\), which largely accounted for the detection of multiple secondary association signals (Supplementary Table S3). Among frequent haplotypes encompassing rs12407485 and 10 other SNPs (five in each direction), only one had the rs12407485-A allele (Supplementary Table S5).

**Confirmatory findings from several sensitivity analyses and multivariable models**

The negative association between rs12407485-A and HIV-1 acquisition was highly consistent when analyses were done for separate cohorts and subgroups within each cohort, with ORs ranging from 0.59 to 0.71 \((P = 0.007–0.243)\) when age, sex and genetic ancestry were treated as covariates (Table 2). The magnitude of effect size attributable to rs12407485-A was persistently modest in both cohorts, with a 39% protection against seroconversion for Rwandans and 36% for Zambians. Statistical power for detecting such minor effects in the study cohorts was adequate (Supplementary Table S6).

Two additional analyses revealed that the negative association of rs12407485-A with HIV-1 acquisition was not apparent if analyses relied on subjects identified as HIV-1 seronegative (HIV\(-\)) or seropositive (HIV\(+\)) at the time of enrollment (no follow-up)
or after 2 years of regular testing for HIV-1 infection (Table 3). For example, when 478 HIV+ individuals were compared with 799 HIV− subjects at baseline, rs12407485-A was slightly less frequent in the HIV+ group (adjusted OR = 0.71), with a corrected P-value of 0.475 (0.005 × 95). The relevance of demographic features (age and sex) and geography to HIV-1 acquisition was not apparent either (∼0.015–0.742) when the interim analyses were conducted at baseline or 2 years into follow-up.

In analyses restricted to HESNs and SCs with confirmed transmission source partners, an expanded multivariable model further indicated that the negative association of rs12407485-A with HIV-1 acquisition was independent of other factors (adjusted OR = 0.52, 95% confidence interval = 0.37–0.75; Table 4). Indeed the relative effect size (OR) for rs12407485-A improved when viral and host factors applicable to both cohorts were evaluated simultaneously (Table 4).

Figure 2. Patterns of LD for SNPs around rs12407485 (a non-coding SNP in IL19). Candidate SNPs within a 3-kb region beyond rs12407485 (indicated by asterisks) are analyzed for HIV-1-infected Rwandans (top panel) and Zambians (bottom panel). The pairwise r² values are boxed. Almost identical results are seen in analyses of HESN subjects.
Table 2. Association of rs12407485 minor allele A with resistance to HIV-1 acquisition, as revealed by adjusted logistic regression models in two separate cohorts

| Subgroups in modela | rs12407485-A in Rwandans | rs12407485-A in Zambians |
|---------------------|--------------------------|--------------------------|
| n                   | OR 95% CI                 | OR 95% CI                 |
| HESNs versus (SCs+SPs) | 515 0.61 0.43–0.88 0.008 | 762 0.64 0.47–0.88 0.007 |
| HESNs versus SCs     | 289 0.71 0.39–1.27 0.243 | 510 0.65 0.45–0.95 0.025 |
| HESNs versus SPs     | 454 0.59 0.40–0.86 0.007 | 510 0.63 0.44–0.92 0.016 |

Abbreviations: CI, confidence interval; HESN, HIV-1-exposed seronegative; MDS, multidimensional scaling analysis; OR, odds ratio; SC, seroconverter; SP, seroprevalent. aThe three subgroups are defined in Table 1 and text; OR and 95% CI have been adjusted for age, sex and regional genetic ancestry (MDS1–MDS4, see text). OR > 1.0 is unfavorable (being seropositive) for HIV-1 infection.

Table 3. Sensitivity analyses: alternative logistic regression models for 1277 subjects (515 Rwandans and 762 Zambians)

| Factors in model | Baseline model: 478 seropositives versus 799 seronegativesb | Interim model: 678 seropositives versus 599 seronegativesc |
|------------------|-------------------------------------------------------------|-------------------------------------------------------------|
|                   | OR 95% CI Adjusted P                                        | OR 95% CI Adjusted P                                        |
| Age, >40 years    | 1.18 0.84–1.68 0.343                                        | 0.85 0.61–1.20 0.363                                        |
| Sex, female       | 1.07 0.85–1.37 0.557                                        | 1.20 0.95–1.51 0.129                                        |
| Country, Rwanda   | 3.50 1.27–9.62 0.015                                        | 1.18 0.44–3.13 0.742                                        |
| rs12407485-A      | 0.71 0.55–0.90 0.005d                                        | 0.69 0.55–0.87 0.002e                                        |

Abbreviations: CI, confidence interval; HIV, human immunodeficiency virus; MDS, multidimensional scaling analysis; OR, odds ratio; SP, seroprevalent. aAs defined in Table 1 and text; OR and CI are further adjusted for genetic ancestry (MDS1–MDS4, see text). OR > 1.0 is unfavorable (being seropositive). bUp to 2 years of follow-up after enrollment, with 200 confirmed seroconversion events (as in Figure 3). cP > 0.19 when corrected for 95 independent tests.

Table 4. Multivariable logistic regression model: inclusion of other covariates that are applicable to 708 subjects (HESNs and SCs) from both cohorts

| Factors in model | n | OR 95% CI Adjusted P |
|------------------|---|----------------------|
| Age, >40 years   | 91 | 0.53 0.30–0.95 0.031 |
| Sex, female      | 370 | 0.94 0.65–1.34 0.721 |
| Region, Rwanda   | 222 | 0.34 0.14–0.86 0.023 |
| Donor VL, >100 000 copies per mlb | 257 | 2.08 1.42–3.03 <0.0001 |
| Donor VL, <10 000 copies per mlb | 166 | 0.37 0.22–0.61 <0.0001 |
| HLA-B, P2-Metc   | 314 | 1.28 0.91–1.81 0.152 |
| rs12407485-A     | 263 | 0.52 0.37–0.75 <0.001 |

Abbreviations: CI, confidence interval; GUI, genital ulcer/inflammation; HESN, HIV-1-exposed seronegative; HIV, human immunodeficiency virus; HLA, human leukocyte antigen; OR, odds ratio; SC, seroconverter; VL, viral load. aAs defined in Table 1 and excluding 24 subjects with missing information for GUI or HIV-1 VL in cohabiting index (donor) partners. OR and CI are further adjusted for genetic ancestry (MDS1–MDS4, see text). OR > 1.0 is unfavorable (being HIV+). bPlasma VL in known index partners (the reference group has medium plasma VL, that is, between 10 000 and 100 000 copies per ml). cOther host genetic factor that is applicable to both cohorts, especially in terms of time to HIV-1 acquisition (Table 5).

Results from alternative analyses: rs12407485 genotypes and time to seroconversion

In univariable analyses restricted to 222 Rwandans and 486 Zambians who were HIV− at enrollment, individuals with rs12407485-A displayed a slower rate of seroconversion when compared with other subjects (Figure 3). A multivariable Cox proportional hazards model (Table 5) confirmed the independent association of rs12407485-A with delayed HIV-1 acquisition (adjusted hazard ratio (HR) = 0.65, P = 0.001). Previously identified factors associated with HIV-1 seroconversion, including GUI and HIV-1 viral load (VL) in known (cohabiting) index partners, were also independent factors in this select group of subjects (Table 5). Overall, GUI was the most prominent predictor of HIV-1 acquisition (HR = 2.75, P < 0.0001), while low VL (< 10 000 copies per ml) in index partners was the most favorable factor in the model (HR = 0.42, P < 0.0001).

Insights from bioinformatics analyses

Data from HaploReg revealed that rs12407485 is within an enhancer element with predicted binding affinity for three transcription factors (Supplementary Table S7 in Supplementary Materials). A single tagging SNP, rs17016339 (not covered by the ImmunoChip or by the Illumina 1 M Duo chip, Illumina), has been seen in Africans (r2 = 0.99). This intronic SNP is also relevant to gene transcription (Supplementary Table S7). A second SNP, rs4347211, can partially tag rs17016339 (r2 = 0.77). Although rs4347211 is covered by the Illumina 1M Duo chip, existing data sets at dbGaP (https://www.ncbi.nlm.nih.gov/gap, last accessed on 19 December 2014) did not include two GWAS based on African natives.

The NCBI Global Cross-database has no existing entries for IL19-related GWAS data or gene expression eQTLs (eQTLs), but the SNP and CNV Annotation (SCAN) database indicates that both rs12407485 and rs17016339 are probable trans-acting eQTLs.
(P = 8 × 10^{-5} in Yoruban samples) for \textit{GARS} (chromosome 7p15), a gene that encodes glycyll-tRNA synthetase (Supplementary Table S8).

The newly established Finemapping Data Portal (also based on the ImmunoChip)\textsuperscript{31} lists five IL10 SNPs (rs1518110, rs1518111, rs3024493, rs3024495 and rs3024490) that are predicted to be causal variants previously associated with four autoimmune disorders, but neither rs12407485 nor rs17016339 is in strong linkage disequilibrium with these \textit{IL10} SNPs of interest.

Other observations

Among other SNPs analyzed in this study, rs1800896 (−1082 G > A) and rs1800872 (−592C > A) in \textit{IL10} and rs2981572 in \textit{IL20} have been highlighted in various studies related to HIV-1 acquisition or disease progression.\textsuperscript{12,13,15-19} The two \textit{IL10} SNPs, rs1800896 and rs1800872 (tagged by rs1518111, rs1800871, and three other SNPs in Africans), had borderline statistical significance (P = 0.050 and 0.057, respectively) in screening models (Supplementary Table S3); these minor associations were readily dismissed by the FDR values (0.242 for both) and by inconsistencies with the directions of reported associations. The \textit{IL20} SNP, rs2981572, could not be confirmed either (P = 0.340).

Meanwhile, several cohort-specific observations (adjusted P < 0.01) were noted, but again the false discovery rates (q > 0.37) were too high to warrant further attention.

**DISCUSSION**

In contrast with earlier reports, new data from two African cohorts failed to confirm the potential roles of several \textit{IL10} and \textit{IL20} SNPs in HIV-1 acquisition. Instead, promising evidence from primary and secondary models persistently and consistently suggested that a non-coding SNP (rs12407485) in \textit{IL19} (1q32.2) might be an independent host factor negatively associated with incident and prevalent HIV-1 infection. Although these epidemiological findings still fell short of a genome-wide significance, the consensus results and their underlying biology are promising in light of existing data from several public databases.

By extending the usual comparison of SCs (incident cases) versus HESNs (contemporary controls)\textsuperscript{26,30} to further analyses of HESNs against SPs alone or SPs combined with SCs, we were able to demonstrate that the immunogenetic relationships seen with rs12407485 in \textit{IL19} were remarkably similar in all analyses (for example, Tables 2 and 4). These separate and joint analyses of seroincident and seroprevalent HIV-1 infection may benefit future studies, as identification of pathogenetic mechanisms relevant to HIV-1 prevention remain critical in many parts of Africa where treatment as prevention\textsuperscript{27} is still unaffordable, and where internal and external validation of immunogenetic relationships are quite rare.

Our findings here can attest to the daunting challenges presented by an evolving and relative trait such as resistance to HIV-1 acquisition. In populations with high rates of HIV-1 transmission, the HIV-1 infection status of seronegative subjects can change rapidly (in months), which should mandate the need for longitudinal testing before resistance to HIV-1 infection is ascertained.\textsuperscript{22} Unfortunately, most studies can only compare HIV-1-infected subjects with HIV− individuals in a typical case–control study design, often without reference to (i) the characteristics of source/donor partners, (ii) assessment of exposure based on follow-up data (especially biological evidence) and (iii) verification of HIV-1 infection status in the initially HIV− subjects. Confirmation of HIV-1 infection through repeated testing is particularly important as positive serology took months to develop after acute infection.\textsuperscript{24} Our study was able to overcome these apparent deficiencies and further established that even an extra 2 years of follow-up after enrollment could have led to false-negative findings (Table 3). Findings based on the Multicenter...
AIDS Cohort Study also suggested that ascertainment of high risk for HIV-1 infection after enrollment is critical to the identification of host genetic factors that delays or prevents HIV-1 acquisition. Lack of supporting evidence form earlier efforts\textsuperscript{7–10} might also reflect their inability to target enough IL19 SNPs or infer IL19 genotypes based on known genetic structure (Supplementary Table S7). The rs12407485 SNP in IL19 is a case in point: it was not covered by previous GWAS arrays; its only tagging SNP (rs17016339) was overlooked by GWAS as well (Supplementary Tables S3 and S7). Indeed, many causal SNPs identified by the ImmunoChip are beyond the scope of routine genotyping for GWAS because they are typically distant (~14-kb away) from coding sequences.\textsuperscript{3} By focusing on immune response genes, especially regions with unambiguous genetic associations from independent genome-wide scans and meta-analyses, the ImmunoChip is advantageous in terms of locus-specific SNP coverage and fine mapping.\textsuperscript{31}

Non-genetic factors (for example, co-infections) that mediate HIV-1 acquisition\textsuperscript{36} are expected to obscure modest, immunogenetic contributions, especially as risk for exposure can evolve (for example, the motivation for child-bearing diminishes as couples age). A prerequisite to any claim for a determinant of HIV-1 resistance is the proof of adequate exposure.\textsuperscript{3,37} Frequent diagnosis of GUI in our cohorts provided clear evidence that high-risk factors did exist in HESNs and their partners (Table 1). In index partners and confirmed transmission source partners, HIV-1 VL was a stronger parameter than is known to improve statistical modeling (Table 5), although not all subjects (for example, SPs) could be analyzed this way. These practical issues can be addressed in cohorts with frequent testing and adequate exposure.

As far as rs12407485 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=12407485) in IL19 is concerned, its novel association with resistance to HIV-1 infection may reflect the importance of various immunoregulatory cytokines, including IL-10 that controls the clearance of viral, bacterial and parasitic infections.\textsuperscript{25,38} Many pathogens produce IL-10-like molecules to subvert host immunity and maintain latency.\textsuperscript{39} IL-19 itself is closely related to IL-10 in terms of amino-acid sequence homology and function, and both are expressed by a wide range of cell lineages to suppress inflammation and cell adhesion.\textsuperscript{40,41} or to facilitate tissue repair,\textsuperscript{42,43} usually in the contraction phase of immune responses. A connection between rs12407485 and cytokine-mediated, immunoregulatory functions is probable because rs12407485 variants are predicted to alter an enhancer-like sequence motif.

The rs12407485 locus is already known to be a trans-acting eQTL for the glycol-tRNA synthetase gene (GARS). In African populations, the lone tagging SNP (rs17016339) for rs12407485 is another probable trans-acting eQTL for GARS, with multiple functional properties in cell lines. These eQTL relationships can be more complex than expected, as the tRNA synthetase gene family is notorious for splicing products that commercial gene expression arrays cannot distinguish.\textsuperscript{44} Follow-up studies may need to take advantage of high-resolution techniques, especially RNA-Seq, to fully elucidate the potential function of rs12407485 or its tagging SNP (r17016399) as a trans-acting eQTL for GARS (http://www.ncbi.nlm.nih.gov/gene/2617).

In other epidemiological studies, IL19 is one of the 139 genes in the human genome that have been associated with virus diversity in human populations,\textsuperscript{46} and several IL19 gene variants have been associated with ulcerative colitis, a disease that is characterized by excessive inflammation of intestinal mucosa.\textsuperscript{46} The two SNPs identified in a Japanese cohort, rs2243188 (intronic) and rs2243193 (3‘ untranslated region), have minor alleles that appear to be protective, but both actually tag many functionally relevant loci in Asians, some of which contribute to enhancer-related sequence motifs as well. A second study\textsuperscript{47} based on analyses of Europeans has identified an IL19 haplotype as being protective against palmoplantar pustulosis, a skin disease caused by chronic inflammation. If the importance of IL-19 to skin and mucosal health is generalizable, manipulation of IL-19 expression could be an attractive option for intervention.

**PATIENTS AND METHODS**

**Study populations**

Subjects available to this study were members of HIV-1 discordant couples assembled by the Rwanda–Zambia HIV-1 Research Group between 1995 and 2012 for voluntary testing and counseling. The overall study design and quarterly follow-up strategies have been described in detail elsewhere.\textsuperscript{24,26,30,46} For this study, analyses focused on comparing HESNs versus SCs and SPs with adequate follow-up and DNA samples for genotyping (Table 1). To minimize the potential bias for phylogenetically related viruses\textsuperscript{24,25} among the HIV-1-infected subjects (SCs and SPs), SPs known as transmission source partners (donor partners for SCs) were excluded, although their characteristics, especially GUI and VL\textsuperscript{20,48} were treated as covariates in multivariable models whenever possible. A small number of SCs with unknown source partners (viruses not confirmed in suspected source partners) were not analyzed for timing of HIV-1 infection after enrollment, as several factors in donor partners have strong influences and thus must be treated as covariates.\textsuperscript{20,48}

**Ethical aspects**

The research outlined in this study was approved annually by the institutional review boards at clinical sites in Africa (Lusaka, Zambia and Kigali, Rwanda) and two collaborating institutions in the United States (Emory University and University of Alabama at Birmingham). All subjects (legally married adults) gave written informed consent for participation in voluntary testing and counseling and related research activities. Interventions for combination antiretroviral therapy, medical male circumcision, use of condoms and other applicable measures (for example, diagnosis and treatment of sexually transmitted infections) followed national or regional guidelines whenever possible. All visits beyond combination antiretroviral therapy initiation, either self-reported or confirmed by presence of antiretrovirals in blood samples, were excluded from analyses.

**Factors already known to mediate HIV-1 transmission**

Various interim analyses demonstrated that heterosexual HIV-1 transmission in both cohorts was associated with several donor (index) and recipient (initially seronegative) characteristics,\textsuperscript{24,27,29,30,48} including (i) GUI in both partners, (ii) plasma VL in index partners (SPs) and (iii) male circumcision. Several host genetic factors from classic human leukocyte antigen genes and the KIR2DS4 locus have been reported for the Zambian cohort,\textsuperscript{27,28,30,49} but so far only the dimorphism for position 2 (P2) of the 9mer human leukocyte antigen-B leader peptide has been shown to impact HIV-1 acquisition in both cohorts when time to HIV-1 acquisition was assessed in Cox proportional hazard models.\textsuperscript{28} Potential confounding by these known factors was considered in multivariable modeling, whenever applicable.

**Genotyping using the ImmunoChip**

Genomic DNA samples from eligible subjects were used for SNP genotyping at a genomics core facility (University of Alabama at Birmingham). SNP alleles were assigned using the joint calling and haplotype phasing algorithm implemented in BEAGLE\textsuperscript{50}. For each cohort, we completed a series of data cleaning and quality control procedures for ~195 000 SNPs on the ImmunoChip, excluding loci based on the following criteria: (i) known duplication, (ii) missingness (call rate < 98.5%), (iii) minor allele frequency < 0.02 in either cohort or (iv) deviation from Hardy–Weinberg equilibrium (P < 10\textsuperscript{-6}). In the end, 190 candidate SNPs spanning the IL10-IL24 gene cluster (~3-kb region at each end of the cluster) were retained for analyses in both cohorts, while 78 monomorphic loci and ~40 rare SNPs were excluded (Supplementary Tables S1 and S2 in Supplementary Materials).
Additional data trimming based on ImmunoChip results

Subjects with ImmunoChip data were further evaluated against three additional (and occasionally overlapping) quality control criteria: (i) overall call rates < 95%, (ii) failing sex determination (n = 0) or (iii) inferred familial relatedness based on kinship coefficients estimated using the KING software package.51,52 Subjects were selectively excluded in a pairwise fashion for up to third degree kinship.

Population stratification

To control for population stratification within the study cohorts, we performed a multidimensional scaling analysis (MDS) using a subset of the ImmunoChip 490 SNPs and the KING software package.21 When SNPs mapping to known regions of extended linkage disequilibrium within European populations12,23,25 were excluded, 34,390 SNPs with pairwise r2 < 0.20 were retained for MDS. The top four MDS vectors (MDS1-4) were then used as covariates in subsequent association analyses.

Assessment of linkage disequilibrium and haplotype structures in study populations

For each cohort stratified by HIV-1 infection status (positive or negative), local haplotypes in the IL10-IL24 gene cluster were inferred using HaploView.22 The ability of individual SNPs to tag adjacent loci was determined by r2 values. To facilitate an alternative strategy for Bonferroni correction of P-values and for assessment of statistical power in various analytical approaches (Supplementary Table S6), the number of independent SNPs within the candidate loci was calculated using SimpleM.5,6

Screening and statistical modeling of genetic associations with HIV-1 acquisition

To maximize statistical power (Supplementary Table S6) and to minimize the number of random testing, association analyses began with the analyses of aggregated SNP data from both cohorts, in which HESNs (controls) were compared with SCs+SPs (cases). The PLINK program (v1.07)26 was used to test all candidate SNPs that passed the quality control procedures, with an emphasis on consensus results. The accompanying QQ plot and Manhattan regional plot were created using the R package (qqman at http://cran.r-project.org/web/packages/qqman/index.html) and LocusZoom (http://csg.sph.umich.edu/locuszoom/),27 respectively. Recombination rates in the Manhattan regional plot were based on SNP data from Africans (Yorubans) in the International HapMap Project.23 In these screening models, the P-values and OR estimates (including 95% confidence interval) for individual SNPs were adjusted for four potential confounders (sex, age, country of origin and regional genetic ancestry) that were applicable to all subjects (Supplementary Table S3). Preliminary association signals (P < 0.05 after Bonferroni correction) were subjected to three sets of sensitivity analyses, that is, (i) HESNs versus SCs alone in each cohort, (ii) HESNs versus SPs alone in each cohort and (iii) multivariable models that further accounted for previously established covariates, including three categories of index partners’ plasma VLs (low, < 10,000 copies per ml; medium, 10,000–100,000 copies per ml and high, > 100,000 copies per ml; not applicable to SCs with unknown transmission source partners).28,29,30 Moreover, Kaplan–Meier curves were used to compare time-dependent events of HIV-1 seroconversion during the study interval (a secondary outcome). Using SAS version 9.3 (SAS Institute, Cary, NC, USA), the HRs (mean and 95% confidence interval) for HIV-1 acquisition were based on Cox proportional hazards models, with the HRs (mean and 95% confidence interval) for HIV-1 seroconversion as the outcome.

Bioinformatics: surveys of four public databases (last accessed on 19 December 2014)

SNPs displaying association signals in our analyses were first queried in HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php), for known linkage disequilibrium with SNPs uncovered by the 1000 Genomes Project (regardless of coverage by the ImmunoChip) or annotated by the ENCODE project.8,9,59 Additional surveys took advantage of the NCBI Global Cross-database (http://www.ncbi.nlm.nih.gov/), which captures SNPs associated with various human traits, diseases and/or cellular gene expression profiles.60 Further information from the SNP and CNV Annotation (SCAN) database (http://www.sncdb.org/newinterface/index.html) and the Finemapping Data Portal http://www.broadinstitute.org/pubs/finemapping/9,10 helped to identify cis- and trans-acting SNPs of functional importance. Overall, these searches were expected to provide a broad and comprehensive view of local genetic structure and SNP functions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported primarily by the National Institute of Allergy and Infectious Disease through R01 AI071906 (and its supplemental award) to RAM/ST and R01 AI064600 to EH. SA and EH received additional funding from the International AIDS Vaccine Initiative (Protocol C), with further support (UL1 RR025088) from the Clinical Translational Science Award program, National Center for Research Resources. We are grateful to members of the Rwanda–Zambia HIV-1 Research Group for their valuable contributions to patient enrollment. We also thank Paul Farmer, Naw Htee Khu, Hallin Lu and Travis R Porter for their assistance with sample inventories, genotyping and data management.

REFERENCES

1 CDC National Center for HIV/AIDS, VH, STD, and TB Prevention. HIV transmission risk: 2014. http://www.cdc.gov/hiv/pdf/policies_transmission_risk_factsheet.pdf.
2 Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG et al. Identification and characterization of transmitted and early virus envolunteers in primary HIV-1 infection. Proc Natl Acad Sci USA 2008; 105: 7552–7557.
3 Carlson JM, Schaefer M, Monaco DC, Batorsky R, Clabotire D, Prince J et al. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 2014; 345: 1254031.
4 Kaslow RA, Dorak T, Tang JJ. Influence of host genetic variation on susceptibility to HIV type 1 infection. J Infect Dis 2005; 191(Suppl 1): 568–577.
5 Lederman MM, Alter G, Daskalakis DC, Rodriguez B, Sieg SF, Hardy G et al. Determinants of protection among HIV-Exposed seronegative persons: an overview. J Infect Dis 2010; 202(Suppl 3): S333–S338.
6 Teleni A, McLaren P. Genomic approaches to the study of HIV-1 acquisition. J Infect Dis 2010; 202(Suppl 3): 3582–3586.
7 Petrovski S, Fellay J, Shaaniva KN, Carpenetti N, Kumwenda J, Kamanga G et al. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. AIDS 2011; 25: 513–518.
8 Lingappa JR, Petrovski S, Kalhe E, Fellay J, Shaaniva K, McElrath MJ 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: e26832.
9 Limou S, Delanoeo O, van Manen D, An P, Sezgin E, Le Clerc S et al. Multicohort genomewide association study reveals a new signal of protection against HIV-1 acquisition. J Infect Dis 2012; 205: 1155–1162.
10 Lane J, McLaren PJ, Dorrell L, Shaaniva KN, Stemke A, Pelak K et al. A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. Hum Mol Genet 2013; 22: 1903–1910.
11 McLaren PJ, Coupleses C, Rijke S, van den Berg L, Buchbinder S, Carrington M 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: e1003155.
12 Shin HD, Winkler C, Stephens JC, Bream J, Young H, Goedert JJ et al. Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. Proc Natl Acad Sci USA 2000; 97: 14467–14472.
13 Vasilescu A, Heath SC, Ivanova R, Hendel H, Do H, Mazoyer A et al. Genomic analysis of Th1-Th2 cytokine genes in an AIDS cohort: identification of IL4 and IL10 haplotypes associated with the disease progression. Genes Immun 2003; 4: 441–449.
14 Wang C, Song W, Lobashovsky E, Wilson CM, Douglas SD, Myllinen J et al. Cytokine and chemokine gene polymorphisms among ethnically diverse North Americans with HIV-1 infection. J Acquir Immune Defic Syndr 2004; 35: 446–454.
15 Chandler S, Stratdidee SA, Galal N, Oleksyk T, Fallin MD, Mehta S et al. Behavioral risk exposure and host genetics of susceptibility to HIV-1 infection. J Infect Dis 2006; 193: 16–26.
16 Erikstrup C, Kallestrup P, Zinyama-Gutsire RB, Gomo E, Butterworth AE, Pedersen BK et al. Reduced mortality and CD4 cell loss among carriers of the interleukin-10-1082G allele in a Zimbabwean cohort of HIV-1-infected adults. AIDS 2007; 21: 2283–2291.

© 2015 Macmillan Publishers Limited
et al.

34 Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention.

et al.

38 Couper KN, Blount DG, Wilson MS, Hafalla JC, Belkaid Y, Kamanaka M et al. IL-10 from CD4+CD25+Foxp3+CD127+ regulatory T cells modulates parasite clearance and pathology during malaria infection. PLoS Pathog 2008; 4: e1000004.

et al.

42 Ouyang W, Rutz S, Crelini NK, Valdez PA, Hymowitz SG. Regulation and functions of the IL-10 family of cytokines in inflammation and disease. Annu Rev Immunol 2011; 29: 71–109.

et al.

220 © 2015 Macmillan Publishers Limited

Host genetics and HIV-1 acquisition

X Li et al

Supplementary Information accompanies this paper on Genes and Immunity website (http://www.nature.com/gene)