The influence of host genetic factors on HIV-1 infection
Manen, D.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
MULTI-COHORT GENETIC ASSOCIATION STUDY REVEALS CXCR6 AS A NEW CHEMOKINE RECEPTOR INVOLVED IN AIDS LONG-TERM NON-PROGRESSION

Sophie Limou¹,²,³,⁴, Cédric Coulanges¹,³, Joshua T. Herbeck⁵, Daniëlle van Manen⁶, Ping An⁷, Sigrid Le Clerc¹,²,³, Olivier Delaneau¹, Gora Diop¹,⁴, Lieng Taing¹, Matthieu Montes¹, Angélique B. van’t Wout⁶, Geoffrey S. Gottlieb⁵, Amu Therwath⁸, Christine Rouzioux³, Jean-François Delfraissy³, Jean-Daniel Lelièvre², Yves Lévy⁶, Serge Hercberg⁹, Christian Dina¹⁰, John Phair¹¹, Sharyne Donfield¹², James J. Goedert¹³, Susan Buchbinder¹⁴, Jérôme Estaquier²,¹⁵, François Schächter¹, Ivo Gut⁴, Philippe Froguel¹⁰,¹⁶, James I. Mullins⁸*, Hanneke Schuitemaker⁶*, Cheryl Winkler⁷* and Jean-François Zagury¹,²,³#

¹Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France ²Université Paris 12, INSERM U955, Créteil, France ³ANRS Genomic Group (French Agency for Research on AIDS and Hepatitis), Paris, France ⁴CEA/Institut de Génomique, Centre National de Génotypage, Evry, France ⁵University of Washington School of Medicine, Department of Microbiology, Seattle, WA, USA ⁶Department of Experimental Immunology, Sanquin Research, Landsteiner Laboratory, Center for Infectious Diseases and Immunity Amsterdam (CINIMA) Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands ⁷Laboratory of Genomic Diversity, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD, USA ⁸Laboratoire d’Oncologie Moléculaire, Université Paris 7, Paris, France ⁹UMR U557 Inserm/U1125 Inra/CNAM/UP13; Centre de Recherche en Nutrition Humaine Ile-de-France, SMBH Paris 13, Bobigny, France ¹⁰UMR CNRS 8090, Institut Pasteur de Lille, Lille, France ¹¹Feinberg School of Medicine, Division of Infectious Diseases, Northwestern University, Chicago, IL, USA ¹²Department of Biostatistics, Rho, Chapel Hill, NC, USA ¹³Infections and Immunoepidemiology Branch, NCI-Bethesda, Division of Cancer Epidemiology and Genetics, Rockville, MD, USA ¹⁴San Francisco Department of Public Health, HIV Research Section, San Francisco, CA, USA ¹⁵Assistance Publique Hôpitaux de Paris, Hôpital Henri Mondor, Créteil, France ¹⁶Genomic Medicine, Hammersmith Hospital, Imperial College London, London, UK

* These authors have equally contributed to the work

Journal of Infectious Diseases 2010; 202(6):908–915
ABSTRACT

The compilation of previous genomewide association studies of AIDS shows a major polymorphism in the *HCP5* gene associated with both control of the viral load and long-term nonprogression (LTNP) to AIDS.

To look for genetic variants that affect LTNP without necessary control of the viral load, we reanalyzed the genomewide data of the unique LTNP Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort by excluding “elite controller” patients, who were controlling the viral load at very low levels (<100 copies/mL).

The rs2234358 polymorphism in the *CXCR6* gene was the strongest signal ($P = 2.5 \times 10^{-7}$; odds ratio, 1.85) obtained for the genomewide association study comparing the 186 GRIV LTNPs who were not elite controllers with 697 uninfected control subjects. This association was replicated in 3 additional independent European studies, reaching genomewide significance of $P_{\text{combined}} = 9.7 \times 10^{-10}$. This association with LTNP is independent of the combined *CCR2–CCR5* locus and the *HCP5* polymorphisms.

The statistical significance, the replication, and the magnitude of the association demonstrate that CXCR6 is likely involved in the molecular etiology of AIDS and, in particular, in LTNP, emphasizing the power of extreme-phenotype cohorts. CXCR6 is a chemokine receptor that is known as a minor coreceptor in human immunodeficiency virus type 1 infection but could participate in disease progression through its role as a mediator of inflammation.
INTRODUCTION

Previous genomewide association studies (GWASs) of AIDS have revealed a major association involving a genetic polymorphism within the human leukocyte antigen region, the rs2395029 $HCP5$ single-nucleotide polymorphism (SNP), which tracks $HLA-B^*5701$. This SNP was associated with viral load control through analysis of human immunodeficiency virus type 1 (HIV-1) seroconverters [1,2], and by the comparison of patients with long-term nonprogression (LTNPs) with uninfected control subjects as well [3]. LTNPs are a small percentage (1%–5%) of HIV-1 seroconverters [4-6] and thus constitute a powerful contrasting tool to unravel new genetic factors associated with AIDS progression. Of the LTNPs in the Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort, patients carrying the $HCP5$ rs2395029-G allele exhibited a significantly lower viral load than the rest of the cohort [3]. Only a minority (ie, 25%) of the GRIV LTNPs exhibited effective viral load control (ie, a very low viral load of <100 copies/mL). Because viral load is known to account for only 34% of the variability in the time to a CD4 T cell decrease of <200 cells/µL [7], we decided to perform a new analysis of the genomewide data for GRIV LTNPs by excluding these “elite controller” subjects (subjects who had a viral load of <100 copies/mL). The aim of the current study was thus to focus on genetic variations affecting LTNP without necessarily controlling the viral load at a very low level. The result is that we have indeed identified a new specific signal in the $CXCR6$ gene and have replicated this finding in 3 additional independent cohorts of European descent.

METHODS

The GRIV Study: Participants, Genotyping, and Analysis

The GRIV study cohort and methods were described in detail in previously published work on the genomewide association study of LTNPs [3]. The GRIV cohort was established in France in 1995 to generate a large collection of DNAs for genetic studies seeking to identify host genes associated with rapid and LTNP to AIDS [8-11]. Only white people who were of European descent and were living in France were eligible for enrollment, to reduce confounding by population substructure. The LTNPs were seroprevalent subjects who were included on the basis of their main clinical outcomes, CD4 T cell count, and time to disease progression: asymptomatic HIV-1 infection for 18 years, no receipt of antiretroviral treatment, and a CD4 T cell count consistently 500 cells/mm³. Among those in the LTNP group ($n = 275$), viral load (ie, the plasma HIV-1 RNA concentration) at the time of inclusion could be assessed for 248 individuals. Of these 248 individuals, 186 had a viral load >100 copies/mL. All subjects provided written, informed consent before their enrollment in the GRIV genetic association study.
The control group

The Data from an Epidemiological Study on Insulin Resistance Syndrome (DESIR) program was a 9-year follow-up study designed to clarify the development of the insulin resistance syndrome. During 1994–1996, subjects were recruited from volunteers insured by the French social security system, which offers periodic health examinations free of charge [12]. This control group was comprised of 697 nonobese and normoglycemic individuals, and all were French, of European descent, and HIV-1 seronegative.

Genotyping method and quality control

The GRIV cohort and the control group were genotyped using the Illumina Infinium II HumanHap300 BeadChips (Illumina). Genotyping quality was assessed using BeadStudio software (version 3.1; Illumina). Missing data (>2%), low minor allele frequency (<1%), and deviations from Hardy-Weinberg equilibrium in the control group ($P < 1.0 \times 10^{-3}$) were excluded from analysis during these quality control steps. Moreover, identification of potential population stratification was identified using Structure software (version 2.2) [13], by producing a quantile-quantile plot (see Figure A1A in the Appendix, which appears only in the electronic version of the Journal) and by computing the genomic inflation factor $\lambda$. Overall, little effect of stratification was observed, and 283,637 SNPs could be tested statistically for association with LTNP.

Statistical analysis

For each SNP, we performed a standard case-control analysis, using Fisher’s exact tests (with Plink software [14]) to compare allelic distributions between LTNP and the control subjects. Bonferroni corrections were made to account for multiple comparisons.

SNP imputation

Untyped SNPs present in the HapMap database of chromosome 3 were imputed for all GRIV patients and control subjects, by use of Impute software (version 2.1) [15], and the HapMap release 21 phased data for the white population (CEU) as the reference panel [16].

CXCR6 genotyping by PCR sequencing

Primers and conditions used for polymerase chain reaction (PCR) amplifications were standard. Sequencing reactions were performed according to the Dye Terminator method by use of an ABI Prism 3730XL DNA Analyzer (Applied Biosystems). Alignment, SNP discovery, and genotyping were performed using the software Genalys, which was developed by the Commissariat à l’Énergie Atomique/Centre National de Génotypage [17].
Haplotype inference was obtained using the rapid and accurate Shape-IT algorithm [18].

Bioinformatics exploration
To further explore the associations observed, we tried to identify modifications in messenger RNA (mRNA) expression (Genevar [19] and Dixon [20] databases), splicing (NetGene2 [21]), polyadenylation (polyAH [22] and polyApred [23]), and transcription factor–binding sites (SignalScan [24], TESS [25], and TFSearch [26], derived from the TRANSFAC database).

Replication in the Amsterdam Cohort Study: Participants, Genotyping, and Analysis
The Amsterdam Cohort Study (ACS) participants and methods were described in detail elsewhere [27]. In the present study, 335 HIV-1–infected homosexual men from ACS were analyzed for the course of HIV-1 infection using AIDS-related death as an end point. AIDS-related death is defined as death with AIDS-related malignancy, death with AIDS opportunistic infections, or death with an AIDS-related cause not specified by the treating physician. The ACS rs2234358 genotyping data were obtained using Illumina Infinium II HumanHap300 BeadChips (Illumina). Quality control filters were applied to ensure reliable genotyping data. Potential population stratification was also analyzed using Structure software (version 2.2) [13], and 19 participants were thus excluded from further analyses (n = 316) because they differed significantly from the HapMap white population. Statistical analysis was performed by Kaplan-Meier survival analysis and determination of the log rank P value under the genotypic model, by use of SPSS software (version 16.0; SPSS) and the R package [28]. Because the viral load (ie, the plasma HIV-1 RNA concentration) and the CD4 T cell count were assessed during routine clinical follow-up, we could identify the ACS LTNPs who matched the GRIV definition and exhibited a viral load of >100 copies/mL (n = 31). LTNP status was easily determined for seroconverters because the date of seroconversion was known, and this was also the case for seroprevalent subjects, because the time of seropositivity was imputed (on average, at 18 months before enrollment).

Replication in the Multicenter AIDS Cohort Study: Participants, Genotyping, and Analysis
Multicenter AIDS Cohort Study (MACS) participants and methods previously have been described in detail elsewhere [29]. GWAS data were collected from 156 HIV-1–infected white homosexual men, with time to clinical AIDS used as an end point. This panel was chosen to be enriched with extreme AIDS progression phenotypes. The MACS rs2234358 genotyping data were obtained using the Affymetrix GeneChip Human Mapping 500K Array (Affymetrix), in which the rs2234358 SNP is tagged by rs4682799 (r² = 1). Quality control filters were applied to ensure reliable genotyping data, and population stratification was also checked. Statistical analysis was performed...
by Kaplan-Meier survival analysis and Cox proportional regression determination of the $P$ value under the genotypic model using the R package. As with the ACS, viral load (the plasma HIV-1 RNA concentration) and CD4 T cell count were assessed during routine clinical follow-up. We could extract 59 MACS LTNPs, selected from among seroconverters and seroprevalent subjects, who matched the GRIV definition and exhibited a viral load of $>100$ copies/mL.

Replication in the USA HIV-1 Cohort: Participants, Genotyping, and Analysis

USA HIV-1 cohort patients and methods previously were described in detail elsewhere [30]. For this study, 556 HIV-1 seroconverters of European ancestry were collected from 4 USA-based natural history HIV/AIDS longitudinal cohorts (MACS, San Francisco City Clinic Cohort, Multicenter Hemophilia Cohort Study, and Hemophilia Growth and Development Study), with AIDS-related death used as an end point. Of importance, the 556 USA HIV-1–infected participants did not include subjects overlapping with the 156 MACS participants. The USA HIV-1 cohort rs2234358 genotyping data were obtained using commercial TaqMan genotype assays (with assay ID C_1929536_1; Applied Biosystems). Conformity to the genotype frequencies expected under Hardy-Weinberg equilibrium was checked. Statistical analysis was performed by Kaplan-Meier survival analysis and Cox proportional regression for determination of the $P$ value under the genotypic model using the statistical SAS package (version 9.13; SAS Institute).

Independence from the CCR2–CCR5 Locus and HCP5 Polymorphisms

The genotypic data were available for the known CCR2–CCR5 locus and HCP5 polymorphisms in the GRIV, MACS, and USA HIV-1 cohorts, and it was thus possible to assess the independence of the rs2234358 SNP from these polymorphisms. For the GRIV cohort, multivariate logistic regression analysis was used to adjust effects of covariates CCR2–64I, CCR5-Δ32, CCR5-P1, and HCP5 rs2395029. The same approach was done for the MACS and USA HIV-1 cohorts but by fitting to the data a linear model instead of a logistic model. The independent effect of the rs2234358 SNP on disease phenotype was confirmed by adjusting the model with these covariates: the $P$ values that were obtained were similar with and without the covariate analysis.

RESULTS

After quality control tests, a case-control analysis using Fisher’s exact tests was performed to compare allelic distributions of the 283,637 SNPs between the GRIV LTNPs exhibiting a detectable viral load ($>100$ copies/mL) ($n = 186$) and uninfected controls ($n = 697$) (see Methods). The strongest association was found for rs2234358, with a $P$ value close to the $1.7 \times 10^{-7}$ Bonferroni threshold for genomewide significance (see Figure A1B in the Appendix, which appears only in the electronic version of
the *Journal*): $P = 2.5 \times 10^{-7}$ (odds ratio [OR], 1.85 [95% confidence interval [CI], 1.46–2.36]). The rs2234358-T allele was associated with not being an LTNP (36.83% in LTNP vs. 51.94% in controls) (Figure 1A). This allele was not associated with acquisition of HIV-1 infection, because its frequency was similar in seropositive and control groups: 51.16% among GRIV rapid progressors [31], 48.59% in the ACS, 51.92% in the MACS, 48% in the USA HIV-1 cohort, 54.9% in the Dutch control population (CTRACS), 48.3% in the Illumina controls (CTR_Illumina), and 50.9% in HapMap CEU (see Figure A2 in the Appendix, which appears only in the electronic version of the *Journal*).

The rs2234358 SNP lies within the *CXCR6* gene in a region of chromosome 3 that is rich in genes encoding chemokine receptors, and it is notably positioned 422 kb from the well-known *CCR5* gene [32] (Figure 2A). To eliminate possible tracking effects, we evaluated a potential association between the *CXCR6* signal and the *CCR2–CCR5* haplotypes (Δ32, P1, and 64I) previously associated with the course of HIV-1 disease [9,32,33]. First, the rs2234358 SNP had no linkage disequilibrium (LD, $r^2 > 0.1$) with any of the CCR5-Δ32, CCR5-P1, or CCR2–64I haplotypes. Second, we could not find any epistatic effects between rs2234358 and any of these haplotypes, by use of either Plink software [14] or logistic regression using *CCR2–CCR5* haplotypes as covariates (version 2.1) (see Methods). Third, the HapMap LD for whites did not reveal any SNP with a high LD ($r^2 > 0.9$) beyond the *CXCR6* locus. Of note, we also did not observe an epistatic effect between rs2234358 and the chromosome 6 rs2395029 *HCP5/HLA-B*5701 signal. This *CXCR6* signal thus represents a new association with LTNP, independent from the well-known *CCR2–CCR5* and *HCP5/HLA-B*5701 associations.

Of interest, we inferred the SNP distribution over the entire chromosome 3, using Impute software (see Methods). Instead of the 20,000 genotyped SNPs present in the Illumina HumanHap300 BeadChip in chromosome 3, a total of 176,000 SNPs could be imputed for which we could not identify a $P$ value better than the one exhibited for rs2234358. The attributable risk for rs2234358-T variant is very strong, because it explains 12% of the prevention of LTNP. For comparison, the attributable risk for CCR5-Δ32 is 5.1% in the GRIV LTNP cohort.

The rs2234358 signal was replicated in 3 independent additional cohort studies of white people of European descent that also evaluated for AIDS progression phenotype (after removal of stratification outliers; see Methods): (1) the European ACS cohort ($n = 316$) ($P = 2.3 \times 10^{-2}$) (Figure 1B), (2) a European descent subgroup of the American MACS cohort enriched in extreme phenotypes ($n = 156$) ($P = 4.2 \times 10^{-2}$) (Figure 1C), and (3) a pool of European American HIV-1 cohorts ($n = 556$) ($P = 8.6 \times 10^{-3}$) (Figure 1D and Table A1, the latter of which may be found in the Appendix, which appears only in the electronic version of the *Journal*). As shown in Figure 1, the rs2234358-T allele favored progression in all of these cohorts, which is in agreement with a prevention of LTNP. Overall, the combined $P$ value computed by the Fisher method between the 4 cohorts (GRIV, ACS, MACS, and USA HIV-1) passed the Bonferroni genomewide significance threshold: $P_{\text{combined}} = 9.7 \times 10^{-10}$. 


It was surprising to observe significant but rather weak $P$ values in all cohorts except the GRIV cohort, so we assessed whether the effect was specifically amplified in the LTNP subpopulation. We identified 31 and 59 LTNP fulfilling the GRIV definition and with a detectable viral load (>100 copies/mL) in the ACS and MACS cohorts,
respectively. In these groups, the rs2234358-T allele frequency was ~40%, which is similar to that found among the GRIV LTNPs (Figure 1A). Because no LTNP from these 3 cohorts differed significantly from the HapMap white population, according to the Structure analysis [13] (see Figure A3B in the Appendix, which appears only in the electronic version of the *Journal*), the ACS and MACS LTNPs were added to those in the GRIV cohort, and we computed a *P* value comparing this extended LTNP case group (*n* = 276) with the control group (*n* = 697). The *P* value again reached genomewide significance: *P* = 2.1 x 10^-8 (OR, 1.77 [95% CI, 1.44–2.18]), confirming the association of rs2234358- T with prevention of LTNP (Table A1, which may be found in the Appendix, which appears only in the electronic version of the *Journal*). Of importance, several additional control groups were tested and exhibited a similar allele frequency for rs2234358-T (see Figure A2 in the Appendix, which appears only in the electronic version of the *Journal*).

To further explore this association, we resequenced the entire CXCR6 gene to detect additional variants (Figure 2B and Table A2, the latter of which may be found in the Appendix, which appears only in the electronic version of the *Journal*): rs2234358 remained the SNP exhibiting the strongest association. Interestingly, using Shape-IT software (version 2.0) to compute haplotypes [18], we found several haplotypes comprising CXCR6 promoter SNPs in high LD (*r^2* = 0.97) with rs2234358 (Figure 2B).

The rs2234358 SNP is located in the 3' untranslated region of CXCR6, located 42 bp downstream from the termination codon (Figure 2B), and could thus influence gene expression, mRNA stability, mRNA regulation, or mRNA splicing. According to the Dixon or Genevar mRNA expression databases, none of the genotyped SNPs are predicted to affect CXCR6 or any other chromosome 3 gene expression, and

---

**Figure 2.** A, Genetic map of the CXCR6 gene region. CXCR6 is localized within the 14th intron of a predicted gene, FYCO1. B, Genetic map of the CXCR6 gene. Exons and untranslated regions are symbolized by full and empty rectangles, respectively. The positions of the ATG and STOP codons are indicated by a triangle (▲) and an asterisk (*), respectively. All single-nucleotide polymorphisms (SNPs) covered by the polymerase chain reaction sequencing study are represented, and the rs2234358 SNP of interest is shown in boldface type. The 3 promoter haplotypes in high linkage disequilibrium (LD) with rs2234358 (*r^2* = 0.97) correspond to 2-SNP haplotypes composed of the rs2234350 SNP (°), with either one of the SNPs denoted by the + symbol (these 3 latter SNPs exhibit *r^2* = 0.99).
bioinformatic methods failed to predict a modification of splicing or polyadenylation sites (see Methods). Nevertheless, we identified several putative transcription factor–binding sites containing the SNPs included in promoter haplotypes in high LD with rs2234358 (see Methods). Further experiments are required to determine the causative genetic variants and the biological mechanisms at stake.

**DISCUSSION**

Because the major signal identified in previous AIDS GWASs was associated with control of viral replication, we reanalyzed the genomewide data of the French GRIV LTNP cohort by excluding elite controller patients (ie, patients with a viral load of <100 copies/mL). The comparison of 186 LTNP patients with a viral load of >100 copies/mL and 697 uninfected controls highlighted a strong association for the CXCR6 rs2234358 ($P = 2.5 \times 10^{-7}$). This new signal was replicated by a candidate SNP approach in 3 additional independent European descent cohorts (including 316, 156, and 556 subjects), and the combined $P$ value of the 4 cohorts reached the genomewide significance threshold: $P_{\text{combined}} = 9.7 \times 10^{-10}$. This chromosome 3 combined association is independent from the well-known neighboring CCR2–CCR5 locus, is not linked with the control of viral load (the GRIV LTNP groups carrying the various rs2234358 genotypes exhibit a similar mean viral load) ($P = .72$, data not shown), and exhibits a high attributable risk of LTNP of 12%.

This study presents the first non-HLA–replicated association obtained through a GWAS approach. The $P$ value for rs2234358 is very strong in the GRIV cohort, and this signal was confirmed in 3 independent cohorts but with weaker $P$ values. The specific design of the LTNP phenotype can explain this discrepancy. Indeed, the extraction of LTNP patients with a viral load >100 copies/mL from the ACS and MACS cohorts confirmed the strength of this common SNP association with LTNP: ~40% versus ~50% in several uninfected control groups (see Figure 1A and Figure A2 in the Appendix, the latter of which appears only in the electronic version of the Journal). It emphasizes the critical importance of cohort design and the particular power of extreme phenotypes [5,34,35], particularly in light of a recent powerful GWAS involving >2500 patients, which solely identified chromosome 6–related signals [36].

The finding of a new chemokine receptor genetic variant contributing to a differential progression to AIDS is not so much of a surprise, because the chemokine system is a major weapon of the early host defense system against infectious diseases and comprises >100 members. An exonic CXCR6 variant present in African Americans (but absent in Europeans) was previously associated with Pneumocystis carinii pneumonia–mediated progression to AIDS [37]. Our CXCR6 genetic variant is not exonic, and its biological effect should rather be a modulation of CXCR6 expression. CXCR6, known as a minor HIV-1 coreceptor [38] and mediator of inflammation [39,40], is notably expressed in organs (thymus, gut, and bone marrow) and in immune cells [41], which are important for HIV-1 infection. It is involved in the trafficking of effector T cells mediating type 1
inflammation [40] and in the activation and homeostasis of natural killer T cells [42], known to be an important bridge between innate and adaptive immune responses. Interestingly, in simian immunodeficiency virus (SIV) infection, it has been proposed that interleukin-17–secreting natural killer T cells could compensate for the Th17 defect in the gut, because they are essential for controlling mucosal barrier integrity and microbial translocation [43,44]. These hypotheses are compatible with a major role of CXCR6 as an inflammation mediator in AIDS [39,40], but they deserve further functional/biological research to enhance our understanding of the molecular pathways to HIV-1 LTNP.

At a time when HIV-1 entry inhibitors such as CCR5 and CXCR4 antagonists are being developed, the identification of a molecular mechanism of AIDS pathogenesis involving a new chemokine receptor is of particular interest and opens new insights for therapeutic drug targets and prediction of HIV-1 progression.

**ACKNOWLEDGMENTS**

We thank all the patients and medical staff who have kindly collaborated with the Genomics of Resistance to Immunodeficiency Virus project. S.L. benefits from a fellowship from the French Ministry of Education, Technology and Research, and S.L.C. benefits from a fellowship of Agence Nationale de Recherches sur le SIDA et les Hépatites Virales Genomic Group.

Agence Nationale de Recherche sur le SIDA, Sidaction, Innovation 2007 program of Conservatoire National des Arts et Métiers, AIDS Cancer Vaccine Development Foundation, Neovacs SA, Vaxconsulting, and R37 (grant AI047734) from the US National Institutes of Health. This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health (NIH; contract HHSN261200800001E) and in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The authors acknowledge funding from the Netherlands Organization for Scientific Research (TOP; registration number 9120.6046). The Hemophilia Growth and Development Study is funded by the National Institutes of Health, National Institute of Child Health and Human Development, 1 R01 HD41224. The Amsterdam Cohort Studies on HIV infection and AIDS, a collaboration between the Amsterdam Health Service, the Academic Medical Center of the University of Amsterdam, Sanquin Research, and the University Medical Center Utrecht, are part of the Netherlands HIV Monitoring Foundation and are financially supported by the Netherlands National Institute for Public Health and the Environment. The authors have declared no conflict of interest.
REFERENCES


