

Genome-Wide Association Study Implicates *PARD3B*-Based AIDS Restriction

Jennifer L. Troyer,¹ George W. Nelson,¹ James A. Lautenberger,² Leslie Chinn,² Carl McIntosh,² Randall C. Johnson,¹ Efe Sezgin,² Bailey Kessing,¹ Michael Malasky,¹ Sher L. Hendrickson,² Guan Li,¹ Joan Pontius,¹ Minzhong Tang,² Ping An,¹ Cheryl A. Winkler,¹ Sophie Limou,³ Sigrid Le Clerc,³ Olivier Delaneau,³ Jean-François Zagury,³ Hanneke Schuitemaker,⁴ Daniëlle van Manen,⁴ Jay H. Bream,⁵ Edward D. Gomperts,⁶ Susan Buchbinder,⁷ James J. Goedert,⁸ Gregory D. Kirk,⁹ and Stephen J. O'Brien²

¹Laboratory of Genomic Diversity, SAIC-Frederick, Inc., and ²Laboratory of Genomic Diversity, National Cancer Institute, Frederick; Departments of ³Molecular Microbiology and Immunology and ⁴Epidemiology, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, ⁵Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; ⁶Department of Hematology, Children's Hospital Los Angeles; ⁷San Francisco Department of Public Health, HIV Research Section, California; ⁸Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France; and ⁹Department of Experimental Immunology, Sanquin Research, Landsteiner Laboratory and Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center, University of Amsterdam, the Netherlands

Background. Host genetic variation influences human immunodeficiency virus (HIV) infection and progression to AIDS. Here we used clinically well-characterized subjects from 5 pretreatment HIV/AIDS cohorts for a genome-wide association study to identify gene associations with rate of AIDS progression.

Methods. European American HIV seroconverters ($n = 755$) were interrogated for single-nucleotide polymorphisms (SNPs) ($n = 700,022$) associated with progression to AIDS 1987 (Cox proportional hazards regression analysis, co-dominant model).

Results. Association with slower progression was observed for SNPs in the gene *PARD3B*. One of these, rs11884476, reached genome-wide significance (relative hazard = 0.3; $P = 3.370 \times 10^{-9}$) after statistical correction for 700,022 SNPs and contributes 4.52% of the overall variance in AIDS progression in this study. Nine of the top-ranked SNPs define a *PARD3B* haplotype that also displays significant association with progression to AIDS (hazard ratio, 0.3; $P = 3.220 \times 10^{-8}$). One of these SNPs, rs10185378, is a predicted exonic splicing enhancer; significant alteration in the expression profile of *PARD3B* splicing transcripts was observed in B cell lines with alternate rs10185378 genotypes. This SNP was typed in European cohorts of rapid progressors and was found to be protective for AIDS 1993 definition (odds ratio, 0.43, $P = .025$).

Conclusions. These observations suggest a potential unsuspected pathway of host genetic influence on the dynamics of AIDS progression.

AIDS has taken over 30 million lives in the 30 years since it first emerged as a cluster of rare infections and cancers among men who have sex with men (MSM) and recipients of human immunodeficiency virus (HIV)-contaminated blood products in the early 1980s. HIV

has evaded most public health prevention campaigns, and efforts to develop a vaccine have been largely unsuccessful. Highly active antiretroviral therapy (HAART) has been developed, which delays progressive immune deficiency and improves survival. However, HAART does not clear HIV infection, can be difficult to tolerate, and occasionally elicits serious adverse reactions [1]. Genetic association studies of candidate genes have revealed several common genetic variants that modulate AIDS progression, including *CCR5-Δ32*, *HLA-homozygosity*, *HLA-B27* and *-B57*, *IL10-5'A*, *MYH9*, and *KIR* variants [2–6]. One of these AIDS Restriction Genes (*CCR5-Δ32*) encouraged the pharmaceutical development of HIV entry inhibitors, a new class of anti-HIV drugs [7–9] and, in one case, was

Received 30 September 2010; accepted 22 December 2010.

Potential conflicts of interest: none reported.

Presented in part: Conference on Retroviruses and Opportunistic Infections, San Francisco, California, 15–19 February 2010. Poster board #458.

Correspondence: Jennifer L. Troyer, MD, Laboratory of Genomic Diversity, Bldg 560, Rm 21-56, NCI-Frederick, Frederick, MD 21702 (troyerj@mail.nih.gov).

The Journal of Infectious Diseases 2011;203:1491–502

Published by Oxford University Press on behalf of the Infectious Diseases Society of America 2011.

0022-1899 (print)/1537-6613 (online)/2011/20310-0018\$14.00

DOI: 10.1093/infdis/jir046

employed in a successful bone marrow transplantation to a patient with AIDS who remained clear of HIV-1 for >2 years [10].

There are specific challenges to genome-wide association studies with susceptibility and resistance to infectious diseases generally and HIV/AIDS associations in particular. Because genetic variation is not causative, but rather modulates the host's response to an infectious agent, effects are expected to be moderate. In addition, operative protective or susceptible alleles may be rare in a population. Finally, few clinically well-defined pretreatment HIV/AIDS cohorts exist, and clinical outcomes and progression measures are varied. Therefore, power for any one test is limited. Nevertheless, a number of genome-wide association studies (GWAS) have been applied with some success to AIDS cohorts [11–15]. Genome-wide significant findings ($P < 5 \times 10^{-8}$) have emphasized the importance of the human leukocyte antigen (HLA) region in viral load (VL) set point and progression. There have also been a few intriguing suggestions of new associations with AIDS progression that approach but do not achieve a strict Bonferroni correction for significance (with P values in the range of 10^{-5} to 10^{-7}), replicate in additional cohorts, and have as-yet-unexplored biological or functional links to AIDS. Most of these studies have been limited to small numbers of patients (ranging from <50 to a few hundred) with extreme and well-defined categorical progression phenotypes (for instance, rapid and slow progressors [14–18]). To target as-yet-undiscovered associations, we have interrogated a different and larger set of SNPs (those present on the Affymetrix 6.0 genotyping array) and have taken advantage of the detailed clinical data from 6 prospective American AIDS cohorts, previously used for discovery of 32 candidate gene associations with HIV acquisition, AIDS progression, AIDS defining condition, or HAART efficacy [3, 19]. Subjects (a total of 1526 European American [EA] and 437 African American [AA] study participants) were genotyped for 934,968 SNP variants represented on the Affymetrix 6.0 genotyping array; after quality control (QC) filters that screen for SNP veracity and utility, 700,022 SNPs were analyzed. Here, we present association with progression to AIDS 1987 (Centers for Disease Control and Prevention [CDC] definition[20]), a complex but clinically relevant phenotype, in all of the successfully genotyped EA seroconverters (755 total subjects).

After correction for multiple SNPs, we found a statistically significant signal for a single SNP within the *PARD3B* gene region on Chromosome 2. Additional linked SNPs in the region approached genome-wide significance (P values of 10^{-5} to 10^{-8}) and define a protective haplotype (odds ratio [OR], 0.3; $P = 3.2 \times 10^{-8}$). One of these SNPs was assessed for functional consequence by quantification of *PARD3B* transcripts (spliced and unspliced variants) in lymphoblastoid cell lines with alternative *PARD3B* SNP genotypes and was interrogated in additional cohorts for association with AIDS (1993 definition [21]).

METHODS

Subjects

EA and AA subjects from 6 cohorts—Multicenter AIDS Cohort Study (MACS, principal investigator [PI] J. Phair), Multicenter Hemophilia Cohort Study (MHCS, PI J. Goedert), D.C. Gays (DCG, PI J. Goedert), Hemophilia Growth and Development Study (HGDS, PI E. Gomperts), San Francisco City Clinic Study (SFCC, PI S. Buchbinder), and AIDS linked to Intravenous Experience (ALIVE, PI G. Kirk)—were selected for GWAS on the basis of the clinical relevance and availability of high-quality high-molecular weight DNA. Seroconverters, defined as subjects who entered their respective studies with a seronegative HIV status and had an interval of ≤ 3 years between their last negative and first positive HIV test result, were available from 5 cohorts (EA: 405 from MACS, 222 from MHCS, 40 from DCG, 75 from SFCC, and 13 from ALIVE; AA: 39 from MACS, 9 from MHCS, 2 from DCG, 6 from SFCC, and 226 from ALIVE); seroconversion date was set as the midpoint between last negative and first positive HIV test result.

Genotyping

DNA was extracted from lymphoblast cell lines[22] and run on the Affymetrix genome-wide human SNP array 6.0 genotyping platform. DNA (325 ng/sample) was prepared for both Sty1 and Nsp1 restriction enzyme digestion for this assay, an increase of 125 ng over Affymetrix recommendations, which greatly increased our success rate. The remainder of the assay was performed as per manufacturer's instructions.

Sample QC

First-pass QC of samples was performed in Genotyping Console 3.0.2. All samples that passed contrast QC (CQC) ($CQC > .4$) were genotyped for 934,968 SNPs using Affymetrix Power Tools (apt-probeset-genotype 1.10.0). Samples were genotyped by batch (200–500 samples), and individuals with <90% call rates were excluded from further analysis. Four controls were run a total of 39 times, and 145 duplicate samples and 8 Centre d'Etude du Polymorphisme Humain (CEPH) parent-offspring trios were run within and between plates and verified for concordance across runs (concordances $\geq 99\%$). Once acceptable genotyping standards were met, Affymetrix genotypes were compared with previously produced Taqman, Illumina, and/or Perlegen genotypes at a mean of 85 (range, 10–142) sites per sample. Samples with <90% concordance with their previous genotypes were discarded. Affymetrix sex calls (based on Y chromosome markers, X chromosome heterozygosity, and the intensity of the sex chromosome invariant probes) were compared with clinical data files, and inconsistent individuals were discarded.

PLINK[23] was used to assess cryptic identity and cryptic relatedness and to verify sex calls. Identical subjects were investigated and discarded (if clinical data did not match) or were

Table 1. Explained Fraction of AIDS Restriction Genes (ARGs) for Progression to AIDS in 755 European American Seroconverter Subjects

Locus	Variant	Model	Frequency, %	<i>P</i>	RH	EF (%)
<i>CCR5</i>	Δ32	D	20.1	.038	0.744	0.56
<i>CCR2</i>	64I	D	18.0	.025	0.723	0.6
<i>SDF1</i>	3'A	R	3.4	.061	0.488	0.5
<i>CCR5</i>	+/P1/+ haplotype	R	12.4	.001	1.667	1.18
<i>IL10</i>	5'A	D	41.9	.170	1.165	0.2
<i>HLA B</i>	B35Px	D	8.4	<.001	2.365	2.63
<i>HLA B</i>	B27	D	9.8	.002	0.513	1.65
<i>HLA B</i>	B57	D	8.5	.033	0.559	0.86
<i>HLA A, B, C</i>	homozygosity at 1 gene	-	22.6	.664	1.067	0.03
<i>HLA A, B, C</i>	homozygosity at 2 or 3 genes	-	6.5	.001	2.019	1.02
<i>RANTES</i>	Haplotype 3 or 5	D	5.2	.004	1.813	0.84
<i>KIR</i>	3DS1	C	21.8	>.990	1	0
<i>KIR3DS1-HLA Bw480I</i>	epistasis	-	-	.013	0.623	1.01
<i>PAR3B</i>	G at rs11884476	C	6.3	3.370×10^{-9}	0.29	4.52
Combined ARGs						15.6

NOTE. RH, relative hazard.

reduced to a single genotype file (if clinical data, including birth date, matched). Related subjects were flagged.

SNP QC

Individuals that passed subject QC were re-clustered and interrogated for all 934,968 SNPs as a single batch. SNPs were then filtered on the basis of multiple criteria. Autosomal perfect-match SNPs supported by Affymetrix ($n = 868,157$) were evaluated for Mendelian inheritance in CEPH trios (389 dropped), SNP call rates of >95% (39,220 dropped), Hardy Weinberg equilibrium (HWE; $P > .001$; 6512 dropped), and minor allele frequency >0.01 (121,403 dropped), resulting in a total of 700,022 SNPs for association analyses in EAs.

Population Structure

Population structure was assessed using the Principal Components Analysis (PCA) module of *Eigensoft* software [24]. EA and AA subjects segregated with no discrepancies between clinical and genetic assignment. Within EA subjects, the top 10 most significant eigenvectors were calculated, and the top 2 were plotted (data not shown). The distribution of EA study subjects along the top 2 eigenvectors demonstrated a topology that mirrored that seen in several other studies of the European population structure [25], including a cline that probably reflects Northern European to Southern European descent and a cluster that most likely represents Ashkenazi Jews in our cohorts. AIDS outcomes were not significantly associated with population substructure. We also corrected for population structure with the genomic control method [26]. The λ value was 1.01, and a λ correction did not alter the association results significantly. Both the PCA and genomic control analyses suggested that population stratification was not a confounding factor in our HIV progression analyses.

All SNPs that contributed to population structure ($P < 10^{-7}$) were flagged in subsequent tests (displayed as an "X" in the Manhattan plot; Figure 1B). For all SNPs that indicated a potential association, the first 2 eigenvectors were used as covariates in the applicable tests, and *P* values were adjusted accordingly. None of the SNPs presented in this manuscript demonstrated significant contribution to population structure, and correction for structure did not alter the *P* values significantly.

Association Analysis

Cox model analysis was performed for all SNPs passing the QC filters, using an additive model taking the number of copies of the rare allele, 0, 1 or 2, as the explanatory variable. The analysis used PROC PHREG in Statistical Analysis Software (SAS) [27]; the significance was evaluated using the log likelihood test. We performed a strict Bonferroni correction for the 700,022 SNPs used in the GWAS [28–30]. All SNPs identified as interesting by statistical analysis were examined visually for quality assurance. SNP rs11884476 was further investigated to calculate explained fraction (EF) [3, 31, 32] using the SAS-based program, SUREV [27]. Haplotypes were inferred by implementation of an expectation-maximization algorithm.

Both rapid progressors (AIDS, by the CDC 1993 definition, within 3 years of seroconversion) and slow progressors (CD4+ cell count >500 cells/mL and no clinical symptoms of AIDS at least 8 years after seroconversion) from the French GRIV cohort were typed using the Illumina HumanHap300 genotyping beadchips [14, 15], as were rapid progressors and controls in the Amsterdam Cohort Study (ACS [33]). 376 Dutch and 695 French general population seronegative controls were typed on the same platform. One SNP in the *PAR3B* region of

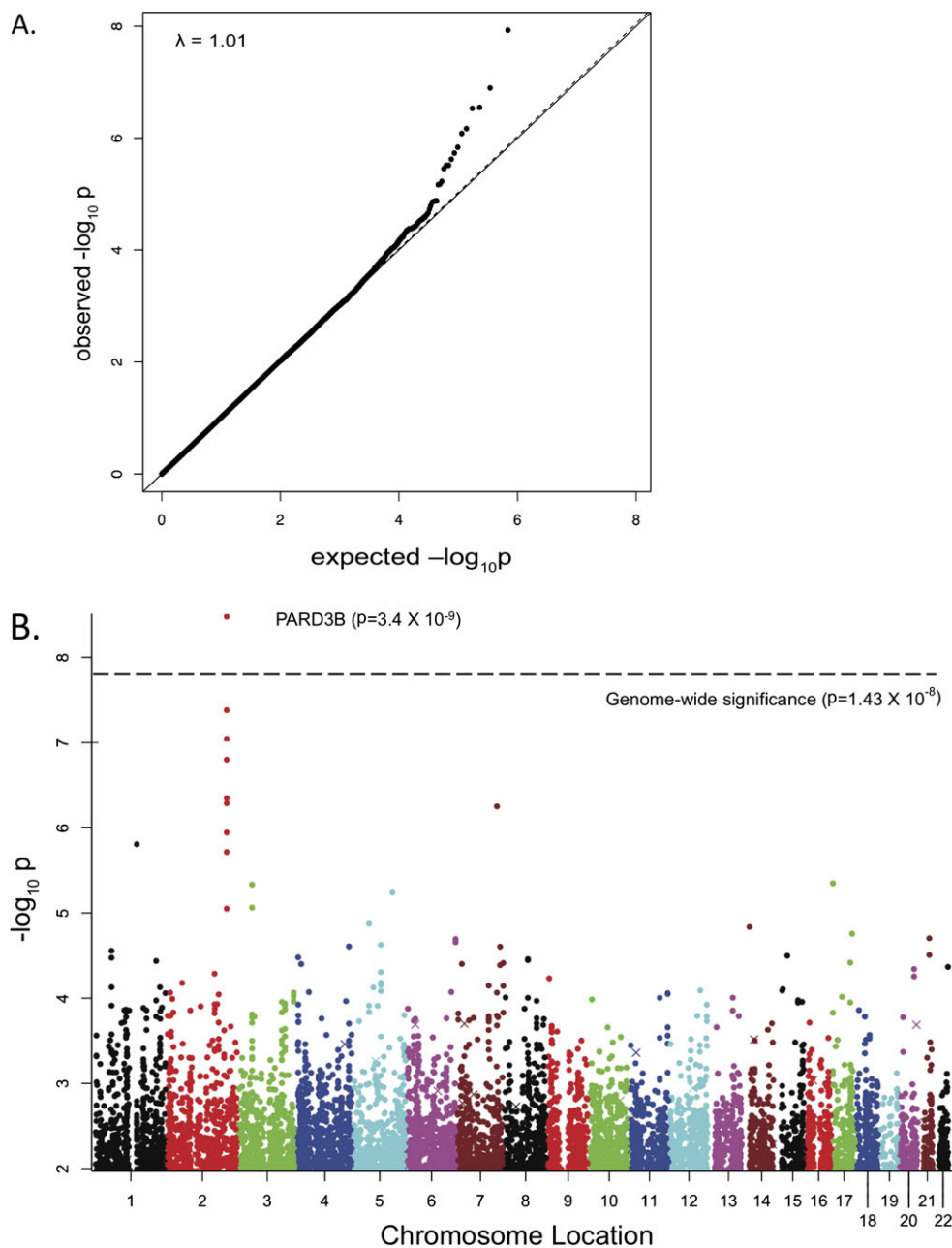


Figure 1. *A*, Quantile-quantile plot for the expected versus observed P values for this study. λ is the slope of the first 95% of the data. *B*, Manhattan plot showing the distribution of $-\log_{10} (P$ values) for time to AIDS87 (co-dominant model) along the autosomes for 700,022 single-nucleotide polymorphisms (SNPs). Each chromosome is represented by a different color and numbered sequentially along the x-axis. SNPs are ordered by chromosomal position. Values for SNPs that contributed significantly to population sub-structure are designated with an X rather than a dot. The horizontal black line marks the level of genome-wide significance at the .01 significance level ($P = 1.370 \times 10^{-8}$).

significance was the same on both Illumina and Affymetrix platforms (rs10185378) and was tested as a candidate SNP in a categorical case-control analysis in a combined set of GRIV and ACS rapid progressors ($n = 123$) versus HIV-negative control subjects of similar ethnicity ($n = 1071$). Eigenstrat, implemented in Eigensoft, was used to identify outliers in the population who were subsequently removed from the analysis. The first 2 eigenvectors were used as covariates to

correct for any remaining population structure. An attempt was made to impute the other 8 SNPs presented here using IMPUTE 2 [34]. However, these SNPs could not be imputed with high confidence in all of the data; 8%–26% of the samples had <90% confidence for genotype calls, with most SNPs missing ~20% of the calls. Minor alleles were absent from most of the imputed data, so statistics could not be reliably performed for these SNPs.

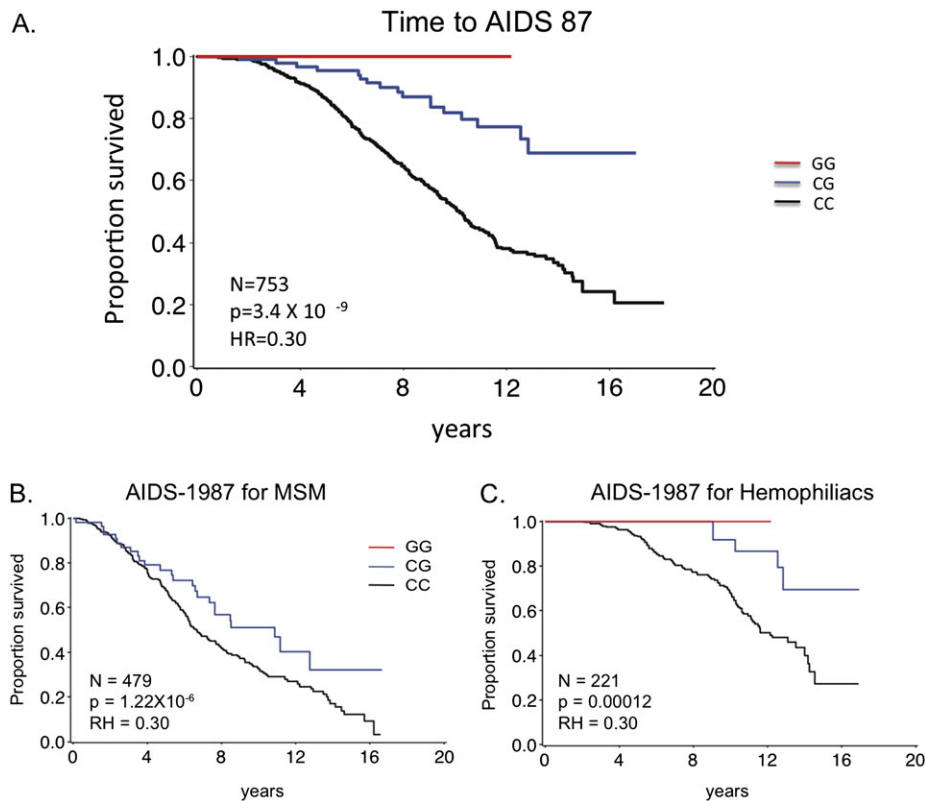


Figure 2. Kaplan-Meier survival curve, demonstrating the co-dominant effect of rs11884476 genotype on time to AIDS87 for all 755 seroconverters (A), 479 men who have sex with men (MSM; B), and 221 hemophiliacs (C).

RNA Isolation and Quantification of *PARD3B* Expression

RNA was extracted from lymphoblastoid cell lines of individuals with alternative *PARD3B* genotypes using the RNeasy Mini kit (Qiagen) and quantified using a NanoDrop spectrophotometer (NanoDrop Technologies). First-strand complementary DNA synthesis was performed using SuperScript III (Life Technologies), according to the manufacturer's instructions. Gene expression was measured by quantitative reverse-transcription polymerase chain reaction (PCR) on a LightCycler 480 (Roche Applied Science). Primers were designed to span the junction between *PARD3B* exons 19 and 20 (isoforms B and C) or exons 19 and 21 (isoform A) and included 1 forward primer (5'-AAACATGGTGGCCTGAGAGA-3') and 2 reverse isoform-specific primers (isoform A: 5'-CGTAGGACGGCCACTTTC-3', isoforms B and C: 5'-TGATGTTTTGCTCCAATCCTT-3'). The FAM-labeled probe used to detect both *PARD3B* isoforms was Universal ProbeLibrary probe 78 (Roche Applied Science). Quantification of *PARD3B* expression was performed in multiplex reactions with the human β -actin reference gene assay (Roche Applied Science). Each reaction contained 1 × LightCycler Master, 100 nM each probe, and 200 nM each primer, in a final volume of 10 μ L. Reactions were performed in triplicate. PCR conditions were 5 min at 95°C, followed by 45 cycles of 10 s at 95°C, 20 s at 60°C, and 1 s at 72°C. *PARD3B* isoform

expression was normalized to β -actin expression for each reaction, and samples were excluded if the coefficient of variation was $>.04$ for the triplicate. The relative effect of genotype at rs10185378 on differential splicing was tested by linear regression using R, version 2.11.1, functions `lm`, and `estimable` (in library `gmodels`).

RESULTS

We present data from 700,022 SNPs that passed QC filters (see Methods) in the 755 EA seroconverters from 5 HIV/AIDS natural history cohorts. Of 803 EA seroconverters from 5 cohorts, 24 failed Affymetrix QC or the initial call rate cutoff of 90%, 5 were excluded for inconsistencies with their clinical file, and 15 were excluded due to sample mix-up (cryptic identity or discordance with previous genotypes). Once these 759 subjects were re-clustered and recalled; all but 4 had call rates of $>95\%$, with an average call rate of 98.8%. When those 4 subjects and SNPs failing filters were removed from the analysis, leaving the 755 seroconverters and 700,022 SNPs considered here, average call rate increased to 99.3%.

A test for progression to the 1987 definition of AIDS[20] using Cox model survival analysis revealed departure of multiple SNPs from expectation (Figure 1A and 1B) including 1 SNP,

rs11884476 on chromosome 2, position 206,026,838, that reached genome-wide significance ($P = 3.370 \times 10^{-9}$; Figure 1B). The minor allele was low frequency in our population (0.063) and protective, with a relative hazard of 0.3 (Figure 2A). When the study cohort was subdivided on the basis of risk group, MSM and hemophiliacs displayed parallel significant associations (for MSM, relative hazard (RH) = 0.43 [$P = .0024$]; for hemophiliacs, RH = 0.3 [$P = .0012$]; Figures 2B and 2C).

The protective SNP is located near exon 20 of the *PARD3B* gene region, which is a homolog of a cell polarity gene identified in *Caenorhobditis elegans* that is associated with tight junctions in *C. elegans* and human cells[35]. To quantify the influence of the *PARD3B* genotype on AIDS progression, we calculated the EF of progression to AIDS 1987[3, 31, 32] for *PARD3B* rs11884476 along with 13 genetic factors previously identified as affecting AIDS progression (Table 1). When compared with the 13 previously described AIDS restriction genes re-assessed in this group of subjects, *PARD3B* factor rs11884476 had the largest EF, 4.52%. The overall EF for all 14 factors considered together was 15.6%.

Eight additional SNPs clustered around the exon 20 region of *PARD3B* were among the top 10 hits, with $P \leq 1.9 \times 10^{-6}$ and a protective effect (Table 2; Figures 1B and 3A). Additionally, a total of 14 *PARD3B* SNPs fell within the lowest 1% of GWAS *P* values (data not shown). The 9 highest-ranking *PARD3B* SNPs,

including the top 5 hits overall, were all included in a single protective haplotype within *PARD3B* ($P = 3.2 \times 10^{-8}$; hazard ratio [HR], 0.29; Figure 3A). These signals did not replicate in a cohort of 282 AA seroconverters.

Only 1 SNP (rs10185378) in the protective linkage group was in the coding region: a missense (C to T) in exon 20 that results in a change from a Threonine to an Isoleucine. This SNP has also been genotyped in 2 well-defined cohorts of rapid progressors; from the GRIV cohort [14, 15] and the Amsterdam Cohort Study [33] ($n = 84$ and $n = 39$, respectively). We examined the effect of this SNP in an analysis of rapid progressors compared with uninfected controls from both cohorts and found that it had a significant protective influence ($P = .025$; odds ratio [OR], 0.47; co-dominant model), consistent with our findings (Figure 3). In all cohorts (GRIV, ACS, and the discovery NCI - LGD combined cohorts), not a single rapid progressor was homozygous for the minor allele, and the overall frequency of this allele within rapid progressors was half that of the general population (data not shown). Although this amino acid change is not predicted to have a profound effect on protein function, this SNP is within a predicted splice site enhancer for exon 20 [36].

To explore this potential functional basis of *PARD3B*-based association, we developed a real-time PCR assay to detect expression of alternative splice forms in lymphoblastoid cell lines (LCLs) with alternative *PARD3B* genotypes for the codon-altering SNP, rs10185378 [37] (Figure 4A). *PARD3B* produces 3

Table 2. Twenty Lowest *P* Values for Survival Analysis to AIDS-1987, Codominant Model

Rank	rs number	chr	Position	Gene	MAF	<i>P</i>	HR ^b
1	rs11884476	2	206,026,838	PARD3B	.0631	3.370E-09	0.301
2	rs11896375	2	206,015,793	PARD3B	.0642	4.174E-08	0.347
3	rs10185378	2	206,013,401	PARD3B	.0609	9.234E-08	0.349
4	rs10210768	2	205,998,621	PARD3B	.0689	1.585E-07	0.387
5	rs11886997	2	206,039,636	PARD3B	.0663	4.496E-07	0.388
6	rs13410658	2	206,050,327	PARD3B	.0589	5.143E-07	0.362
7	rs6467710	7	136,854,359	DGKI	.0159	5.596E-07	4.302
8	rs880343	2	206,023,936	PARD3B	.0675	1.134E-06	0.416
9	rs11239930	1	145,027,188	NBPF14	.4184	1.556E-06	0.681
10	rs7565154	2	206,034,837	PARD3B	.0742	1.923E-06	0.443
11	rs7217319	17	38,924	RPH3AL	.3714	4.482E-06	0.677
12	rs6441975	3	46,403,194	CCRL2	.3138	4.643E-06	1.483
13	rs477687	5	134,495,599	H2AFY	.1807	5.744E-06	1.548
14	rs1015164	3	46,426,684	LTF	.3088	8.636E-06	1.463
15	rs10932110	2	206,050,138	PARD3B	.0622	8.868E-06	0.429
16	rs1553477	5	53,152,723	ARL15	.4651	1.335E-05	0.699
17	rs17112431	14	20,027,330	RNASE10	.2078	1.456E-05	0.644
18	rs17225585	17	66,882,025	SOX9	.1245	1.750E-05	0.568
19	rs7275915	21	34,753,498	KCNE1	.1851	1.979E-05	0.627
20	rs4397202	6	168,718,725	SMOC2	.0471	2.024E-05	2.092

NOTE. Rank of *P* values for 700,022 SNPs tested. chr, chromosome; MAF, minor allele frequency; HR, hazard ratio; Chromosome assignment, base-pair positions, and gene associations of single-nucleotide polymorphisms (SNPs) presented here are from Affymetrix SNP Annotation, version 13.1, which is based on National Center for Biotechnology Information human genome build 36.1.

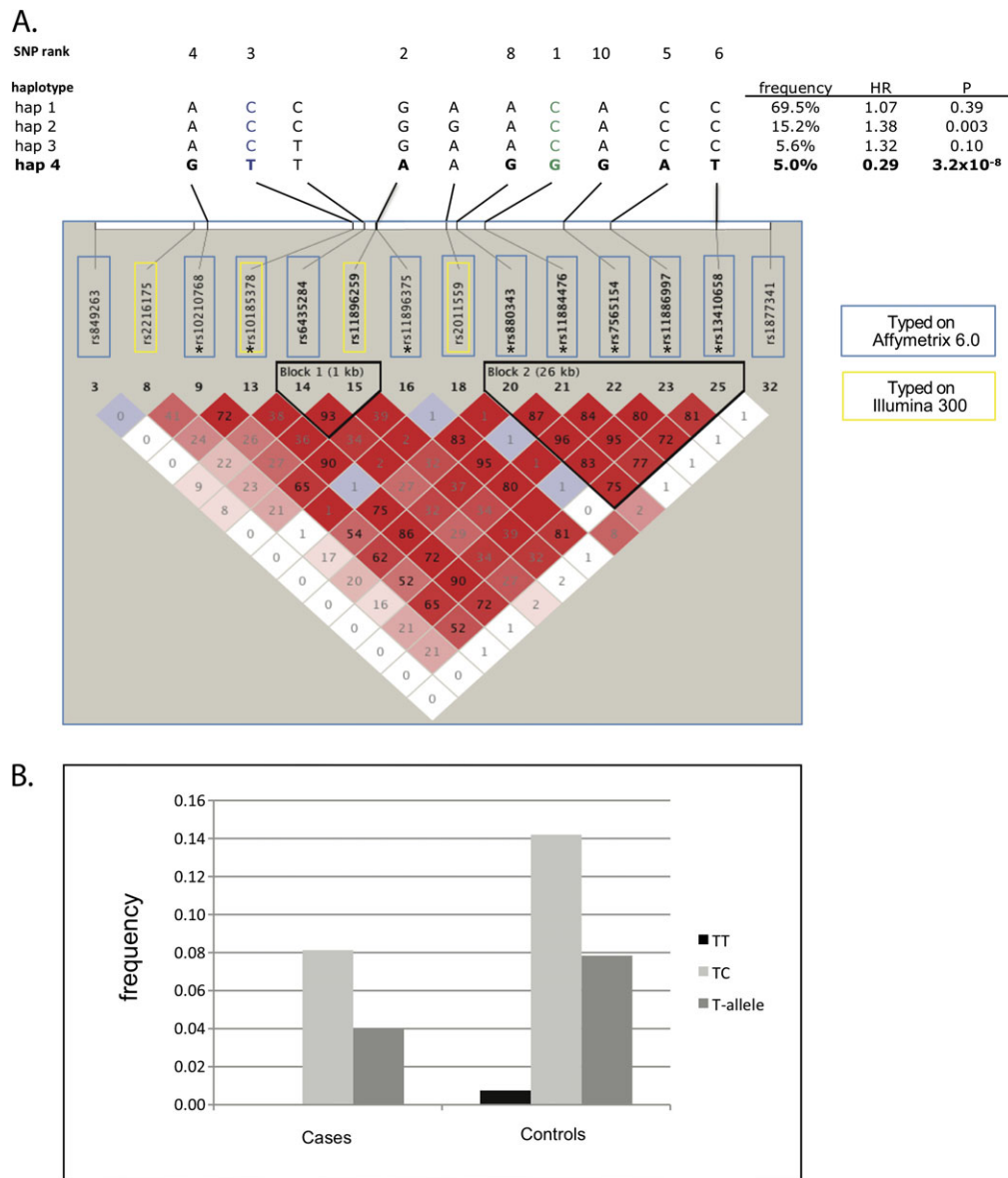


Figure 3. Analysis of *PARD3B* single-nucleotide polymorphisms (SNPs). *A*, Haplotype and LD structure of the SNPs for the *PARD3B* region. SNPs outlined in blue were typed on Affymetrix 6.0 for European American samples and those highlighted with yellow were typed on Illumina300 in the GRIV and Amsterdam Cohort Study (ACS) European cohorts. Only HapMap SNPs are included here; underlying LD structure is from HapMap. SNPs with an asterisk were among the top ranking SNPs in this study. All haplotypes with a frequency >1% in the discovery cohorts are shown. SNPs that rank in the top 10 of the AIDS 93 progression analysis track a single haplotype, hap 4 (tracking SNPs are highlighted in bold), that is itself significant at the genome-wide level. Alleles from the top hit in this study (rs11884476) are shown in green; alleles from the ranking SNP typed on both platforms (rs10185378) are shown in blue. The three other SNPs in this region that are typed in the European cohorts (yellow) do not track the significant haplotype. *B*, Distribution of rs10185378 genotypes in the combined GRIV and ACS cohorts. Genotypes and allele frequencies for in control subjects ($n = 1071$) and case patients (rapid progressors; $n = 231$) CT (light gray), and TT (black) genotypes and T allele frequency (dark gray), where T is the minor allele with an overall frequency of .075, a frequency in controls of .078, and a frequency of .041 among rapid progressors.

alternatively spliced reference sequences; 2 of these, isoforms B and C, have an exon (exon 20) that is absent in isoform A [38] (Figure 4B). When we compared RNA expression of these isoforms in LCLs bearing different *PARD3B* genotypes (CC, $n = 12$; CT, $n = 8$; TT, $n = 9$), we observed an increased average expression of isoform A and a decreased expression of

isoforms B and C in cell lines homozygous (TT) for the minor (protective) allele (Figure 4B). There is a significant shift in relative expression of the 2 splice forms ($P = .006$; co-dominant model; Figure 4C). These data lend in vitro support to the in silico prediction that this enhancer influences *PARD3B* splicing [36].

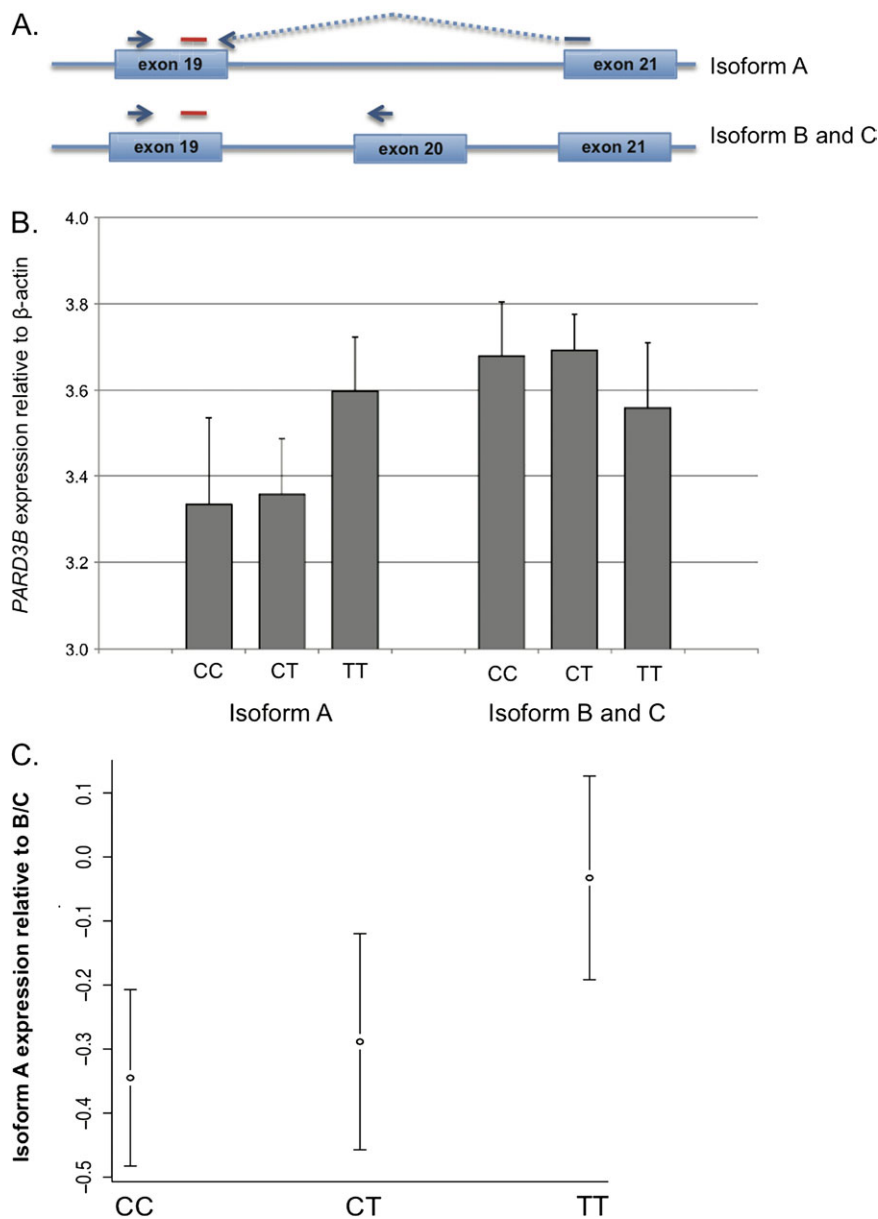


Figure 4. A, Primer probe design for the expression of alternative splice forms. Introns are represented as thin blue lines, and exons are blue rectangles with the exon number designated in black. Probes are represented as red lines, and primers are blue with arrows. B, Expression of alternative *PARD3B* splice forms by rs10185378 genotype. The arithmetic means of *PARD3B* normalized transcript expression grouped by genotype (CC, $n = 12$; CT, $n = 8$; TT, $n = 9$) are shown. Error bars indicate the standard error of the mean. Differences between expression profiles of the alternate splice forms were consistent with the predicted effect for this SNP. C, Relative difference between *PARD3B* expression of isoform A and isoforms B/C with 95% confidence intervals plotted by genotype ($P = .006$, co-dominant model).

DISCUSSION

The advance from candidate gene studies to genome-wide association studies for AIDS, as for other diseases, has the notable advantage of being able to reveal previously unsuspected genetic influences. The GWAS results presented here implicate rare genetic variants in the *PARD3B* gene as conferring slower progression to clinical AIDS among HIV-infected EAs.

PARD3B is a homolog of a *C. elegans* cell polarity determinant that localizes to tight junctions in human epithelial cells [35]. It may be relevant that human *PARD3B* gene products interact with several members of the *SMAD* (mothers against decapentaplegic homolog) family. *SMAD* proteins are signal transducers and transcriptional modulators that mediate multiple signaling pathways and are involved in a range of biological activities, including cell growth, apoptosis, morphogenesis,

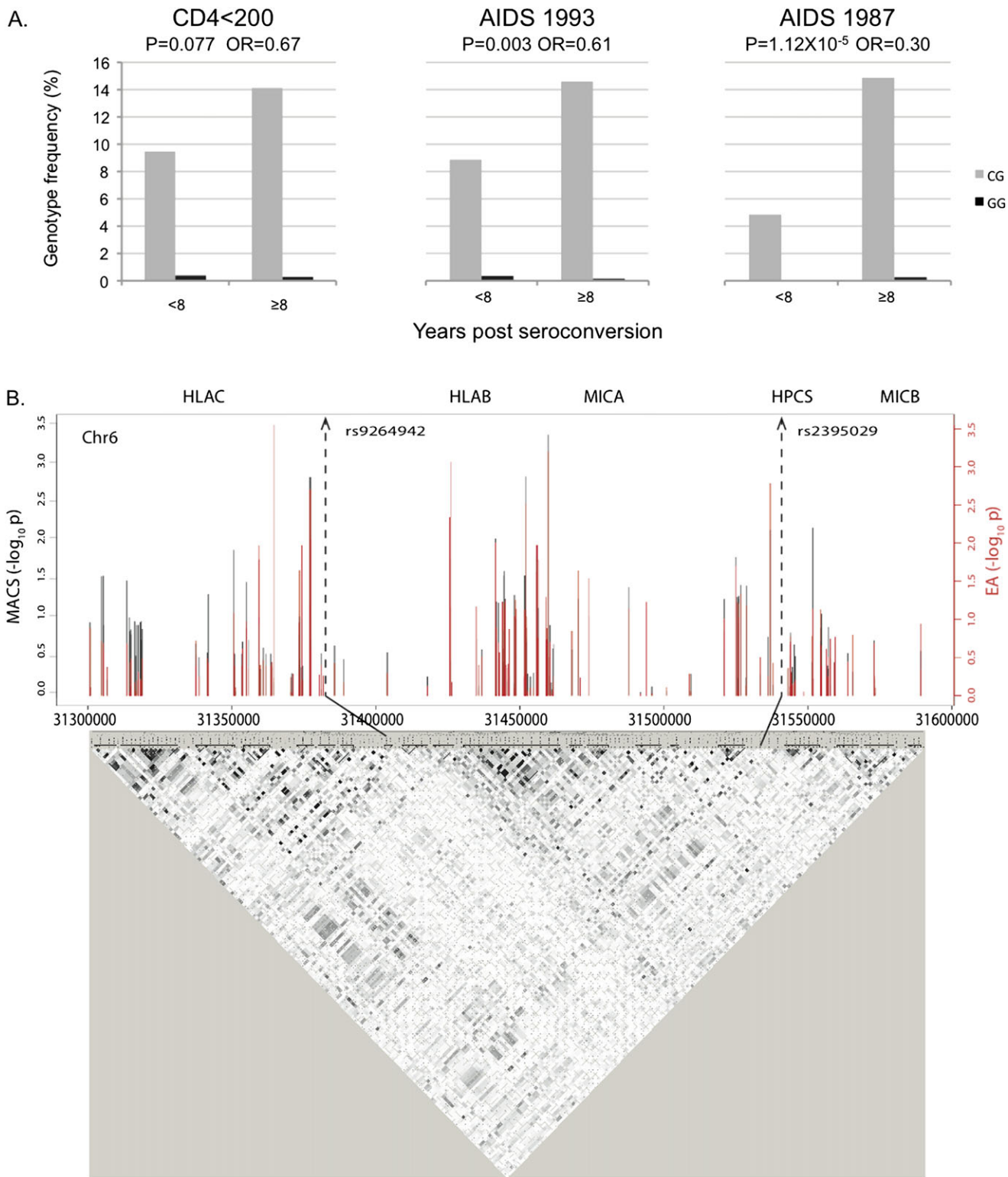


Figure 5. Additional analysis of the GWAS data. *A*, Categorical analysis of rs11884476 for different clinical outcomes. Shown here are CD4+ cell count <200 cells/mL and AIDS 1993 definition in <8 years or ≥8 years and AIDS 1987 with a cutoff value of 10 years. *B*, Results of survival analysis (AIDS-1987) for the HLA region on chromosome 6. Negative log *P* values are graphed with respect to chromosomal position for MACS seroconvertors alone (*black*) or all European American seroconvertors (*red*). Categorical tests on several cohorts, including MACS samples overlapping those in this study, have identified single-nucleotide polymorphisms (SNPs) on chromosome 6 that are highly associated with viral load set point and, to a lesser extent, progression to CD4+ cell count <400 cells/mL and rapid disease progression. The SNPs with robust association (rs9264942 and rs2395029) are indicated with dotted arrows; neither of these are typed or tracked on the Affymetrix 6.0 platform. No other SNPs in this region reached genome-wide

development, and immune responses. Of particular interest are the interactions of *PARD3B* encoded proteins with *SMAD2*, *SMAD3*, *SMAD4*, and *SMAD7*, all of which interact directly with HIV-1 [39–43]. *SMAD3* and 4 also interact with TGF β to regulate Tat-mediated transcription of HIV-1 LTR. Individually, *SMAD3* increases viral promoter activity, whereas *SMAD4* decreases it. *SMAD4* also suppresses activation by *SMAD3* [41]. These proteins have similar effects on Tat-mediated transcription of *CCL2* (formerly MCP-1; [40]). *SMAD2* and 7 are involved in HIV-1 glycoprotein 120-induced tubular epithelial cell apoptosis [43]. *SMAD7* helps prevent apoptosis, whereas *SMAD2* is involved in downstream signaling [43].

PARD3B is a large gene (with 23 exons coding for 1205 amino acids and spanning 1.07 Mbp) located on chromosome 2. One of the exons (exon 20) is alternately spliced. Three splice variants of the *PARD3B* product have been characterized and shown to have differences in their protein binding properties [44]. The functional significance of these alternative forms is unknown, but they may also have different expression profiles [44]. A SNP in the cluster of association, rs10185378, which introduces a missense mutation in exon 20, also eliminates a predicted splice enhancer for this exon. We tested this same SNP for transcript production using a group of 29 lymphoblastoid cell lines of alternate genotypes and observed a significant increase in messenger RNA levels of the *PARD3B* isoform lacking exon 20 relative to forms with exon 20 with the rare allele (TT) at rs10185378 (Figure 4). It is by no means clear that this is the causative polymorphism; additional studies into *PARD3B* function with respect to HIV-host interactions are indicated.

As with any association result that may produce false-positive results, replication in independent cohorts is essential. Although results here are significant for this association, several factors make independent replication particularly important. The fact that the gene identified falls in an unexpected category demands additional evidence for the association to be validated. By the same token, if confirmed, the identification of *PARD3B* would have the potential to reveal an entirely new avenue of HIV cellular activity, with potential new targets for therapy, and could lead to a new understanding of the process of AIDS pathogenesis. The corroborative association of *PARD3B*-SNP rs10185378 in the European GRIV and ACS rapid progression cohorts provides an initial replication for this

association (Figure 3B). A EA seroconverter cohort with AIDS 1987 progression would be a more valuable, though unavailable, comparison, because the signal in our cohort clearly associates with this particular phenotype and is not evident when comparing rapid and slow progressors from our cohorts for CD4+ cell count <200 cells/mL and does not achieve genome-wide significance for AIDS 1993 definition (Figure 5A). Therefore, it is not surprising that our SNP of highest significance did not replicate for CD4+ cell count <300 cells/mL or AIDS 1993 in a dichotomous analysis of the GRIV rapid versus slow progressors (data not shown). In addition, allele frequency and haplotype structure in African populations is quite different than in EA populations, which might account for the failure to replicate this signal in a small ($n = 282$) AA seroconverter replication cohort.

Although this GWAS demonstrates the strength of this approach for detecting unsuspected AIDS restriction genes, it also illustrates some weaknesses. Although we observed significant unadjusted P values for many of the previously validated ARGs [3], none achieved genome-wide significance in the present GWAS, nor did we detect additional recently reported associations from other HIV/AIDS progression GWAS[11–14] with this phenotype (Figure 5B). Deviations in SNP choice and region coverage from platform to platform account for some of the disparity between results from similar studies (as demonstrated by our difficulty in using Illumina HapMap300 SNPs to impute Affy 6.0 SNPs with high confidence in our *PARD3B* region of significance; see Methods), as do alternative choices in disease phenotype tests and genetic models. The AIDS progression end point phenotype examined here represents a proven, well-powered analysis for progression to clinical AIDS. However, it is a complex phenotype, and further exploration of *PARD3B* association with multiple sequelae is now clearly indicated. A more detailed exploration of genetic associations with all of the clinical data collected on these AIDS cohorts will ultimately allow a much broader view of multiple AIDS phenomena, albeit with statistical penalties for multiple test approaches (J. A. Lautenberger, J. L. Troyer, and S. J. O'Brien, unpublished data). As the interpretation of the full range of data necessarily involves subjective judgment, with a corresponding loss of statistical rigor, it seems useful to extract from the full dataset the statistically straightforward analysis presented here, which implicates a strong new AIDS progression genetic association.

significance for association with time to AIDS 1987. Fellay et al [11] observed differences in association signals when tests for viral load set point were compared with test for CD4+ cell count depletion; AIDS 1987 is a much more complex phenotype that includes a range of clinical conditions that may have complex and conflicting interactions with genetic variants within this region. Therefore, the lack of signal in this study does not conflict with or contradict previous reports.

Funding

The Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research; the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E. 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. Funding for genetic studies on the ACS was provided by the Netherlands Organization for Scientific Research (TOP, registration number 9120.6046).

Acknowledgments

We gratefully acknowledge and assign collaborative credit investigators who have assembled the AIDS study cohorts: Multicenter AIDS Cohort Study (MACS, PI J. Bream), Multicenter Hemophilia Cohort Study (MHCS, PI J. Goedert), D.C. Gays (DCG, PI J. Goedert), Hemophilia Growth and Development Study (HGDS, PI E. Gomperts), San Francisco City Clinic Study (SFCC, PI S. Buchbinder) and AIDS linked to Intravenous Experience (ALIVE, PI G. Kirk). We also acknowledge the excellent work of Michelle Hall, Mary McNally, Lisa Maslan, David Wells, and Jamie Troxler, who worked tirelessly to produce the genotypes, and the LGD student interns who provided sample preparation and technical assistance. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

1. Pereira CF, Paridaen JT. Anti-HIV drug development—an overview. *Curr Pharm Des* **2004**; 10:4005–4037.
2. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med* **2003**; 54:535–551.
3. O'Brien SJ, Nelson GW. Human genes that limit AIDS. *Nat Genet* **2004**; 36:565–574.
4. Heeney JL, Dalgleish AG, Weiss RA. Origins of HIV and the evolution of resistance to AIDS. *Science* **2006**; 313:462–466.
5. Kopp JB, Smith MW, Nelson GW, et al. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* **2008**; 40:1175–1184.
6. An P, Winkler CA. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet* **2010**; 26:119–131.
7. Manfredi R, Sabbatani S. A novel antiretroviral class (fusion inhibitors) in the management of HIV infection. Present features and future perspectives of enfuvirtide (T-20). *Curr Med Chem* **2006**; 13:2369–2384.
8. Aquaro S, D'Arrigo R, Svicher V, et al. Specific mutations in HIV-1 gp41 are associated with immunological success in HIV-1-infected patients receiving enfuvirtide treatment. *J Antimicrob Chemother* **2006**; 58:714–22.
9. Westby M, van der Ryst E. CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection. *Antivir Chem Chemother* **2005**; 16:339–54.
10. Hutter G, Nowak D, Mossner M, et al. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* **2009**; 360:692–8.
11. Fellay J, Ge D, Shianna KV, et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* **2009**; 5:e1000791.
12. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* **2007**; 317:944–7.
13. Dalmaso C, Carpentier W, Meyer L, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *LoS One* **2008**; 3:e3907.

14. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* **2009**; 199:419–26.
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**; 200:1194–201.
16. Herbeck JT, Gottlieb GS, Winkler CA, et al. Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. *J Infect Dis* **2010**; 201:618–26.
17. Le Clerc S, Coulonges C, Delaneau O. Screening low frequency SNPs from genome wide association study reveals a new risk allele for progression to AIDS. *J Acquir Immune Defic Syndr*. **2010**. Epub ahead of print, 23 November.
18. Limou S, Coulonges C, Herbeck JT, et al. Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. *J Infect Dis* **2010**; 202:908–15.
19. Hutcheson HB, Lautenberger JA, Nelson GW, et al. Detecting AIDS restriction genes: from candidate genes to genome-wide association discovery. *Vaccine* **2008**; 26:2951–65.
20. Centers for Disease Control. Revision of the CDC Surveillance Case Definition for Acquired Immunodeficiency Syndrome. *MMWR Morb Mortal Wkly Rep* **1987**; 36.
21. From the Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *JAMA* **1993**; 269:729–30.
22. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **1996**; 273:1856–62.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **2007**; 81:559–75.
24. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **2006**; 38:904–9.
25. Price AL, Butler J, Patterson N, et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* **2008**; 4:e236.
26. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* **1999**; 55:997–1004.
27. Schaubel DE, Wei G. Fitting semiparametric additive hazards models using standard statistical software. *Biom J* **2007**; 49:719–30.
28. Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* **2008**; 4:e1000130.
29. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **2008**; 9:356–69.
30. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* **1996**; 273:1516–7.
31. Nelson GW, O'Brien SJ. Using mutual information to measure the impact of multiple genetic factors on AIDS. *J Acquir Immune Defic Syndr* **2006**; 42:347–54.
32. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics* **2000**; 56:249–55.
33. van Manen D, Kootstra NA, Boeser-Nunnink B, Handulle MA, van't Wout AB, Schuitemaker H. Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS* **2009**; 23:19–28.
34. Howie BN, Donnelly P, Marchini J. IMPUTE2. http://mathgen.stats.ox.ac.uk/impute/impute_v2.html. Accessed 17 February 2011.
35. Kohjima M, Noda Y, Takeya R, Saito N, Takeuchi K, Sumimoto H. PAR3beta, a novel homologue of the cell polarity protein PAR3, localizes to tight junctions. *Biochem Biophys Res Commun* **2002**; 299:641–6.

36. Yuan HY, Chiou JJ, Tseng WH, et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* **2006**; 34:W635–41.
37. National Center for Biotechnology Information. PARD3B par-3 partitioning defective 3 homolog B (*C. elegans*) [*Homo sapiens*]. <http://www.ncbi.nlm.nih.gov/gene/117583>. Accessed 17 February 2011.
38. National Center for Biotechnology Information. PARD3B par-3 partitioning defective 3 homolog B (*C. elegans*) [*Homo sapiens*]. <http://www.ncbi.nlm.nih.gov/gene/117583>. Accessed 17 February 2011.
39. Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* **2009**; 37:D417–22.
40. Abraham S, Sawaya BE, Safak M, Batuman O, Khalili K, Amini S. Regulation of MCP-1 gene transcription by Smads and HIV-1 Tat in human glial cells. *Virology* **2003**; 309:196–202.
41. Abraham S, Sweet T, Sawaya BE, Rappaport J, Khalili K, Amini S. Cooperative interaction of C/EBP beta and Tat modulates MCP-1 gene transcription in astrocytes. *J Neuroimmunol* **2005**; 160: 219–27.
42. Coyle-Rink J, Sweet T, Abraham S, et al. Interaction between TGFbeta signaling proteins and C/EBP controls basal and Tat-mediated transcription of HIV-1 LTR in astrocytes. *Virology* **2002**; 299:240–7.
43. Kapasi AA, Fan S, Singhal PC. p300 modulates HIV-1 gp120-induced apoptosis in human proximal tubular cells: associated with alteration of TGF-beta and Smad signaling. *Nephron Exp Nephrol* **2006**; 102: e30–8.
44. Gao L, Macara IG, Joberty G. Multiple splice variants of Par3 and of a novel related gene, Par3L, produce proteins with different binding properties. *Gene* **2002**; 294:99–107.