In vitro whole-genome analysis identifies a susceptibility locus for HIV-1

LOEUILLET, Corinne, et al.

Abstract

Advances in large-scale analysis of human genomic variability provide unprecedented opportunities to study the genetic basis of susceptibility to infectious agents. We report here the use of an in vitro system for the identification of a locus on HSA8q24.3 associated with cellular susceptibility to HIV-1. This locus was mapped through quantitative linkage analysis using cell lines from multigeneration families, validated in vitro, and followed up by two independent association studies in HIV-positive individuals. Single nucleotide polymorphism rs2572886, which is associated with cellular susceptibility to HIV-1 in lymphoblastoid B cells and in primary T cells, was also associated with accelerated disease progression in one of two cohorts of HIV-1-infected patients. Biological analysis suggests a role of the rs2572886 region in the regulation of the LY6 family of glycosyl-phosphatidyl-inositol (GPI)-anchored proteins. Genetic analysis of in vitro cellular phenotypes provides an attractive approach for the discovery of susceptibility loci to infectious agents.

Reference

LOEUILLET, Corinne, et al. In vitro whole-genome analysis identifies a susceptibility locus for HIV-1. PLOS Biology, 2008, vol. 6, no. 2, p. e32

DOI: 10.1371/journal.pbio.0060032
PMID: 18288889

Available at:
http://archive-ouverte.unige.ch/unige:9214

Disclaimer: layout of this document may differ from the published version.
In Vitro Whole-Genome Analysis Identifies a Susceptibility Locus for HIV-1

Corinne Loeuillet1*, Samuel Deutsch2*, Angela Ciuffi1, Daniel Robyr2, Patrick Taffé3, Miguel Muñoz1,
Jacques S. Beckmann4*, Stylianos E. Antonarakis2*, Amalio Telenti1,3*

1 Institute of Microbiology, University Hospital, University of Lausanne, Lausanne, Switzerland, 2 Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospital of Geneva, Geneva, Switzerland, 3 Swiss HIV Cohort Study, Lausanne, Switzerland, 4 Department of Medical Genetics, University Hospital and University of Lausanne, Lausanne, Switzerland

Advances in large-scale analysis of human genomic variability provide unprecedented opportunities to study the genetic basis of susceptibility to infectious agents. We report here the use of an in vitro system for the identification of a locus on HSA8q24.3 associated with cellular susceptibility to HIV-1. This locus was mapped through quantitative linkage analysis using cell lines from multigeneration families, validated in vitro, and followed up by two independent association studies in HIV-positive individuals. Single nucleotide polymorphism rs2572886, which is associated with cellular susceptibility to HIV-1 in lymphoblastoid B cells and in primary T cells, was also associated with accelerated disease progression in one of two cohorts of HIV-1--infected patients. Biological analysis suggests a role of the rs2572886 region in the regulation of the LY6 family of glycosyl-phosphatidyl-inositol (GPI)–anchored proteins. Genetic analysis of in vitro cellular phenotypes provides an attractive approach for the discovery of susceptibility loci to infectious agents.

Citation: Loeuillet C, Deutsch S, Ciuffi A, Robyr D, Taffé P, et al. (2008) In vitro whole-genome analysis identifies a susceptibility locus for HIV-1. PLoS Biol 6(2): e32. doi:10.1371/journal.pbio.0060032

Introduction

Some individuals do not become infected to the HIV-1 virus despite repeated exposures, and among those that do, there is marked variation in the clinical course and progression to AIDS [1]. Although a number of host genetic determinants of susceptibility to HIV-1 have been identified through the analysis of candidate genes—most notably CCR5 Δ32 and HLA alleles—only a fraction of the observed phenotypic variation can be explained by variation at these loci [2,3]. Thus, there is a considerable interest in applying unbiased methods such as whole-genome analysis for the identification of novel susceptibility loci to human pathogens [1]. This hunt is, however, plagued by numerous confounding factors such as the lack of ascertainment of informative patient cohorts and difficulties to control for the variability of the infectious agent. Whole-genome mapping for viral susceptibility has been reported in mice for the murine adenovirus type 1 [4], and in mosquitoes for the dengue-2 virus [5]. Recently, the first genomewide association analysis for determinants of host control of HIV-1 in humans has been completed [6].

Whole-genome scans can also be performed through the analysis of family data using linkage analysis, an approach widely used to map monogenic disorders [7,8]. The need for family-based data has limited the use of this approach in the HIV-1 field because of the rarity, beyond instances of vertical transmission, of multi-case family infections. Studies of host genetic susceptibility to HIV-1 are also confounded by differences in virulence of the infecting viral strain [1].

To circumvent these limitations, we established an in vitro system to address the genetic control of cellular susceptibility to HIV-1 using cell lines from multi-generation families [9,10]. We used families from the Centre d’Etude du Polymorphisme Humain (CEPH) resource (up to four grandparents and an average of eight children per family), consisting of Epstein-Barr virus (EBV)–immortalized lymphoblastoid B cell lines (LCL). CEPH LCLs have been extensively genotyped, and the data are publicly available (http://snpdata.cshl.edu/population_studies/linkage__maps/). CEPH LCLs were previously used to identify genomic loci influencing sodium–lithium counter transport [11], natural variation in gene expression [12–15], transcriptional response to ionizing radiation [16], susceptibility to chemotherapy [17], and the relative impact of nucleotide and copy number variation on gene expression [18,19]. We hypothesized that CEPH LCLs could allow genome-wide investigation of interindividual variation of cellular susceptibility to infection with an isogenic virus, under standardized conditions and a controlled environment.

Results/Discussion

We designed the study to progress through consecutive steps including: (i) the identification by linkage and association of candidate markers associated with in vitro cellular susceptibility to an HIV-1-based vector, (ii) in vivo validation of candidate polymorphisms in humans infected with HIV,
Author Summary

Individuals differ in their susceptibility to the HIV-1 virus, and the determinants of susceptibility are encoded in the human genome. Genetic variants influencing this trait have been identified by investigating candidate genes thought likely to be involved in HIV-1 pathogenesis or by whole-genome association studies, which type more than 500,000 genetic variants per individual (genome-wide association studies) to see which ones associate with susceptibility. We have addressed the issue of identification of new genetic variants influencing susceptibility to HIV-1 by a novel strategy based on the in vitro infection of cells. For this, immortalized B lymphocytes from 15 families (198 cell lines) were infected by a HIV-based vector. Differences in cellular susceptibility to infection—a genetic trait—could be mapped to a precise region on Chromosome 8, suggesting a role of the LY6 family of GPI-anchored proteins in HIV-1 infection. Genetic analysis of in vitro standardized cellular phenotypes provides a new approach to the discovery of the basis of genetic susceptibility to infectious agents.

To identify genomic loci that contribute to the variation in cellular permissiveness to HIV.GFP, we performed a quantitative genome-wide linkage analysis using a panel of 2,600 SNP markers with an effective resolution of 3.9 cM. Calculations were performed using the variance components analysis option from Merlin [21]. A region on HSA8q24 (marker rs1398296) showed the highest multipoint linkage score (logarithm of the odds [LOD] = 2.89, $p = 1.3 \times 10^{-14}$; Figure 2A). To determine the significance of this finding, we performed 500 simulations in which genotypes were randomly chosen, but the phenotype was kept constant, so as to preserve the heritability of the trait, the trait density, and the missing data patterns [22]. The distribution of maximum LOD scores of the simulations revealed that the observed HIV.GFP linkage peak is significant on a genome-wide basis at the 95% significance level (Figure 2A).

In order to independently confirm and fine-map the linkage analysis result, we assayed LCLs from 56 unrelated CEPH individuals that have been genotyped at a high density in the frame of the HapMap project [23]. The association analysis was performed using 521 tag SNPs in a 3-Mb region centered on the initial linkage assignment. A single SNP, rs2572886G>A, was strongly associated with HIV.GFP permissiveness ($p = 1.8 \times 10^{-5}$), and statistical significance was maintained after Bonferroni correction for multiple testing and permutation analysis ($n = 10,000$) (Figure 2B). Allele A of marker rs2572886 is associated with an average 1.4-fold increase in susceptibility to the HIV.GFP in LCLs from unrelated individuals, $p = 0.001$ (Figure 2C). Similar steps were taken for the secondary study phenotypes unrelated to lentiviral cellular susceptibility, which led to the precise identification (by linkage and followed by association) of a region involved in cis-regulation of CD39 expression (Figure S2). In contrast, no locus was identified affecting LMP1 expression, suggesting a more complex control of this trait by multiple genes.

Because observations were all made on B cells, we next assessed the potential role of rs2572886 as a susceptibility factor for HIV-1 infection in CD4+ T cells. We genotyped the
SNP in a collection of purified CD4⁺ T cells obtained from 128 Caucasian healthy blood donors. CD4⁺ T cells were infected with a replicating HIV-1, and permissiveness was assessed by p24 antigen production [3]. A significant association was again obtained for SNP rs2572886 and cellular susceptibility to HIV-1 on this independent sample (p = 0.019) using a biological system that more closely resembles the in vivo situation. Consistent with the results of transduction of B cells with a HIV-1–based vector, in CD4⁺ T cells, the allele A of marker rs2572886 is associated with a 1.6-fold increase in susceptibility to infectious HIV-1 virus than CD4⁺ T cells of noncarriers, as assessed in a 7-d replication kinetics analysis (Figure 2D). The size effect associated with rs2572886 in vitro is comparable to that identified for other genetic variants influencing the HIV life cycle [3].

Since the previous results were obtained from in vitro assays, we set out to assess the potential association of rs2572886 with disease progression in HIV-1 infected individuals. The rs2572886 SNP has a minor allele frequency of 7% in Caucasians (Utah CEPH individuals), 19% in West Africans (Yoruba Hapmap sample), and 23% in Asians (Han Chinese and Japanese, HapMap sample). We genotyped 805 individuals recruited in the frame of the genetic project of the Swiss HIV Cohort Study (http://www.shcs.ch) who provided informed consent. These patients contributed consecutive CD4⁺ T cell data (n = 4,999 measurements) and viremia (n = 1,926 measurements) over an average follow-up period of 7 y in the absence of anti-retroviral drug treatment. The rs2572886A allele was associated with greater viral load, and faster progression of immunosuppression, as defined by the

Figure 2. Genome Scan and Fine Mapping of an HIV-1 Susceptibility Locus, and In Vitro Validation of the Candidate SNP Marker

(A) Linkage analysis with 15 families and 2,600 markers identifies a quantitative trait locus (QTL) on HSA8q24.3. The highest multipoint LOD score of 2.89, p = 1.3 × 10⁻⁴ (marker rs1398296) was significant at genome-wide level as determined by permutation analysis (95% significance threshold of 2.83, dotted line).

(B) Association analysis in 56 unrelated individuals, using 521 SNPs from the HapMap project centered 3 Mb around the initial QTL, identifies marker rs2572886 (p = 1.8 × 10⁻⁴). The association remained significant after correction for multiple testing by Bonferroni (dashed line) or permutation analysis (dotted line).

(C) The candidate marker is associated with a significant 1.4-fold increase in susceptibility to the HIV.GFP.

(D) rs2572886A is associated with a significant 1.6-fold increase in susceptibility of CD4⁺ T cells from healthy blood donors to replicating HIV-1, differences between the alleles remain significant (p = 0.0331) after removing the outliers.

doi:10.1371/journal.pbio.0060032.g002
slope of CD4+ T cells depletion over time (Figure 3A and 3B). Individuals homozygous for the minor allele variant (n = 7) exhibited, as a group, a faster disease progression, but no conclusions can be made with such a small sample size. The same trends were also present in the incident cohort of 259 individuals identified within a 1-y interval of seroconversion (Figure 3C and 3D), although the limited numbers precluded significant association. These results are consistent with the in vitro data, since the A allele, which was associated with higher susceptibility of infection in the cellular systems, was associated with greater viral load and faster progression in vivo.

As an additional validation step, we genotyped a second independent cohort including 189 individuals with a precise date of seroconversion (Figure 3E and 3F), which was collected in the context of a whole-genome association analysis [6]. No association was detected in this cohort; however, the power to detect association in a sample of this size is estimated to be around 25%. These patients were recruited by eight different cohorts, while the original results were established using Swiss HIV cohort data. When pooling the discovery incident subcohort and the validation cohort, the association of rs2572886 on both the CD4 and viral load did not reach significance despite the increased number of samples to 448. However, the combined sample is far from homogeneous. Thus this association should be considered as suggestive at this point.

The association results should be discussed in the frame of the recently published genome-wide association analysis of host determinants of viral setpoint [6]. First, the marker identified in the current study, rs2572886, is neither present nor tagged by the Illumina HumanHap550 BeadChip used in

---

**Figure 3.** Assessment of rs2572886A in an HIV-1–Infected Human Cohort in the Absence of Therapy

(A) Pattern of viral load in the full study population; each gray point represents a single viral load determination; the red (rs2572886A heterozygous), black line (individuals with the common allele), and orange lines (minor allele variant homozygous) represent the trend trajectories.

(B) Pattern of CD4+ T cell depletion in the full study population; each gray point represents a single CD4+ T cell determination.

(C) Pattern of viral load in infected individuals with known date of infection (incident sub-cohort).

(D) Pattern of CD4+ T cell depletion in the incident sub-cohort.

(E) Pattern of viral load in the validation incident cohort.

(F) Pattern of CD4+ T cell depletion in the validation incident cohort.

doi:10.1371/journal.pbio.0060032.g003
the paper by Fellay et al. [6]. In addition, the design and premises of the genome-wide association and those of the genome scan reported herein are different: (i) the manuscript by Fellay et al. led to the identification of acquired/innate immunity loci (in major histocompatibility complex (MHC)) that cannot be captured in a cellular assay that investigates the viral life cycle; (ii) the study design in the paper by Fellay was powered to detect only very strong and sufficiently common genetic determinants; (iii) the standardized infection conditions and the study endpoint (expression of a reporter) that were used in the current study are very different from the conditions encountered in a population of HIV-infected individuals. It is through the nature of these profound study design differences that we aimed at generating complementary information to that provided by the genome-wide association analyses.

To provide a reference parameter that would allow comparison with genetic variants identified in other studies, we estimated the proportion of variation explained by rs2572886 and compared it to the contribution of CCR5 Δ32 in the same study population. Including CCR5 Δ32 into the model increased the proportion of variation explained by 1.9% for the CD4 cell count and 0.4% for the viral load. For rs2572886A, the estimates were 0.8% and 1%, respectively. For comparison, age at infection contributed to an increase in the proportion of variation explained of 1.4% for the CD4 cell count and 0.05% for the viral load, whereas the increase was 3% and 3.1%, respectively, for gender. These values are comparable to estimates indicated in the literature for various genetic variants influencing HIV pathogenesis [24], although direct comparisons are difficult due to different study designs. In contrast, the proportion of variation explained by rs2572886 or CCR5 Δ32 is considerably smaller than that for the genetic determinants reported in the study by Fellay et al., reflecting the fact that this genome-wide association study was powered to detect strong genetic effects. Thus, HCP5 and HLA-C variants explained 9.6% and 6.5% of the total variation in viral load, respectively, and an SNP near the RNF39 and ZNDR1 genes explained 5.8% of the total variation in disease progression [6].

The rs2572886 SNP is located in a nonconserved intergenic region on the telomeric end of Chromosome 8q (Figure S4). It is flanked on both sides by genes of the LY6/uPAR family (Figure S3). The LY6 genes are characterized by conserved cysteine-rich domains with specific disulfide bonding patterns but with little homology (20%–30% amino acid conservation among family members); members are either glycosyl-phosphatidyl-inositol–anchored cell-surface receptors or secreted cytotoxins. Eight genes are located at 8q24.3: LY6K, SLURP1, LYPD2, LYNX1, LY6D, GML, LY6E, and LY6H. The functions of the encoded proteins are diverse but not well understood [25]. None has been associated with HIV-1 in the past, although the LY6H gene was reported to be up-regulated upon HIV infection [26]. A related protein of the LY6/uPAR family, the urokinase-type plasminogen activator receptor, coded by a gene in Chromosome 19, has been reported up-regulated in HIV-infected individuals, and proposed to participate in the innate immunity to HIV-1 through an interferon (IFN)-like mechanism [27,28].

The SNP rs2572886 is located in a recombination hot spot between two linkage disequilibrium blocks. We resequenced the surrounding region (~13 kb) in 30 chromosomes to identify additional SNPs in linkage disequilibrium with rs2572886 that might point toward a biological function. Although two closely positioned SNPs—rs12546765 and rs12546801—were associated with rs2572886 in this limited resequencing dataset, they are not found in linkage disequilibrium in HapMap (pairwise $r^2 = 0.03$). The region (~1 kb) where rs2572886 is located is only present in the human genome. We downloaded the homologous region of chimpanzee and Rhesus macaque and sequenced it in seven additional primates (bonobo, gorilla, orangutan, nomascus gibbon, siamang gibbon, baboon, and African green monkey). rs2572886G>A was particularly variable among primates, with “G” representing the ancestral nucleotide in Old World monkeys, “A” the ancestral residue in hominoids, and “T” in gibbons.

To identify candidate genes that could be functionally related to the rs2572886 SNP, we performed quantitative SC (chromatin conformation capture) [29], with the goal of detecting potential chromatin interactions between the SNP region and neighboring genes. We tested 11 regions by Taqman real-time PCR, spanning a distance of 190 kb surrounding the SNP. We focused primarily on the upstream areas (promoters) of genes in the locus. Results from cross-linked cells were compared to randomly ligated BAC DNA from the same region to correct for interassay differences and potential ligation biases. As expected, we observed a high level of enrichment with a region located 3.1 kb from the SNP (positive control) due to random chromatin interactions that have been reported to occur between regions separated by less than 5 kb [30]. The trend for enrichment rapidly decreased with increasing distance. Interestingly, we detected higher than background peaks of enrichment on the upstream areas of two genes—LY6D and LYPD2 (Figure 4)—suggesting that these are good candidates for functional interaction with the associated SNP. There are no apparent interactions between the SNP and the nearby GML promoter, despite its relative proximity (12 kb) in comparison to the LY6D (35 kb) and LYPD2 (70 kb) genes.

We prioritized the following proteins for additional biological assessment: LY6D and LYPD2 on the basis of chromatin conformation capture analysis, SLURP1 based on its unique status of secreted protein, and GML because of the proximity to the genetic marker. First, we overexpressed each of the four proteins from several vector backgrounds in 293T and HeLa cells to assess whether this would influence transduction by HIV.GFP. No significant changes in cellular infectivity were detected upon overexpression in these cell lines (Table S1).

In general, all eight genes of the LY6/uPAR family show detectable, but very low levels of expression as assessed by quantitative RT-PCR (unpublished); this precludes gene expression variation analysis to determine whether the genotype at rs2572886 correlates with expression levels of nearby genes. LY6D and LYPD2, which showed relatively higher expression levels, were silenced in HeLa cells by small interfering RNA (siRNA). Silencing with three different siRNA was successful for LY6D and suboptimal for LYPD2. After transduction with HIV.GFP, we observed minor modifications in rates of cellular infection (Table S1). These findings are interesting, but given the cell type used and the harsh treatment of the cells, they are not conclusive enough to make a functional link for these genes. Overall, the
biological basis for a role the LY6uPAR family of proteins in HIV-1 cellular susceptibility remains elusive after this first line of biological screening. Additional analyses will be required to convincingly demonstrate a role for these proteins in the HIV life cycle.

In summary, by using a multi-step procedure involving a whole-genome linkage scan followed by association studies, we identified a locus on HSA8q24 that influences cellular susceptibility to HIV-1, and possibly progression of HIV-1 infection in vivo. Although the initial findings were based on transduction of transformed B lymphoblastoid cells with a HIV-1-based vector, subsequent experiments first on primary CD4+ T cells infected with replicating HIV, and second on a cohort of untreated HIV patients, supported the initial observations of association. In addition, although quantitative 3C data suggest a possible participation of genes of the LY6uPAR family, further work is required to decipher the biological mechanism underlying this association.

Materials and Methods

Plasmids. Members of the LY6uPAR family—LY6D (GenBank NM_003695), LYPD2 (NM_205545), GML (NM_002066, SC303114; Origene), Lynxlc (NM_177457), Slurpl (NM_029427), and Slurp2 (NM_177458)—were amplified by RT-PCR from RNA extracted from human cell lines (HeLa or 293T), and cloned into pCI-neo.

Cell culture. 293T cells were cultivated in Dulbecco’s modified Eagle Medium (DMEM; Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 50 µg/ml gentamycin. 293T cells were infected with HIV-based particles (25 ng p24 equivalent) in the presence of 15% fetal calf serum (FCS, Inotech). They were maintained by replacing half medium twice a week. Pedigrees studied were numbers 102, 884, 1328, 1331, 1332, 1333, 1334, 1430, 1341, 1345, 1346, 1347, 1362, 1408, and 13292.

HIV-1 Susceptibility Locus

Transduction of transformed B lymphoblastoid cells with a HIV-1-based vector, subsequent experiments first on primary CD4+ T cells infected with replicating HIV, and second on a cohort of untreated HIV patients, supported the initial observations of association. In addition, although quantitative 3C data suggest a possible participation of genes of the LY6uPAR family, further work is required to decipher the biological mechanism underlying this association.

Figure 4. Long-Range Interactions between rs2572886 and LY6D/uPAR Gene Promoters

Curves show results of quantitative chromatin conformation capture (3C) for four independent CD4+ cell lines. The y-axis corresponds to interaction enrichment relative to background (furthest point) and corrected according randomly ligated BAC. The x-axis shows relative distance to rs2572886. doi:10.1371/journal.pbio.0060032.g004

HIV-1 Susceptibility Locus

Transduction of CEPH cells and of the B cells of healthy blood donors (0.5 × 10⁶ cells in 96 wells) was performed by spinoculation with HIV-based particles for 3 h at 1500 g and 22 °C. After 72 h, cells were harvested and expression of GFP protein was monitored by fluorescence activated cell sorting (FACS) analysis with mock transduced cells as control. This was performed twice in triplicate at 1-wk intervals. Similarly, transduced 293T cells (in six-well plates) were infected with HIV-based particles (25 ng p24 equivalent) in the presence of 10 µg/ml DEAE-dextran in 1 ml D-10 for 5 h. Culture medium was replaced and cells were incubated for 24 h before FACS analysis of GFP expression. Typically, 20%–30% of GFP+ cells were obtained for controls.

HIV infection. Healthy blood donors’ CD4+ T cells (10⁶ cells) were infected with NL4-3 BaL virus (1,000 pg of p24 antigen) in a 1-ml final
HIV-1 Susceptibility Locus

volume for 2 h at 37 °C in 5% CO2. Cells were washed and cultured for 7 d. Virus-containing supernatant was harvested, and p24 antigen production was monitored by an enzyme-linked immunosorbent assay (ELISA) (Abbott).

Phenotyping. For cell surface molecule staining, 10^6 CEPH cells were stained, resuspended in PBS/0.5% BSA (Sigma) and incubated with primary antibodies (mAbs) or isotype controls at 4 °C for 1 h, and then stained with secondary antibodies (IgG1) for 30 min. After two washes with Permix, cells were fixed (CellFix, Becton Dickinson) and analyzed using a FACS Calibur system for 10,000 events. Positive events were defined as a fluorescence level superior to that of isotypic control. Determination of EBV copy number was carried out by real-time PCR by using specific probes as described [33].

Heritability, linkage, simulations, and associations studies. Heritability calculations (h^2) were performed using the polygenic screen command from SOLAR software [34]. SNP genotyping data, consisting of 2,688 autosomal SNPs were downloaded from the SNP Consortium database (http://snpdata.cshl.edu/population_studies/ped_check/ped_check.html) and subsequently resuspended by the HapMap project (http://www.hapmap.org/cgi-perl/hgbrowser/hgbrowser/hapmap/). HapMap public release number 19.

Cohort analysis. Data from both incident (patients identified during primary infection; or who have had a negative and positive test for HIV infection within a narrow time interval, 1 y in this study, in which the date of infection is estimated as the mid-point), as well as data from prevalent cases (i.e., individuals already HIV-seropositive by the time they entered the study, unknown date of infection) were analyzed longitudinally by modeling the CD4 T cell count and HIV-1 RNA load (log scale) and the CD4 (square root scale) trajectories post infection. The analysis was conducted using population-averaged marginal modeling [39], because the focus of the study was to investigate the effect of specific genetic factors on disease progression at the population level. In a marginal model, the mean regression function is modeled independently from the variance–covariance matrix. We used fractional polynomials to assess the best-fitting functional form. The viral load (log scale) and the CD4 (square root scale) trajectories post infection were linear and appeared stationary. Therefore, linear functions of time, along with interactions with polymorphisms and time, were included in the model. The proportion of explained variation was assessed [40]. A multivariate distribution was fitted to the data by score-like methods (generalized estimating equations) [41]. The correlation structure was assumed to be exchangeable, and an autoregressive process of order 1. To limit the impact of frailty selection, only the data for the first eight years since seroconversion were considered. The analysis was repeated considering, in turn, only the incident, prevalent, and both cohorts. Subanalyses were also performed considering the Caucasian group only. In prevalent cases, an estimate of the unknown date of infection was obtained using the markers data and defining for each patient an infection window based on his or her last negative and first positive available HIV tests. The date of infection was then imputed using a methodology that extends published methods [42,43] to accommodate multiple marker measurements per individual [P. Taffé and M. May, unpublished data and [44]. A second incident cohort, recruiting individuals from various European countries, was used to validate results obtained from the analysis of the incident cohort recruited within the Swiss HIV Cohort study. Statistical analyses were conducted using SAS version 9.1 for Windows, as well as STATA 9.2.

Re-sequencing in human and nonhuman primates. The region (~1 kb) around the candidate marker was resequenced by using forward primer SG2000 (5'-AGTCTACACCCCTTTGCGAGTTTGG-3') and reverse primer SG1853 (5'-CTCTACAGCTACTCCCTTGGAGA-3') as well as forward primer SG1829 (5'-TTTCCCTAGCCTGAGGACT-3') and reverse primer SG1853 (5'-CTCTACAGCTACTCCCTTGGAGA-3') to generate overlapping PCR products.

Sequences have been submitted to GenBank for bonobo (EU340898, Pan paniscus), gorilla (EU340899, Gorilla gorilla), orangutan (EU340890, Pongo pygmaeus) and African green monkey (EU340894, Cercopithecus [chloreus] asethus). Quantitative chromatin conformation capture. Approximately 10^7 stimulated PBL T cells [31] were crosslinked in their media for 10 min at RT with 1% formaldehyde (v/v). Crosslinking was quenched with 125 mM glycine prior to two successive washes with 1xPBS. Pelleted cells were resuspended in 5 ml ice-cold lysis buffer (10 mM Tris HCl, pH 8.0, 10 mM NaCl, 0.2% (v/v) NP-40) complemented with protease inhibitors (Complete, Roche) and 0.5 mM PMSE. Lysis of the cells was allowed to proceed for 10 min on ice with mild shaking. Nuclei were recovered by centrifugation (600g for 5 min at 4 °C) and resuspended into 300-μl 1xDPN restriction buffer (NEB). Nuclei were lysed with 0.3% (v/v) SDS for 60 min at 37 °C. SDS was neutralized with 2% (v/v) glycerol for 30 min at 37 °C. Chromatin was subsequently restricted overnight at 37 °C with 500 U DpnII (NEB) in a final reaction volume of 600 μl. After heat inactivation of the restriction enzyme (10 min at 65 °C), chromatin was dialyzed (Slide-A-Lyzer, Pierce) for 1 h against 1.5 l of water at RT and transferred into 7 ml ligation reaction mix (50 mM Tris HCl, pH 8.0, 10 mM MgCl2, 0.5%SDS, 10 mM β-mercaptoethanol, 0.5 mM ATP and 400 U T4 DNA ligase (NEB). The ligation reaction was performed for 4 h at 16 °C followed by another 30 min at RT. Crosslinking was heat-reversed and proteins were degraded (300 μg proteinase K) overnight at 65 °C in a hybridization oven. DNA was purified using phenolchloroformisoamyl alcohol [25:24:1(v/v)] extraction, precipitated with isopropanol, and washed with ethanol 70%. DNA was subsequently resuspended with 200 μl 1xTE pH 8.0 and treated with 50 μg RNaseA for 30 min at 37 °C. Finally, DNA was extracted with 1 volume phenolchloroformisoamyl alcohol [25:24:1(v/v)], ethanol precipitated, and resuspended in 100 μl 1xTE pH 8.0. Crosslinking was independently performed on four CD4^+ lines derived from different individuals, two of whom (2 and 4) were heterozygous for rs2572886.

For quantitative Taqman PCR, we designed 11 assays comprising the PCR primers and a dual-labeled probe sitting at the predicted DpnII junction between the target and bait targets (primer and probe sequences are available upon request). Reactions were set up using a Biomek 2000 robot (Beckman), in a 10-μl volume in 384-well plates. Three replicates per assay per sample were performed. PCRs were run in an ABI 7500 Sequence Detection System (Applied Biosystems) with the following conditions: 2 min for 95 °C, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. Each reaction contained 300 nM of each primer and 250 nM of probe.

For the 3C samples, approximately 200 ng of DNA was used per well, and for the BAC (digested – randomly ligated) samples, 10 ng of DNA was used. Normalization for each assay was performed using the values obtained from BAC experiment (all assays are expected to give the same result, given that the naked BAC was fully digested and religated, and all ligation combinations are expected to be present equimolarly). Enrichment was calculated with respect to the most centromeric probes, which showed very low levels of interaction.

Networks. Recently, the authors [45] identified a large SRNA screen over 250 HIV dependency factors. Among these there were three members of the LYN6PAR family (GML, LY6D, and LYPD4). This new evidence provides independent support for a biological role of the LYN6PAR family in HIV-1 pathogenesis.

Supporting Information

Figure S1. Validation of the CEPH Model to Study HIV-1 Susceptibility

CD4^+ T cells and B cells from same healthy blood donors were transduced with the HIV-GFP. Linear regression of one representative experiment for the percentage of GFP-positive cells is shown. Each point represents the mean of triplicate values.

Found at doi:10.1371/journal.pbio.0060032.sg001 (41 KB PPT).
Figure S2. Expression and Heritability of HIV-1 Susceptibility Trait in CEPH Families
(A) Heritability estimate ($h^2$) for eight study traits (percent of positive cells and mean fluorescence intensity, MFI), and for the EBV copy number (** p < 0.001, *** p < 0.0001).
(B) Representation of interindividual (black dots), and interfamily number (**) cells and mean fluorescence intensity, MFI), and for the EBV copy of 198 lymphoblastoid B cell lines from 15 CEPH pedigrees. Found at doi:10.1371/journal.pbio.0060032.sg002 (68 KB PPT).

Figure S3. Genome Scan and Fine Mapping of CD39 Expression (MFI Trait)
(A) Results of linkage analysis with 15 families. A single locus on HSA10q23 (rs766083) is significant after correction by permutation analysis (n = 500).
(B) Results of association analysis in 56 independent individuals using SNPs from the HapMap project. All SNPs in a 700 kb region surrounding the CD39 gene were used, rs numbers next to peaks indicate significant SNPs after correction for multiple testing. Found at doi:10.1371/journal.pbio.0060032.sg003 (94 KB PPT).

Figure S4. Chromosomal Localization of rs2572886
The SNP is localized in an intergenic region containing genes belonging to the LY6/PAF family. The figure shows images generated through the University of California at Santa Cruz genome browser and Haploview software.

References
1. Teleni A, Goldstein B (2006) Genomics meets HIV. Nat Rev Microbiol 4: 9–18.
2. O’Brien SJ, Nelson GW (2004) Human genes that limit AIDS. Nat Genet 36: 565–574.
3. Bleiber G, May M, Martinez R, Meylan P, Ott J, et al. (2005) Use of a combined ex vivo/in vivo population approach for screening of human genes involved in the Human immunoAIDS virus type 1 life cycle for variants influencing disease progression. J Virol 79: 12674–12680.
4. Welton AR, Chersel EJ, Sturkie C, Jackson AU, Hirsch GN, et al. (2005) Identification of quantitative trait loci for susceptibility to mouse adenosine virus type 1. J Virol 79: 11517–11522.
5. Bennett KE, Flick D, Fleming KH, Jochim R, Beatty BJ, et al. (2005) Quantitative trait loci that control dengue-2 virus dissemination in the mosquito Aedes aegypti. Genetics 170: 185–194.
6. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317: 944–947.
7. Antonarakis SE, Beckmann JS (2006) Mendelian disorders deserve more attention. Nat Rev Genet 7: 277–282.
8. Picard C, Casanova JL, Abel L (2006) Mendelian traits that confer predisposition or resistance to specific infections in humans.Curr Opin Immunol 18: 353–359.
9. Cam HM, de Toma C, Cazes L, Logrand MF, Morel V, et al. (2002) A human genome diversity cell line panel. Science 296: 261–262.
10. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, et al. (1990) Centre d’étude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6: 575–577.
11. Schork NJ, Gardner JP, Zhang L, Fallin D, Thiel B, et al. (2002) Genomic association/linkage of sodium lithium countertransport in CEPH pedigrees. Hypertension 40: 619–628.
12. Cheung VG, Conlin LS, Weber TM, Arcaro M, Jen KY, et al. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet 33: 422–425.
13. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430: 743–747.
14. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, et al. (2004) Genetic inheritance of gene expression in human cell lines. Am J Hum Genet 75: 1094–1105.
15. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, et al. (2005) Mapping determinants of human gene expression by regional and genome-wide association. Nature 437: 1365–1369.
16. Jen KY, Cheung VG (2005) Transcriptional response of lymphoblastoid cells to ionizing radiation. Genome Res 15: 2092–2100.
17. Watters JW, Kraja AM, Meucci MA, Province MA, McLeod HL (2004) Genome-wide discovery of loci influencing chemotherapy cytotoxicity. Proc Natl Acad Sci U S A 101: 11809–11814.
18. Stranger BE, Forrest MS, Dunham M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315: 848–853.
19. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, et al. (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1: e78. doi:10.1371/journal.pgen.0010167.
37. O’Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63: 259–266.
38. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30: 97–101.
39. Diggle PJ (2002) Analysis of longitudinal data. Oxford: Oxford University Press.
40. Zheng B (2000) Summarizing the goodness of fit of generalized linear models for longitudinal data. Stat Med 19: 1265–1275.
41. Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42: 121–130.
42. Munoz A, Carey V, Taylor JM, Chmiel JS, Kingsley L, et al. (1992) Estimation of time since exposure for a prevalent cohort. Stat Med 11: 939–952.
43. Geskus RB (2001) Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. Stat Med 20: 795–812.
44. Arnedo M, Taffé P, Sahli R, Furrer H, Hirschel B, et al. (2007) Evaluation of the contribution of 20 variants of 13 genes to dyslipidemia associated with antiretroviral therapy. Pharmacogenet Genomics 17: 755–764.
45. Brass AL, Dykshoorn DM, Benita Y, Yan N, Engelman A, et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. Science, E-pub ahead of print. PMID: 1818762.