CD4 intragenic SNPs associate with HIV-2 plasma viral load and CD4 count in a community based study from Guinea-Bissau, West Africa.

Branwen J. Hennig[a,b,*], PhD, Digna R. Velez-Edwards[c,d,e,*], PhD, Maarten F. Schim van der Loeff[f], MD, PhD, Cyrille Bisseye[g], MSc, Todd L. Edwards[c,e], PhD, Alessandra Tacconelli[d], MSc, Giuseppe Novelli[d,h], PhD, Peter Aaby[i], DMSc, Steve Kaye[j], PhD, William K. Scott[e], PhD, Assan Jaye[g], DVM, PhD, Hilton C. Whittle[g], MB, ChB, F Med Sci, Scott M. Williams[d], PhD, Adrian V. Hill[b], DPhil, DM, FRCP and Giorgio Sirugo[c,d,g,h], MD, PhD

[a] London School of Hygiene & Tropical Medicine, London, UK

[b] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

[c] Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN, USA

[d] Unita' di Genetica Medica, Ospedale S. Pietro FBF, Rome, Italy

[e] Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA

[f] Public Health Service Amsterdam, and Academic Medical Center, Center for Infection and Immunity Amsterdam, Amsterdam, The Netherlands

[g] Medical Research Council Laboratories (UK), The Gambia

[h] Dipartimento di Biopatologia e Diagnostica per Immagini, Tor Vergata University School of Medicine, Rome

[i] Bandim Health Project, INDEPTH Network, Bissau, Guinea-Bissau

[j] Jefferiss Trust Laboratory, Imperial College London, London, UK

[*] Equal contribution

CORRESPONDING AUTHOR

Giorgio Sirugo

Ospedale S. Pietro FBF, Rome, Italy

Tel: +390633585872; Fax: +390633553512

Email: sirugo.giorgio@fbfrm.it

Data presented previously at the annual 59th meeting of the American Society of Human Genetics, Honolulu, Hawaii, October 20-24 2009, abstract # 682.

This work was supported by the MRC award G0000690 to GS.

RUNNING TITLE: *CD4* SNPs associate with HIV-2 infection in Guinea-Bissau

ABSTRACT

Objectives: The human genetics of HIV-2 infection and disease progression is understudied. Therefore, we studied the effect of variation in two genes that encode products critical to HIV pathogenesis and disease progression: *CD4* and *CD209*.

Design: This cross-sectional study consisted of 143 HIV-2, 30 HIV-1 + HIV-2 and 29 HIV-1 infected subjects and 194 uninfected controls recruited from rural Guinea-Bissau.

Methods: We genotyped 14 *CD4* and 4 *CD209* single nucleotide polymorphisms (SNPs) that were tested for association with HIV infection, HIV-2 plasma viral load (high vs. low), and CD4+ T-cell count (high vs. low).

Results: The most significant association was between a *CD4* haplotype rs11575097-rs10849523 and high viral load (OR = 2.37, 95%CI [1.35-4.19], p = 0.001, corrected for multiple testing), suggesting increased genetic susceptibility to HIV-2 disease progression for individuals carrying the high-risk haplotype. Significant associations were also observed at a *CD4* SNP (rs2255301) with HIV-2 infection (OR = 2.36, 95% CI [1.19-4.65], p = 0.01) and any HIV infection (OR = 2.50, 95% CI [1.34-4.69], p = 0.004).

Conclusions: Our results support a role of *CD4* polymorphisms in HIV-2 infection, in agreement with recent data showing that *CD4* gene variants increase risk to HIV-1 in Kenyan female sex workers. These findings support at least some commonality in HIV-1 and HIV-2 susceptibility.

KEY WORDS: HIV-2, West-Africa, *CD4*, *CD209* (DC-SIGN), SNPs

INTRODUCTION

Human immunodeficiency virus (HIV) infection rates vary across Africa, with the highest adult prevalences in southern and eastern regions (up to approximately 30%) and the lowest rates in northern and western regions (0.1-5.0%).[1] HIV-1 is global, but HIV-2 occurs primarily in West Africa.[2,3] In Caio', rural Guinea-Bissau, where we performed our cohort study, we observed an increasing prevalence of HIV-1 and a decreasing prevalence of HIV-2.[4] Another study from an urban area in Guinea-Bissau confirmed the trends.[5]

Despite characteristics shared with HIV-1,[6] HIV-2 displays differences, such as lower mother to child transmission rates (1-4% vs. 20-40%), lower sexual transmission (1/3 to 1/4 that of HIV-1), lower plasma viral load (0-1000 vs. 10,000-100,000 copies/ml), slower CD4 decline, lower likelihood of progression to AIDS and generally slower progression to AIDS when it does occur (although fast progressors have been described).[6] However, clinical AIDS from HIV-2 is similar to HIV-1 with no substantial difference in mortality.[7]

The host genetics of HIV-1 has received considerable attention and reported associations include *HLA-B*, *HLA-C*, *CCR2*, *CCR5*, *CCL5* (*RANTES*) and *CCL3L1* variants.[8-16] Host genetics of HIV-2 is almost completely unstudied. One study from Senegal described an association of *HLA-B53* with increased risk of HIV-2 disease progression, consistent with previous findings on HIV-1 in Caucasians.[17] Ali and colleagues studied the gene encoding α (1,2) fucosyltransferase (*FUT2*) in HIV-1, HIV-2 and dually infected Senegalese women, and found a suggestive association with secretor status in HIV-1 but not in HIV-2 infection.[18]

Although it has been reported that HIV-2 infection protects against HIV-1 infection *in vitro* and to some extent *in vivo* (summarized by[3]), the epidemiological evidence does not support such an effect.[19]

A large number of biologically plausible candidate genes potentially exist for the study of host genetic factors in HIV infection pathogenesis and disease progression. Our report focuses

on two molecules that play central roles in critical pathophysiological processes relevant to HIV: T-cell antigen T4/Leu3 (encoded by *CD4*) and dendritic cell-specific ICAM3-grabbing nonintegrin (DC-SIGN, encoded by *CD209*). We genotyped single nucleotide polymorphisms (SNPs) in *CD4* and *CD209* in a cohort of HIV-1, HIV-2 and dually infected individuals from Guinea-Bissau.

MATERIALS AND METHODS

Study Area and Subjects

The study was conducted in Caio', a rural area in north-western Guinea-Bissau, comprising nine small, predominantly Manjago settlements (98% of subjects studied). Ethnicity was thus not adjusted for in the analyses. Since 1989 several HIV serosurveys have been conducted in this area.[19,20] All HIV infected persons and a similar number of age-, sex- and settlement-matched HIV uninfected controls were recruited for a cohort study, examined and sampled in 1991, 1996, and 2003. Our study is based on the 2003 study round, which comprised 547 eligible subjects. None of the study participants were on antiretroviral therapy (ART). As with any seroprevalent cohort or cross-sectional study, the HIV subjects are by definition survivors, and in this cohort there is an effect of HIV status on survival.[21] This study was approved by the MRC Laboratories/Gambia Government Joint Ethics Committee, the London School of Hygiene & Tropical Medicine Ethics Committee, and the Research Committee of the Ministry of Health of Guinea-Bissau.

All participants gave informed written consent.

Laboratory methods

Sample preparation. An 8 ml blood sample was taken from a cubitous vein in EDTA tubes. Two ml of the sample were used for routine laboratory tests and six ml were centrifuged

for 5 minutes at 2000 rpm, the plasma was drawn off, divided into three aliquots and frozen in liquid nitrogen. Lysis buffer was added to the cellular fraction for later DNA purification. Frozen plasma and samples in lysis buffer were transported to the MRC Laboratories in Fajara (The Gambia), where further analyses were performed and the DNA extractions were completed.

HIV sero-diagnosis. Plasma samples were screened for HIV antibodies using the ICE HIV1.2.0 assay (Abbott Murex Diagnostics, Dartford, UK). Positive samples were tested using two monospecific ELISA's: Wellcozyme HIV Recombinant I (Abbott Murex) and ICEHIV-2 (Abbott Murex) to differentiate between HIV-1 and HIV-2 infection. If one mono-specific ELISA was positive and the other negative, an individual was characterized as singly infected. In case of dual reactivity, the sample was diluted (1:10) and subjected to re-testing by the same two mono-specific ELISAs.

HIV diagnostic PCR. All samples that showed dual reactivity in the two mono-specific ELISA's (undiluted) were subjected to separate diagnostic nested PCRs for HIV-1 and HIV-2 in order to confirm dual infection. Both PCRs used primers targeted to highly conserved regions of the LTR of the HIV-1 or HIV-2 genome[22,23] and were used to amplify pro-viral DNA extracted from whole blood samples. Each reaction contained 600 ng of DNA, equivalent to 100,000 cells. Specific products were detected on ethidium bromide stained agarose gels.

HIV viral load. Virus load was measured using an in-house assay. Amplifications included an internal molecular control and products were quantified in an enzyme-linked oligonucleotide assay (ELONA).[22]

Lymphocyte subsets. The routine method was to use rapid field adapted Simultest monoclonal antibodies (Becton Dickinson, Erembodegem, Belgium) to stain and fix the cells using the Q-Prep machine (Beckman Coulter Ltd, High Wycombe, UK) in the field laboratory. A 10 µL blood sample was rapidly processed with Simultest CD3/CD4/CD8 conjugated antibodies (Beckton Dickinson, Erembodegem, Belgium) and fixed. The samples were then transported

overland to the Laboratório Nacional de Saúde Pública in Bissau, where they were analyzed using FACStrack and SimulSet software (Becton Dickinson, San Jose, CA, USA). FACS results were regarded as invalid if the CD3 % of the two runs were more than 10% different from each other, or if the mean of the two CD3 % was < 45%, or if the sum of CD4 % and CD8 % was more than 10% different from the mean CD3 %.

A second, back-up method for lymphocyte subset measuring was also in place. For this method, 100 µl "TransFix" solution (NEQAS, Sheffield, UK)[24] was added to 500 µl fresh blood and roller-mixed for 5 minutes, after which the sample was kept in a refrigerator at 4 ºC and transported to the MRC Laboratories in Fajara, The Gambia. When the routine method failed, the transfixed samples were stained with fluorochrome-conjugated monoclonal antibodies manually rather than by Q-Prep, and analyzed in the FACScalibur, using the MultiSet software (Becton Dickinson, Erembodegem, Belgium). Results were regarded as invalid if the CD3 % was < 45%, or if the sum of CD4 % and CD8 % was more than 10% different from the mean CD3 %.

In order to investigate whether the results of the two methods (routine and back-up) were comparable, 13 samples were processed using both methods and results compared. The agreement was excellent (r^2 = 0.89 for CD4 %). In addition, a Bland and Altman test[25] was performed to compare the two methods and indicated no difference between them (p = 0.3).

DNA extraction and genotyping. DNA was extracted from the packed blood cells, using the Puregene DNA extraction kit (Flowgen, Ashby de la Zouch, UK) or a standard salting-out procedure. SNP selection was based on physical location within the gene, minor allele frequency (MAF) (> 0.10 in the Yoruba population of the HapMap dataset (http://www.hapmap.org), with exception of *CD4* rs11064419), and assay availability. Chromosome band, SNP rs number, marker alleles, reference allele frequencies, amino acid change, and marker function (if known) are reported in Table 1 and were identified using the SNPper database (http://snpper.chip.org), based on NCBI Build 35.1. *CD209* markers were genotyped by standard TaqMan SNP assays

(ABI, Applera International Inc, Foster City, CA, USA) in 10 µl reaction volume, using the Rotor-Gene 3000 (Corbett Robotics Pty Ltd, Brisbane, Queensland, Australia) and the ABI 7500 real-time PCR systems. Fluorescence curves were analyzed with the Rotor-Gene Software version 6 and the 7500 Sequence Detection. *CD4* genotyping was carried out using the Sequenom hME Mass-Array primer extension assay (http://www.sequenom.de/)[26] under standard conditions. Primer sequences for *CD4* genotyping are shown in the supplementary materials (Table 1, Supplemental Content).

Statistical Methods

STATA 9.0 statistical software (STATA 9, College Station, TX, 2007) was used to perform the statistical analyses including the Shapiro-Wilk test, logistic regression and haplotype Odds Ratio (OR) and 95% Confidence Interval (CI) calculation.

Demographic data. Chi square tests were used to compare HIV status by gender, and t-tests were used to test for differences in the mean ages of study participants between HIV groups and controls. Age was tested for normality, using both visual inspection and Shapiro-Wilk tests, age did not differ from normality and as a result a t-test was performed.

Dichotomous analyses of HIV groups versus controls. Tests for deviation from Hardy Weinberg Equilibrium (HWE) were performed using Powermarker software.[27,28] Tests for allele and genotype associations were also performed with Powermarker software. Statistical significance for these analyses was determined using Fisher's exact tests. Logistic regression was performed on all SNPs with statistically significant allele or genotype tests of association modelling additive, dominant, and recessive models with the minor alleles as the risk allele. No correction for multiple comparisons was made at this stage (for further details see below).

Pairwise linkage disequilibrium (LD) was characterized and haplotype frequencies were calculated using Powermarker[27,28] and Haploview[29] statistical softwares. Standard summary

statistics D' and r^2 were calculated using Haploview.[29] Haplotype blocks were assigned, using the D' confidence interval algorithm created by Gabriel et al.[30] Both Powermarker and Haploview use an EM algorithm to determine haplotype frequency distributions when phase is unknown.

The Powermarker haplotype trend analysis was performed for *CD4* and *CD209* using 2- and 3-SNP sliding windows. This analysis is a regression approach to test haplotype-trait association.[27,28] The test for association uses an F test for a specialized additive model. Only haplotypes with a minimum of 5% frequency were considered in the analysis. Haplotype trend analysis produces global p values for tests of haplotype associations and not haplotype specific p values. P values were calculated using the permutation option (1000 permutations) to obtain a more reliable empirical p value. Odds Ratios (OR) and 95% confidence intervals (CI) for haplotype analysis were calculated for each individual haplotype with the referent being the most frequent haplotype. We used permutation tests to correct for multiple testing.

CD4 count and plasma viral load for HIV-2 analyses. CD4 count (cells/μl) and plasma viral load for HIV-2 (pvlhiv2) (copies/ml) were dichotomized into high and low groups for the analysis of associations within the group of HIV-2 infected subjects. The high and low groups for CD4 were ≥ 500 cells/μl versus < 500 cells/μl; the cut-off for CD4 is the lower limit of normal CD4 counts. The high and low groups for pvlhiv2 were ≥ 1000 copies/ml versus < 1000 copies/ml; this cut-off was based on an earlier study that showed that subjects with a viral load below this threshold (and with a normal CD4 count) had a normal life expectancy, i.e. similar to uninfected controls.[31] These data were analyzed for allele and genotype associations with Fisher's exact tests using Powermarker.[27,28]

RESULTS

Subjects

Of the 547 eligible subjects, 26 subjects died prior to enrollment, two had moved away from the area permanently, 97 subjects were away temporarily, and 15 refused to participate. All remaining 407 subjects gave a medical history and had a physical examination. Of these, 402 provided blood samples. Five of these samples did not yield usable DNA and clinical data was lacking for one additional subject, leaving 396 subjects for genetic analyses: 202 HIV-infected cases and 194 controls. The distribution of HIV-1, HIV-2 and dual infection status among cases was 29 (14.4%), 143 (70.8%), and 30 (14.9%), respectively. Analyses of demographic variables (Table 2, Supplemental Content) showed a difference in age between groups, with the HIV-1 group being significantly younger ($p < 0.001$). No statistically significant differences were observed between groups with respect to gender.

Dichotomous association analyses (HIV groups 1, 2, and dual infection versus controls)

HWE was tested in controls and there was one significant deviation from HWE at marker rs874627 ($p<1.00x10^{-3}$). Tests for HWE revealed significant deviations from expected in several, but not all, HIV groups including: rs874627, rs2255301, rs1045261, rs1055141, rs8105483, rs2071081 and rs3213427 (Table 3, Supplemental Content). Single locus tests identified associations in *CD4* at rs2255301, rs2886398, rs11064419, rs874627, rs2071081 with alleles and/or genotypes (Table 2; Table 3, Supplemental Content).

Additive, dominant and recessive logistic regression models for the SNPs that showed the strongest associations are presented in Table 3. For SNP rs2255301 the C/C&C/T vs. T/T had the strongest effect size across all outcomes: for HIV-2 (OR = 2.36 [CI 1.19-4.65]), dual infection (OR = 4.14 [CI 1.19-14.39]), and all HIV (OR = 2.50 [CI 1.34-4.69]). For SNP rs2071081 the strongest effect size was for the A/A&A/C vs. C/C model for HIV-1 (OR = 5.60 [CI 1.38-22.71]) and A/A vs. A/C&C/C5 for all HIV (OR = 1.58 [CI 0.94-2.67]) (Table 3).

Haplotype trend regression analysis revealed a significant 3 marker haplotype association (rs8105483-rs2287886-rs4804803) in *CD209* (p = 0.03) with HIV-1 (Table 4, Supplemental Content). No individual haplotypes associations were detected when comparing each haplotype in cases and controls to the referent (most frequent) haplotype; however CGG was marginally protective, (OR = 0.67 [CI 0.43-1.03] (p = 0.06)). There was no strong evidence of haplotype association for the *CD4* gene.

Analyses of HIV-2 examining CD4 count and plasma viral load in HIV-2 (pvlhiv2)

CD4 count significantly associated with both allelic and genotypic variation at *CD209* rs8105483 in patients who were HIV-2 infected (allele p = 0.02 and genotype p = 0.05). In analyses of plasma viral load, *CD4* markers rs11575097 (p = 0.03) and rs10849523 (p = 0.02) had significant allelic associations and rs11575097 had a significant genotypic association (p = 0.04).

Haplotype analyses showed a significant association for *CD209* rs8105483-rs2287886 (global p = 0.02) (Table 4) with CD4 count. Two individual haplotypes for these markers showed the following association: G-G (OR = 1.98 [CI 1.02-3.98] (p = 0.03)) with a frequency of 0.34 for the high CD4 count group and 0.18 for the low CD4 count group. The G-A haplotype had a frequency of 0.00 in the high CD4 count group and 0.10 in the low group; therefore, an OR was not calculated. Haplotype analyses assessing HIV-2 plasma viral load also identified a significant two-marker haplotype for markers rs11575097-rs10849523 (global p = 0.004) (Table 4). Haplotype G-G was the only significantly associated haplotype, with an OR = 2.37 [CI 1.35-4.19] and a p = 0.001. The frequency of G-G was 0.39 in the high group and 0.20 in the low group.

DISCUSSION

In the present study we examined eighteen SNPs in two host candidate genes, *CD4* and *CD209*, for association with HIV status, as well as plasma viral load and CD4 count in subjects infected with HIV-2. We observed statistically significant associations between individual *CD4* polymorphisms and HIV-2 status, HIV-2 plasma viral load and dual infection (the latter group being small). Although we did not observe evidence for disease status association with single *CD209* SNPs, we did observe some evidence for haplotype specific associations in analyses of the HIV-1 infected group compared to control groups (however this was based on small numbers of HIV-1 infecteds). More convincingly, we found evidence for *CD4* haplotype frequency differences between HIV-2 individuals with high and low viral load (OR = 2.37, 95%CI [1.35-4.19], p = 0.001).

CD4 is essential to immune regulation; it acts as an accessory protein for MHC class-II antigen and is involved in T-cell receptor interaction. CD4 is also a primary receptor for both HIV-1 and HIV-2, even if viral isolates of HIV-2 can infect T cells in the absence of CD4, using the chemokine receptor Fusin/CXCR4.[32] There is some evidence that laboratory-adapted CD4-independent HIV-1 infection also exists.[33] Although co-receptors are required for virus entry into the host cell and HIV-2 can use a wider range of co-receptors than HIV-1,[3] given the CD4 receptor's major role, the *CD4* gene is an obvious candidate for "susceptibility to HIV infection". Nonetheless, few studies have investigated the role of host genetic variation in *CD4*.

A recent study examining Kenyan female commercial sex workers[34] observed an association between HIV-1 infection and the *CD4* variant C868T. The C868T variant encodes a non-synonymous substitution at amino acid residue 240 of Arginine (R) to Tryptophan (W) known to change the tertiary structure of CD4 (the so called "CD4-Trp240" receptor), which could alter either the binding HIV to CD4 or CD4–dependent activation of T cells.[35-37] Oyugi et al (2009) showed that the minor allele of C868T is associated with an increased susceptibility to HIV-1 infection.[34] We did not type this specific SNP in our cohort; however, we did observe an

association with dual infection (p = 0.04) at a SNP (rs11064419) that is located less than 40 bp from C868T. These two SNPs are in strong LD based on HapMap data and our own data on controls suggests an LD block in this region (Figure 1, Supplemental Content). Since the other study did not test for association with HIV-2 or dual infection, their results cannot be directly compared to ours. Nonetheless, association within this small region of the *CD4* gene with both HIV-1 and HIV-2 in the two studies suggests that a variant in this region is of functional importance with respect to susceptibility to infection.

In our study, significant associations, both allelic and genotypic, were observed between SNP rs2255301 and infection status (Table 2). This SNP is located in intron 2, < 200 bases from exon 2 and in the boundary with exon 3. Both of these exons encode the amino-terminal D1 domain of *CD4*, an Ig variable (V)-like domain that binds with high affinity to the gp120 glycoprotein in the viral envelope.[38] Although we did not find evidence for significant LD between rs2255301 and other SNPs we typed around the D1 coding sequence, it is reasonable to hypothesize that rs2255301 is in LD with a functional, but uncharacterised, variation affecting this domain. Our lack of LD in this region is not surprising given the fact that our SNPs in this region are more than 4000 bp from rs2255301. However, an association with high plasma viral load (hence suggesting faster clinical progression) was detected with the rs11575097-rs10849523 G-G haplotype (Table 4) in intron 2, further reinforcing the idea that variation in this part of the gene is linked to increased susceptibility. Functional variation in *CD4* (and D1 in particular) could have a significant impact on susceptibility to HIV, as suggested by evidence of recent positive selection of intragenic polymorphisms.[39] However, a correlation between variants showing signatures of positive selection and those predicting risk has yet to be found.[39] This may be an artefact of how recent the pressure of HIV may be on this gene, the fact that only a minority of the population is exposed, and that different parts of the gene are susceptible to different selection pressures (from one or more infectious agents).

CD209 encodes the dendritic cell-specific ICAM3-grabbing nonintegrin (DC-SIGN), present on both macrophages and dendritic cells (DCs) and it belongs to the C-type lectin receptor family. HIV can subvert DC defense mechanisms and escape immune surveillance, allowing the virus to enter T cells, thus leading to infection of these cells.[40] Reports have associated variation in *CD209* with HIV-1 susceptibility, but findings from different studies and different ethnic groups are inconsistent and comparative genomics data from chimpanzees are unsupportive.[41-48]

The *CD209* role in transporting the virus via immature DCs from the periphery to lymph nodes, where CD4 cells would be activated and infected, could explain a link between *CD209* genetic variation and CD4 count. In our dataset the rs8105483-rs2287886 (in LD, Figure 2, Supplemental Content) G-G haplotype was associated with protection from disease (i.e. with "high" CD4 count; p = 0.03) however the "low count" CD4 subset was based on only five subjects carrying the G-A risk haplotype while no G-A were observed in the "high count" group. Given the small numbers no clear conclusion can be drawn.

Although we did not explicitly correct for multiple tests for single SNPs we did calculate empirical p-values by using permutation testing of haplotypes; similar size effect and level of significance were detected when haplotypes included positively associated single SNPs. This suggests that, despite no correction for multiple testing for single markers, the results are reliable and merit follow-up. *CD4* rs874627 deviated from HWE in controls, however, this SNP was not part of the haplotype associated with viral load in the HIV-2 group.

This study implicates *CD4* as susceptibility gene for HIV-2 infection in West Africans from Guinea-Bissau and, to our knowledge, is the first study examining the role of *CD4* and *CD209* variants in HIV-2 infection. The effect of *CD4* polymorphisms in our HIV-2 cohort, in addition to the association reported in the HIV-1 cohort from Kenya,[31] suggests that sequence variation in *CD4* plays a role at a fundamental biological level in the interaction between HIV

and host and that there may be some commonality in HIV-1 and HIV-2 susceptibility. Elucidation of the precise role this gene plays in HIV susceptibility depends on isolating functional variants that may explain the phenotypic effect.

References

1. UNAIDS. 2008 Report on global AIDS epidemic.
2. De Cock KM, Adjorlolo G, Ekpini E, et al. Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *Jama.* 1993;270:2083-2086.
3. Reeves JD, Doms RW. Human immunodeficiency virus type 2. *J Gen Virol*. 2002;83:1253-1265.
4. van Tienen C, van der Loeff MS, Zaman SM, et al. Two Distinct Epidemics: The Rise of HIV-1 and Decline of HIV-2 Infection Between 1990 and 2007 in Rural Guinea-Bissau. *J Acquir Immune Defic Syndr.* 2009;53:640-647.
5. da Silva ZJ, Oliveira I, Andersen A, et al. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *Aids.* 2008;22:1195-1202.
6. Schim van der Loeff M. Epidemiology, Natural History and Treatment of HIV-2. In: Volberding P, Sande M, Lange J, Greene W, Gallant J, eds. *Global HIV/AIDS Medicine*. Philadelphia: Saunders; 2008.
7. Martinez-Steele E, Awasana AA, Corrah T, et al. Is HIV-2- induced AIDS different from HIV-1-associated AIDS? Data from a West African clinic. *Aids.* 2007;21:317-324.
8. Ahuja SK, Kulkarni H, Catano G, et al. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. *Nat Med.* 2008;14:413-420.

9. Bhattacharya T, Stanton J, Kim EY, et al. CCL3L1 and HIV/AIDS susceptibility. *Nat Med.* 2009;15:1112-1115.

10. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science.* 2007;317:944-947.

11. Hill AV. Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet.* 2006;40:469-486.

12. Thomas R, Apps R, Qi Y, et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat Genet.* 2009;41:1290-1294.

13. Urban TJ, Weintrob AC, Fellay J, et al. CCL3L1 and HIV/AIDS susceptibility. *Nat Med.* 2009;15:1110-1112.

14. Sirugo G, Hennig BJ, Adeyemo AA, et al. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. *Hum Genet.* 2008;123:557-598.

15. Le Clerc S, Limou S, Coulonges C, et al. 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.

16. Pelak K, Goldstein DB, Walley NM, et al. Host determinants of HIV-1 control in African Americans. *J Infect Dis.* 2010;201:1141-1149.

17. Diouf K, Sarr AD, Eisen G, et al. Associations between MHC class I and susceptibility to HIV-2 disease progression. *J Hum Virol.* 2002;5:1-7.

18. Ali S, Niang MA, N'Doye I, et al. Secretor polymorphism and human immunodeficiency virus infection in Senegalese women. *J Infect Dis.* 2000;181:737-739.

19. Schim van der Loeff MF, Aaby P, Aryioshi K, et al. HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau. *Aids.* 2001;15:2303-2310.

20. Wilkins A, Ricard D, Todd J, et al. The epidemiology of HIV infection in a rural area of Guinea-Bissau. *Aids.* 1993;7:1119-1122.

21. van der Loeff MF, Larke N, Kaye S, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. *Retrovirology* 2010;7:46.

22. Berry N, Ariyoshi K, Jaffar S, et al. Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection. *J Hum Virol.* 1998;1:457-468.

23. Berry N, Ariyoshi K, Jobe O, et al. HIV type 2 proviral load measured by quantitative polymerase chain reaction correlates with CD4+ lymphopenia in HIV type 2-infected individuals. *AIDS Res Hum Retroviruses.* 1994;10:1031-1037.

24. Canonico B, Zamai L, Burattini S, et al. Evaluation of leukocyte stabilisation in TransFix-treated blood samples by flow cytometry and transmission electron microscopy. *J Immunol Methods* 2004;295:67-78.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.

26. Jurinke C, van den Boom D, Cantor CR, et al. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol.* 2002;187:179-192.

27. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 2005;21:2128-2129.

28. Zaykin DV, Westfall PH, Young SS, et al. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered.* 2002;53:79-91.

29. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21:263-265.

30. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science.* 2002;296:2225-2229.

31. Hansmann A, Schim van der Loeff MF, Kaye S, et al. Baseline plasma viral load and CD4 cell percentage predict survival in HIV-1- and HIV-2-infected women in a community-based cohort in The Gambia. *J Acquir Immune Defic Syndr.* 2005;38:335-341.

32. Endres MJ, Clapham PR, Marsh M, et al. CD4-independent infection by HIV-2 is mediated by fusin/CXCR4. *Cell.* 1996;87:745-756.

33. Xiao P, Usami O, Suzuki Y, et al. Characterization of a CD4-independent clinical HIV-1 that can efficiently infect human hepatocytes through chemokine (C-X-C motif) receptor 4. *Aids.* 2008;22:1749-1757.

34. Oyugi JO, Vouriot FC, Alimonti J, et al. A common CD4 gene variant is associated with an increased risk of HIV-1 infection in Kenyan female commercial sex workers. *J Infect Dis.* 2009;199:1327-1334.

35. Fleury S, Lamarre D, Meloche S, et al. Mutational analysis of the interaction between CD4 and class II MHC: class II antigens contact CD4 on a surface opposite the gp120-binding site. *Cell.* 1991;66:1037-1049.

36. Hodge TW, Sasso DR, McDougal JS. Humans with OKT4-epitope deficiency have a single nucleotide base change in the CD4 gene, resulting in substitution of TRP240 for ARG240. *Hum Immunol.* 1991;30:99-104.

37. Matthias LJ, Yam PT, Jiang XM, et al. Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat Immunol.* 2002;3:727-732.

38. Harrison SC. CD4: structure and interactions of an immunoglobulin superfamily adhesion molecule. *Accounts of chemical research.* 1993;26:449-453.

39. Ortiz M, Guex N, Patin E, et al. Evolutionary trajectories of primate genes involved in HIV pathogenesis. *Mol Biol Evol.* 2009;26:2865-2875.

40. van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. *Nat Rev Immunol.* 2003;3:697-709.

41. Bashirova AA, Geijtenbeek TB, van Duijnhoven GC, et al. A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. *J Exp Med.* 2001;193:671-678.

42. Li-Jun X, Brockmeyer NH, Bader A, et al. HIV-1 infected Germans have more variations on neck region of DC-specific intercellular adhesion molecule-3-grabbing nonintegrin than HIV-1 infected Chinese. *Eur J Med Res.* 2008;13:59-62.

43. Liu H, Hwangbo Y, Holte S, et al. Analysis of genetic polymorphisms in CCR5, CCR2, stromal cell-derived factor-1, RANTES, and dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin in seronegative individuals repeatedly exposed to HIV-1. *J Infect Dis.* 2004;190:1055-1058.

44. Rathore A, Chatterjee A, Sivarama P, et al. Role of homozygous DC-SIGNR 5/5 tandem repeat polymorphism in HIV-1 exposed seronegative North Indian individuals. *J Clin Immunol.* 2008;28:50-57.

45. Rathore A, Chatterjee A, Sood V, et al. Risk for HIV-1 infection is not associated with repeat-region polymorphism in the DC-SIGN neck domain and novel genetic DC-SIGN variants among North Indians. *Clin Chim Acta.* 2008;391:1-5.

46. Zhang J, Zhang X, Fu J, et al. Protective role of DC-SIGN (CD209) neck-region alleles with <5 repeat units in HIV-1 transmission. *J Infect Dis.* 2008;198:68-71.

47. Martin MP, Lederman MM, Hutcheson HB, et al. Association of DC-SIGN promoter polymorphism with increased risk for parenteral, but not mucosal, acquisition of human immunodeficiency virus type 1 infection. *J Virol.* 2004;78:14053-14056.

48. Santos PR, Michel-Salzat A, Butor C. Chimpanzee DC-SIGN alleles predict the existence of A and B isoforms, but do not support a role for resistance to HIV infection. *AIDS Res Hum Retroviruses.* 2005;21:820-829.

Table 1. Marker information

Gene	SNP rs#	Database Alleles	Reference Allele	Reference allele frequency in HapMap YRI[1]	Function	Amino Acid Change
	rs2855534	C/G	G	0.37	Promoter	-
	rs2707209	C/T	C	0.42	Intron	-
	rs2255301	C/T	T	0.38	Intron (boundary)	-
	rs2886398	A/G		N/A	Intron	-
	rs7957426	A/G	G	0.89	Intron	-
	rs11575097	C/G		N/A	Intron	-
CD4	rs10849523	C/G	C	0.66	Intron (boundary)	-
Chr 12p13.31	rs3782736	A/C	C	0.67	Intron	-
	rs11064419	C/T	T	0.96	Coding exon	F227S
	rs1055141	C/T	T	0.21	Coding exon	S341S
	rs3213427	G/T	T	0.77	3' UTR	-
	rs1045261	C/T	T	0.76	3' UTR	-
	rs874627	A/C	G	0.54	Downstream	-
	rs2071081	A/C	A	0.85	Downstream	-
	rs4804803	A/G	G	N/A	Promoter	-
CD209	rs2287886	A/G	A	0.2	Promoter	-
Chr 19q13.2	rs8105483	C/G	C	0.74	Intron	-
	rs11465413	A/T	A	0.48	3' UTR	-

[1] YRI: HapMap Yoruba data

Table 2. Statistically significant single marker analysis within each HIV subgroup and all groups combined

Gene	Marker	MA[3]	MAF[4] controls (N=194)	MAF Cases by HIV Group 1 (N=29)	2 (N=143)	Dual (N=30)	All (N=202)	Allelic association p Value 1	2	Dual	All	Genotypic association p Value 1	2	Dual	All
	rs2255301[2]	**T**	0.41	0.45	0.47	0.29	0.45	0.62	0.27	**0.008**	0.09	0.14	**0.02**	**0.005**	**0.008**
	rs2886398	**A**	0.34	0.35	0.36	0.5	0.36	0.73	0.54	**0.05**	0.32	0.9	0.48	0.11	0.51
CD4	**rs11064419**	**C**	0.05	0.02	0.06	0.12	0.05	0.47	0.51	**0.04**	0.58	0.46	0.41	0.15	0.44
	rs874627[1,2]	**G**	0.2	0.29	0.16	0.5	0.21	0.2	0.59	**0.04****	0.64	0.43	0.59	0.13	0.56
	rs2071081[2]	**C**	0.15	0.26	0.23	0.11	0.19	0.09	0.05	0.58	**0.05**	**0.03**	0.12	0.84	0.18

[1]This marker deviated from HWE in the controls: rs874627 ($p<1.00x10-3$). [2]These markers had significant deviations (p value ≤ 0.05) from HWE in one of the HIV groups: rs874627 (control group, $p < 1.00x10^{-3}$; HIV-1, p = 0.002: HIV-2, $p < 1.00x10^{-3}$), rs2255301 (all HIV group, p = 0.01; HIV-2, p = 0.01), rs1045261 (all HIV group, p = 0.01; HIV-2, p = 0.04); rs1055141 (all HIV group, p = 0.04; HIV-2, p = 0.02), rs8105483 (all HIV group, p = 0.04), rs2071081 (HIV-1, p = 0.02), and rs3213427 (HIV-2, p = 0.001). [3]MA-minor allele [4]MAF-minor allele frequency * In bold are all statistically significant associations ($p \leq 0.05$). Associations with HIV-2 and dual infection should be interpreted with caution based on the small sample size for these groups and because rs874627 was out of HWE in both cases and controls (although this may be due to chance).

Table 3. Associations between two single loci (*CD4* SNPs rs2255301 and rs2071081) and HIV infection, based on additive, dominant and recessive models

Marker	Outcome	Model	OR[1]	[95% CI][2]	p Value
rs2255301	HIV-2	C/C (referent), C/T, T/T[3]	1.25	0.86-1.84	0.25
		C/C&C/T (referent) vs. T/T[4]	2.36	1.19-4.65	**0.01**
		C/C (referent) vs. CT&T/T[5]	0.9	0.50-1.61	0.72
	Dual HIV	C/C (referent), C/T, T/T[3]	4.08	1.51-11.03	0.006
		C/C&C/T (referent) vs. T/T[4]	4.14	1.19-14.39	**0.026**
		C/C (referent) vs. C/T&T/T[5,6]	-	-	-
	All HIV	C/C (referent), C/T, T/T[3]	1.36	0.96-1.92	0.09
		C/C&C/T (referent) vs. T/T[4]	2.5	1.34-4.69	**0.004**
		C/C (referent) vs. C/T&T/T[5]	1.3	0.60-1.76	0.91
rs2071081	HIV-1	A/A (referent), A/C, C/C[3]	1.75	0.89-3.42	0.1
		A/A&A/C (referent) vs. C/C[4]	5.6	1.38-22.71	**0.02**
		A/A (referent) vs. A/C &C/C[5]	1.46	0.57-3.72	0.43
	All HIV	A/A (referent), A/C, C/C[3]	1.49	0.98-2.27	0.06
		A/A&A/C (referent) vs. C/C[4]	2.01	0.66-6.17	0.22
		A/A (referent) vs. A/C&C/C[5]	1.58	0.94-2.67	0.08

[1]OR-odds ratio [2]CI-confidence interval [3]Additive model [4]Dominant model [5]Recessive model [6]The OR for the dual infection dominant model could not be calculated because there were no cases with the referent genotype (C/C). *In bold are all statistically significant associations ($p \leq 0.05$)

Table 4. Associations between haplotypes and CD4 count or plasma viral load among HIV-2 infected subjects

Outcome	Haplotype	Global p Value	Markers	Relative Frequency of haplotype		OR	[95% CI]	p Value
				High	Low			
CD4 count[1] High n = 88 Low n = 49	***CD209*** **rs8105483- rs2287886**	**0.02**	C-G	0.48	0.51	(Baseline)	-	-
			G-G	0.34	0.18	1.98	1.02-3.98	**0.03**
			C-A	0.18	0.21	0.95	0.47-1.96	0.88
			G-A	0	0.1	-	-	**0.001**
Plasma Viral Load[2] High n = 71 Low n = 72	***CD4*** **rs11575097- rs10849523 ****	**0.004**	G-C	0.6	0.74	(Baseline)	-	-
			G-G	0.39	0.2	2.37	1.35-4.19	**0.001**
			C-G	0.01	0.04	0.25	0.01-2.30	0.18
			C-C	0	0.02	-	-	0.12

[1] CD4 count: ≥ 500 cells/µl (high) versus < 500 cells/µl (low)
[2] Plasma viral load: ≥ 1000 copies/ml (high) versus < 1000 copies/ml (low)
*In bold are all statistically significant associations ($p \leq 0.05$)
**Also significant was the three marker haplotype consisting of rs11575097- rs10849523- rs3782736 ($p = 0.009$), however, this was not the strongest window and therefore those results are not presented.

Supplemental Figure Legends

Supplemental Figure 1. *CD4* HIV-2 subgroup and control haplotype linkage disequilibrium (r^2 and D') plots

Linkage disequilibrium (LD) plots characterizing haplotype blocks in *CD4* in Guinea Bissau. All figures are oriented 5' to 3', right to left, relative to the gene orientation on the minus strand. In the first column are LD plots for pairwise D' between SNPs (control group **A**, HIV-2 subgroup **C**) and in the second column are LD plots for pairwise r^2 between SNPs (control group **B**, HIV-2 subgroup **D**). Both r^2 and D' values are indicated in percentages within squares in the LD plot. Strong LD is indicated by dark gray/red, while light gray/pink and white indicate uninformative and low confidence values, respectively. LD Blocks were created using the default algorithm in HaploView program, version 4.1.

Supplemental Figure 2. *CD209* HIV-1 and HIV-2 subgroup and control haplotype linkage disequilibrium (r2 and D') plots

Linkage disequilibrium (LD) plots characterizing haplotype blocks in *CD4* in Guinea Bissau. LD plots for pairwise r^2 and for pairwise D' between SNPs for control group (**A**, **B**), for HIV-1 subgroup (**C**, **D**) and HIV-2 subgroup (**E**, **F**)**.** Please refer to Supplemental Figure 1 legend for a description of the colour scheme used to define pairwise LD between SNPs. The haplotype blocks were created using the HaploView program, version 4.1.

Supplemental Table 1. SNP primer details

Gene	SNP_ID	Fwd primer	Rev primer	Extension primer
CD4	rs2855534	ACGTTGGATGTGCCATCTTTTTCTTGCCGC	ACGTTGGATGGCCAAAGTCAAGGGAAAAAG	GCTTAAATCAGGAGTCAAA
CD4	rs2707209	ACGTTGGATGATCAGTGACTTTCGCTCCTC	ACGTTGGATGGACATAGAGAGAGGACTTGG	GGCTAACTTAAGCCAAATT
CD4	rs2255301	ACGTTGGATGAAAGGCAAAGGTGGAGGATG	ACGTTGGATGAGGAGCGCTAAGTGGAAAAG	AGGAGTCAGGTCTCA
CD4	rs2886398	ACGTTGGATGGTCATCCACAAATAGGAACAG	ACGTTGGATGTGCCAGTGCAATAAGGCAAG	AGGAAAACTGATCAGAAAG
CD4	rs7957426	ACGTTGGATGTTCTCCACATCCTCATCAGC	ACGTTGGATGAAACCACAATAGGCTATCAC	AATAGGCTATCACTAAACCTATCA
CD4	rs11575097	ACGTTGGATGAACTCCTGCCCTCAGGTGAT	ACGTTGGATGAAAAGAGGTATGGGCCTGTG	CGCCTGTAATCCCAGCA
CD4	rs10849523	ACGTTGGATGCGATCATTCAGCTTGGATGG	ACGTTGGATGCATTCTCTTGCCAGAGGTGC	GCGCCTCAGTCCCCC
CD4	rs3782736	ACGTTGGATGTGAAATCCTGGTTCTGCCAC	ACGTTGGATGTCACAGAGGCTGAAACCAAG	GAGGTTTAGTAACTCACCCAAG
CD4	rs11064419	ACGTTGGATGCCTCCAGCATAGTCTATAAG	ACGTTGGATGCACTGCCCGTCAGCTTTTCA	CCCGTCAGCTTTTCAACTGTA
CD4	rs1055141	ACGTTGGATGATTTGACCTGTGAGGTGTGG	ACGTTGGATGACCTTTGCCTCCTTGTTCTC	CTTGTTCTCCAGTTTCAA
CD4	rs3213427	ACGTTGGATGTCCACGCCATTTCCTTTTCC	ACGTTGGATGTGGGATCCAAATGAGCAGTG	GGAGAGGGTCAGAGAGAAAT
CD4	rs1045261	ACGTTGGATGTGTGCAGAACCTCCTGGAAG	ACGTTGGATGCTGGCACAATCACCTTTGAG	CCTTTGAGGTCCCTT
CD4	rs874627	ACGTTGGATGAAGAGTGCTCAGAGACCATG	ACGTTGGATGTTCAGCCGCAGACCTCTCAT	TTCCCTCCTTCCCTC
CD4	rs2071081	ACGTTGGATGTTCTCCAGTGTCTGTTAGGC	ACGTTGGATGCCTCTCTTGCTATAGCTTTC	CTTTCTTTCACCCAGCTC
CD209	rs11465413, rs8105483, rs2287886, rs4804803	*CD209* SNPs were genotyped employing off-the-shelf TaqMan SNP assays (ABI, Applera International Inc, Foster City, CA, USA).		

Supplemental Table 2. Age and gender of study subjects, by HIV status

HIV group	Gender			Age		
	Female	Male	P [1]	Mean (yrs)	SD[3]	P [2]
All HIV groups	139 (68.8%)	63 (31.2%)	0.34	52.5	15.3	0.80
HIV-1	18 (62.1%)	11 (37.9%)	0.21	39.4	10.4	**0.0001**
HIV-2	96 (67.1%)	47 (32.9%)	0.23	56.1	15.0	0.96
HIV-1 & HIV-2 infection	25 83.3%)	5 (16.7%)	0.24	48.2	12.2	0.16
Control group	142 (73.2%)	52 (26.8%)		52.9	17.9	

[1]P values from chi square tests comparing HIV group to control group

[2]P values from t-test comparing mean ages between HIV group and control group

[3]SD is the standard deviation.

Supplemental Table 3. Single marker analysis within each HIV subgroup and all HIV subgroups combined

Gene	Marker	MA[3]	MAF[4] Controls	MAF Cases by HIV Group				Allelic Association p Value				Genotypic Association p Value			
				1	2	Dual	All	1	2	Dual	All	1	2	Dual	All
CD4	**rs2855534**	**G**	0.33	0.40	0.35	0.18	0.34	0.33	0.69	0.13	0.78	0.40	0.78	0.13	0.97
	rs2707209	**C**	0.40	0.39	0.40	0.29	0.39	0.87	0.93	0.21	0.76	1.00	0.30	0.51	0.4
	rs2255301[2]	**T**	0.41	0.45	0.47	0.29	0.45	0.62	0.27	**0.008**	0.09	0.14	**0.02**	**0.005**	**0.008**
	rs2886398	**A**	0.34	0.35	0.36	0.50	0.36	0.73	0.54	**0.05**	0.32	0.90	0.48	0.11	0.51
	rs7957426	**A**	0.10	0.08	0.08	0.12	0.09	0.79	0.55	0.76	0.61	1.00	0.53	0.59	0.73
	rs11575097	**C**	0.02	0.04	0.04	0.02	0.03	0.30	0.29	0.60	0.34	0.31	0.28	0.59	0.22
	rs10849523	**G**	0.29	0.28	0.31	0.36	0.30	1.00	0.55	0.50	0.52	0.75	0.80	0.44	0.86
	rs3782736	**A**	0.44	0.44	0.41	0.38	0.43	0.88	0.50	0.61	0.46	0.49	0.29	0.56	0.20
	rs11064419	**C**	0.05	0.02	0.06	0.12	0.05	0.47	0.51	**0.04**	0.58	0.46	0.41	0.15	0.44
	rs1055141[2]	**T**	0.12	0.14	0.17	0.12	0.14	0.44	0.10	1.00	0.13	0.52	0.23	0.61	0.39
	rs3213427[2]	**C**	0.13	0.15	0.16	0.11	0.14	0.65	0.33	1.00	0.57	0.55	0.40	0.81	0.68
	rs1045261[2]	**C**	0.30	1.00	0.40	-	0.46	0.07	1.00	-	0.25	0.59	1.00	-	0.38
	rs874627[1,2]	**G**	0.20	0.29	0.16	0.50	0.21	0.20	0.59	**0.04**	0.64	0.43	0.59	0.13	0.56
	rs2071081[2]	**C**	0.15	0.26	0.23	0.11	0.19	0.09	0.05	0.58	**0.05**	**0.03**	0.12	0.84	0.18
CD209	**rs11465413**	**A**	0.44	0.36	0.41	0.42	0.42	0.26	0.40	0.67	0.34	0.57	0.71	0.58	0.54
	rs8105483[2]	**G**	0.32	0.33	0.33	0.40	0.33	0.78	0.73	0.24	0.49	0.18	0.29	0.18	0.11
	rs2287886	**A**	0.26	0.17	0.22	0.25	0.24	0.14	0.27	0.87	0.25	0.33	0.61	1.00	0.43
	rs4804803	**A**	0.47	0.50	0.52	0.50	0.49	0.77	0.15	0.67	0.17	0.28	0.37	0.93	0.36

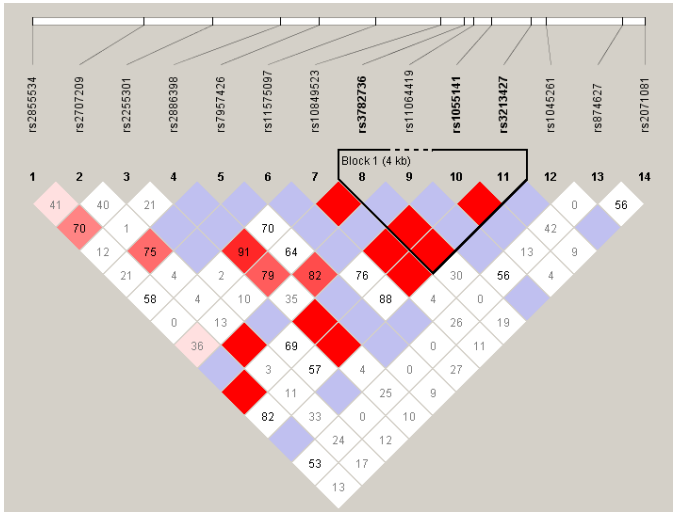
[1]This marker deviated from HWE in the controls: rs874627 (p<1.00x10-3). [2]These markers had significant deviations (p value ≤ 0.05) from HWE in one of the HIV groups: rs874627 (control group $p < 1.00x10^{-3}$; HIV-1 p = 0.002: HIV-2 $p < 1.00x10^{-3}$), rs2255301 (all HIV group p = 0.01; HIV-2 p = 0.01), rs1045261 (all HIV group p = 0.01; HIV-2 p = 0.04); rs1055141 (all HIV group p = 0.04; HIV-2 p = 0.02), rs8105483 (all HIV group p = 0.04), rs2071081 (HIV-1 p = 0.02), and rs3213427 (HIV-2 p = 0.001). [3]MA-minor allele [4]MAF-minor allele frequency * In bold are all statistically significant associations (p≤0.05). Associations with HIV-2 and dual infection should be interpreted with caution based on the small sample size for these groups and because rs874627 was out of HWE in both cases and controls (although this may be due to chance).

Supplemental Table 4. Haplotype analyses in HIV-1 vs. control group

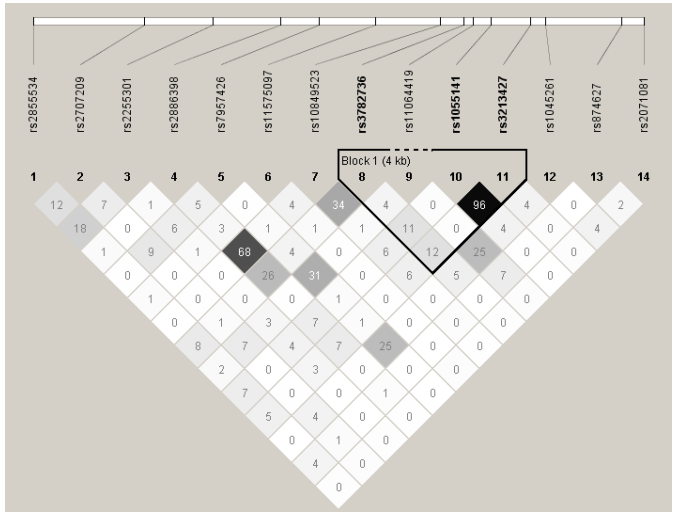
Outcome	Haplotype	Global p Value	Markers	Relative Frequency of haplotype		OR	[95% CI]	p Value
				HIV-1	Control Group			
HIV-1	***CD209*** **rs8105483-rs2287886-rs4804803**	**0.03**	G-G-G	0.32	0.31	(Baseline)	-	-
			C-G-A	0.31	0.23	1.29	0.87-1.89	0.18
			C-A-A	0.20	0.24	0.82	0.55-1.23	0.32
			C-G-G	0.14	0.21	0.67	0.43-1.03	0.06

Supplemental Figure 1. *CD4* HIV-2 subgroup and control haplotype linkage disequilibrium (r^2 and D') plots

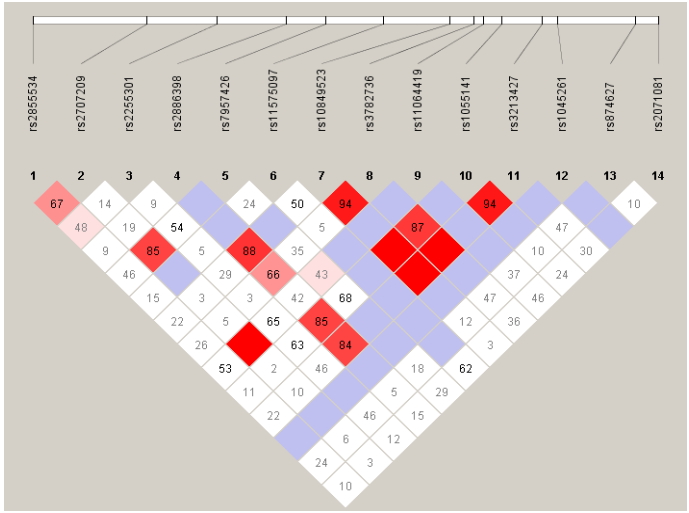
A. *CD4* control group D'



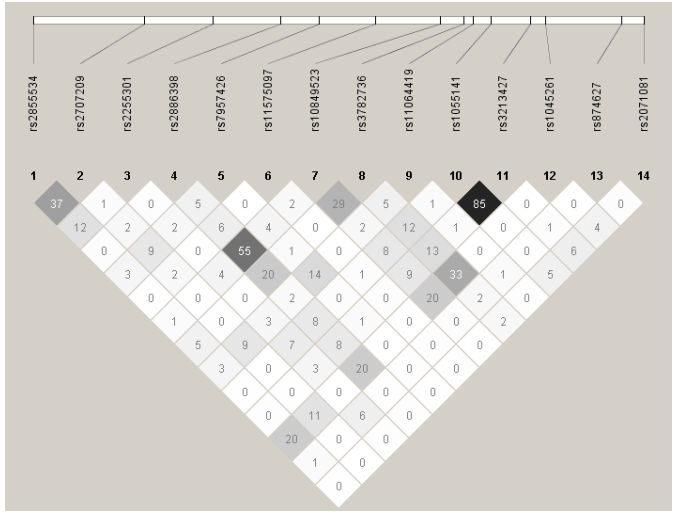
B. *CD4* control group r^2



C. *CD4* HIV-2 D'

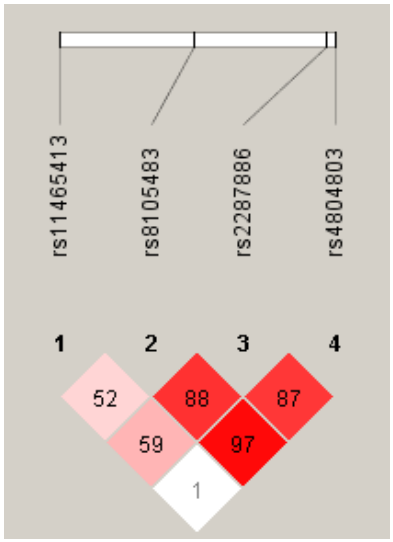


D. *CD4* HIV-2 r^2

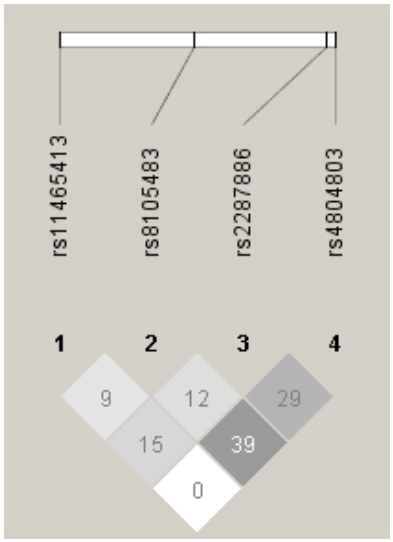


Supplemental Figure 2. *CD209* HIV-1 and HIV-2 subgroup and control haplotype linkage disequilibrium (r^2 and D') plots

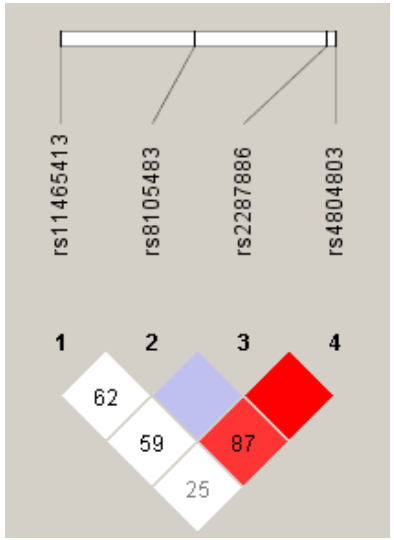
A. *CD209* control group D'



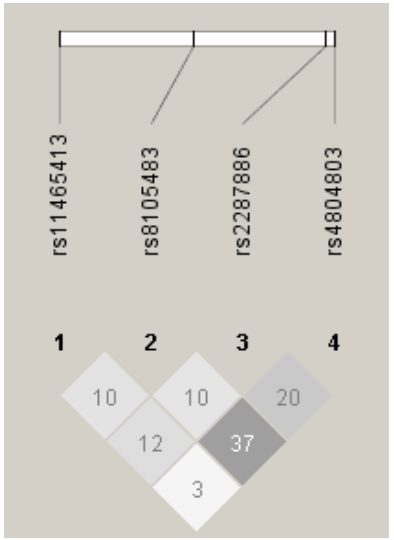
B. *CD209*control group r^2



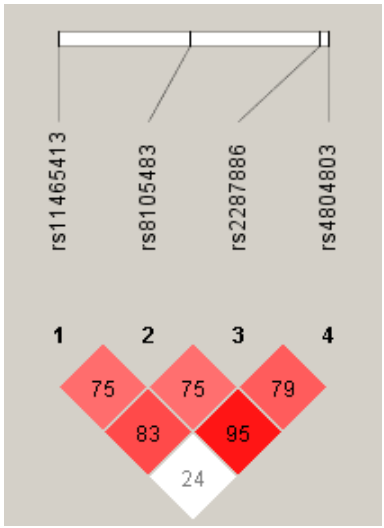
C. *CD209* HIV-1 D'



D. *CD209* HIV-1 r^2



E. *CD209* HIV-2 D'



F. *CD209* HIV-2 r^2