

Screening Low-Frequency SNPs From Genome-Wide Association Study Reveals a New Risk Allele for Progression to AIDS

Sigrid Le Clerc, PhD,*†‡ Cédric Coulonges, MS,*‡ Olivier Delaneau, PhD,*
 Danielle Van Manen, MS,§ Joshua T. Herbeck, PhD,|| Sophie Limou, PhD,*†‡¶ Ping An, PhD,#
 Jeremy J. Martinson, MD,** Jean-Louis Spadoni, PhD,* Amu Therwath, PhD,††
 Jan H. Veldink, PhD,‡‡ Leonard H. van den Berg, PhD,‡‡ Lieng Taing, MS,* Taoufik Labib, MS,*
 Safa Mellak,* Matthieu Montes, PhD,* Jean-François Delfraissy, MD, PhD,‡
 François Schächter, PhD,* Cheryl Winkler, PhD,# Philippe Froguel, MD, PhD,§§|||
 James I. Mullins, PhD,|| Hanneke Schuitemaker, PhD,§ and Jean-François Zagury, MD, PhD*†‡

Background: Seven genome-wide association studies (GWAS) have been published in AIDS, and only associations in the HLA region on chromosome 6 and *CXCR6* have passed genome-wide significance.

Methods: We reanalyzed the data from 3 previously published GWAS, targeting specifically low-frequency SNPs (minor allele frequency <5%). Two groups composed of 365 slow progressors and 147 rapid progressors from Europe and the United States were compared with a control group of 1394 seronegative individuals using Eigenstrat corrections.

Results: Of the 8584 SNPs with minor allele frequency <5% in cases and controls (Bonferroni threshold = 5.8×10^{-6}), 4 SNPs showed statistical evidence of association with the slow progressor phenotype. The best result was for *HCP5* rs2395029 [$P = 8.54 \times 10^{-15}$, odds ratio (OR) = 3.41] in the HLA locus, in partial linkage disequilibrium with 2 additional chromosome 6 associations in

C6orf48 ($P = 3.03 \times 10^{-10}$, OR = 2.9) and *NOTCH4* (9.08×10^{-07} , OR = 2.32). The fourth association corresponded to rs2072255 located in *RICH2* ($P = 3.30 \times 10^{-06}$, OR = 0.43) in chromosome 17. Using *HCP5* rs2395029 as a covariate, the *C6orf48* and *NOTCH4* signals disappeared, but the *RICH2* signal still remained significant.

Conclusions: Besides the already known chromosome 6 associations, the analysis of low-frequency SNPs brought up a new association in the *RICH2* gene. Interestingly, *RICH2* interacts with BST-2 known to be a major restriction factor for HIV-1 infection. Our study has thus identified a new candidate gene for AIDS molecular etiology and confirms the interest of singling out low-frequency SNPs to exploit GWAS data.

Key Words: AIDS, disease progression, genome-wide association study, HIV-1, *RICH2*, SNP

(*J Acquir Immune Defic Syndr* 2011;56:279–284)

Received for publication August 18, 2010; accepted October 20, 2010.

From the *Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, France; †Université Paris, INSERM U955, Créteil, France; ‡ANRS Genomic Group (French Agency for Research on AIDS and Hepatitis), Paris, France; §Department of Experimental Immunology, Sanquin Research, Landsteiner Laboratory, Center for Infectious Diseases and Immunity Amsterdam Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands; ||Department of Microbiology, University of Washington School of Medicine, Seattle, WA; ¶CEA/Institut de Génomique, Centre National de Génotypage, Evry, France; #Laboratory of Genomic Diversity, SAIC-Frederick, Inc, National Cancer Institute-Frederick, Frederick, MD; **Department of Human Genetics University of Pittsburgh, Pittsburgh, PA; ††Laboratoire d'Oncologie Moléculaire, Université Paris, Paris, France; ‡‡Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; §§UMR CNRS 8090, Institut Pasteur de Lille, Lille, France; and |||Genomic Medicine, Hammersmith Hospital, Imperial College London, London, United Kingdom.

Supported by Agence Nationale de Recherche sur le SIDA (ANRS), Sidaction, Fondation de France, Innovation 2007 program of Conservatoire National des Arts et Métiers (CNAM), AIDS Cancer Vaccine Development Foundation, Neovacs SA, Vaxconsulting. Sophie Limou benefits from a fellowship from the French Ministry of Education, Technology and Research and Sigrid Le Clerc benefits from a fellowship of ANRS. 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 financially supported by the Netherlands National Institute for Public Health and the Environment. The authors acknowledge funding from the Netherlands Organization for Scientific Research (TOP, registration number 9120.6046). The MultiCenter AIDS Cohort Study is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute. UO1-AI-35042, UL1-RR025005 (GCRC), UO1-AI-35043, UO1-AI-35039, UO1-AI-35040, UO1-AI-35041. This project has been funded in part with federal funds from the National Institutes of Health, under contract HHSN261200800001E. This Research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

The content of this publication neither does 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.

Correspondence to: Jean-François Zagury, MD, PhD, 292 rue Saint Martin, 75003 Paris, France (e-mail: zagury@cnam.fr).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Copyright © 2011 by Lippincott Williams & Wilkins

INTRODUCTION

In recent years, the development of new DNA technologies has allowed the successful completion of genome-wide association studies, (GWAS) and multiple genetic associations were identified in several diseases such as celiac disease,¹ Schizophrenia,² or Type 2 diabetes.³ In AIDS, several GWAS have been published since 2007,^{4–10} and associations passing genome-wide significance have been found solely in chromosome 6 in the region of HLA (*HCP5*^{4–6}, *HLA-C*^{4,5,11}) and in the *CXCR6* gene.¹² The design of genotyping chips tends to rely mainly on common variants with minor allele frequencies (MAFs) >5%. Moreover, the power to detect low allele frequency SNP associations in AIDS is weakened for 2 complementary reasons: (1) lower frequency means less individuals at stake and thus weaker *P* values, and (2) AIDS-related genomic cohorts have enrolled fewer patients compared with other pathologies such as chronic kidney disease¹³ or Type 2 diabetes.³ However, low-frequency SNPs are predicted to have the potential for greater functional consequences than common alleles and may contribute strongly to genetic susceptibility to common diseases^{14–16}; thus, they constitute very good candidates for genetic association studies. We therefore decided to reanalyze the genome-wide data obtained from the Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort on slow progression⁶ and on rapid progression,⁷ by focusing specifically on low-frequency SNPs (MAF < 5%). To increase the power of the study, we also included in our analysis rapid and slow progressors from the Dutch ACS cohort¹¹ and from the American MultiCenter AIDS Cohort Study (MACS) 156 group.⁸

METHODS

The Case Groups

Slow Progressors and Rapid Progressors were gathered from 3 HIV cohorts:

The Genomics of Resistance to Immunodeficiency Virus Cohort

The GRIV cohort, established in 1995 in France, is a collection of DNA samples to identify host genes associated with slow progression and with rapid progression to AIDS.^{17–20} Only whittens of European descent living in France were eligible for enrollment to reduce confounding by population substructure. These criteria limit the influence of the ethnic and environmental factors (all subjects live in a similar environment and are infected by HIV-1 subtype B strains) and put an emphasis on the genetic make-up of each individual in determination of rapid progression (RP) or slow progression (SP) to AIDS. The RP and the SP were included on the basis of the main clinical outcomes, CD4 T-cell count, and time to disease progression. SP were defined as asymptomatic HIV-1-infected individuals for more than 8 years, no treatment and a CD4 T-cell count above 500 cells per cubic millimeter. The SP group (*n* = 270) was composed of 200 males and 70 females aged at inclusion from 19 to 62 (mean = 35). Rapid progressors (RP) were stringently defined as having a 2 or more CD4 T-cell counts below 300 cells per cubic millimeter less than 3 years

after the last seronegative testing. The RP group (*n* = 84) was composed of 72 males and 12 females aged at inclusion from 21 to 55 (median = 32). DNA was obtained from fresh peripheral blood mononuclear cells or from EBV-transformed cell lines. All patients provided written informed consent before enrollment in the GRIV genetic association study.

The Amsterdam Cohort Studies Cohort

The ACS (Amsterdam Cohort Studies) cohort is composed of 316 HIV-1 homosexual men and 100 HIV-1 drug users. This cohort was established to follow the course of HIV-1 infection using various endpoints related to HIV-1 infection and AIDS. The ACS participants were described in detail previously.¹¹

Since CD4 T-cell count was assessed during routine clinical follow-up, we could extract the ACS SP and RP patients respecting the GRIV criteria (SP: *n* = 36, RP: *n* = 41). The SP and RP status was easily determined among seroconverter subjects since the date of seroconversion was known. We could also extract SP from seroprevalent subjects when the time of seropositivity was known to be higher than 8 years.

The MultiCenter AIDS Cohort Study 156 Group

The MACS156 study comprises a subset of 156 HIV-1 homosexual men enrolled in the cohort, a prospective cohort originally established to investigate the natural history of HIV infection.²¹ This subset of MACS European American participants was chosen to be enriched with the extremes AIDS progression phenotypes.⁸ The MACS156 participants were described in detail previously.⁸

Because CD4 T-cell count was assessed during routine clinical follow-up, we could extract the MACS SP and PR respecting the GRIV definition (SP: *n* = 59, RP: *n* = 22). As with the ACS cohort, the SP and the RP were selected from seroconverter subjects. SP were also selected from seroprevalent subjects.

Control Groups

Three white control groups from France, The Netherlands, and United States were merged and used as a control group.

The Data from an Epidemiological Study on Insulin Resistance Syndrome 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. Subjects were recruited from 1994 to 1996 from volunteers insured by the French social security system, which offers periodic health examinations free of charge.²² This control group comprised 694 participants both nonobese and normoglycemic of the DESIR trial, all French and HIV-1 seronegative. It was composed of 281 males and 413 females aged from 30 to 64 years.

Dutch Controls (CTR-ACS)

This control group corresponds to 376 Dutch subjects genotyped with HumanHap300 BeadChips.²³

ILLUMINA CONTROL GROUP

This control group corresponds to white subjects genotyped with HumanHap300 BeadChips from the Illumina Genotyping Control Database (www.illumina.com). There were 324 individuals.

Genotyping Method, Quality Control

The GRIV cohort, the ACS cohort, and the control groups were genotyped using the Illumina Infinium II HumanHap300 BeadChips (Illumina, San Diego, CA). The genotyping quality was assessed for each group using the BeadStudio software (version 3.1, Illumina). Missing data >2%, MAF (<1%), and deviations from Hardy-Weinberg equilibrium in the control groups ($P < 1.0 \times 10^{-3}$) were removed during these quality control steps.^{6,7,11} The MACS156 group genotype data were obtained through the Affymetrix GeneChip Human Mapping 500K Array (Affymetrix, Santa Clara, CA). Different quality control filters were applied to ensure reliable genotyping data.⁸ For all the cohorts, we removed outliers exhibiting nonwhite ancestry, cohort by cohort, using the Eigenstrat method.²⁴

SNP Selection

We considered 2 pooled case groups (SP from GRIV, ACS, and MACS156 group on the one hand, RP from GRIV, ACS, and MACS156 on the other hand) and the pooled control groups (DESIR, CTR-ACS, Illumina-CTR). We retained the 8584 SNPs exhibiting a MAF < 5% for the slow progressors versus controls (SP-CTR) comparison (Bonferroni 5.8×10^{-6}), and 10295 SNPs exhibiting a MAF < 5% for the RP-CTR comparison (Bonferroni 4.8×10^{-6}). It was important to choose SNPs with low frequency in either groups because we were looking for factors either promoting progression (low MAF in SP compared with CTR or low MAF in CTR compared with RP) or preventing progression (low MAF in RP compared with CTR or low MAF in CTR compared with SP). The choice to screen specifically low-frequency SNPs stems from 2 complementary reasons: (1) a biological reason: HIV-1 infection is a multi-factorial disease with several genetic factors impacting disease. Several groups have pointed out that most signals involved in complex diseases should deal with the low-frequency variants spectrum.¹⁴⁻¹⁶ Indeed, this observation was confirmed in AIDS because the main signal found up to now was in *HCP5* with a low-frequency variant.^{4-6,9} (2) a statistical fact: in our case-control configuration, a low frequency either in the CTR or in the CASE group means systematically a weaker *P* value for a given odds ratio (OR which measures the real biological impact of the SNP). For instance, for an OR of 0.5 in the dominant mode, the *P* values obtained are: 0.02 with a SP MAF of 2% and a CTR MAF=3.9%, 1.5×10^{-5} with a SP MAF of 5% and a CTR MAF = 9.5%, 9.98×10^{-8} with a SP MAF of 10% and CTR MAF = 18.2%. Moreover, in the Illumina genotyping chips used, SNPs with low MAF are underrepresented in genotyping chips compared with SNPs with higher MAF (data not shown). Overall, for a biological effect measured by a given OR, SNP associations are thus more difficult to identify in the low-frequency spectrum because they exhibit weaker *P* values by essence and because they are underrepresented

and thus artificially penalized through global Bonferroni corrections.

Statistical Analysis

We performed a case-control analysis by comparing either the SP group ($n = 365$) or the RP group ($n = 147$) consisting in GRIV, ACS, and MACS patients with the control group consisting of DESIR, CTR-ACS, and Illumina control individuals ($n = 1394$). The statistical analysis was performed by a logistic regression (with SNPtest software²⁵) in the dominant mode, taking into account stratification by adding the 2 first Eigenstrat PC axes as covariates using EIGENSOFT. Testing for association under the dominant model was appropriate since we lacked power to test for associations under the recessive model and additionally, in this context, the dominant model is identical to the additive mode. For each SNP passing the Bonferroni threshold, we recomputed the regression by adding the *HCP5* SNP (rs2395029) as a covariate to check for nonindependence due to linkage disequilibrium (with SNPtest software²⁵).

SNP Imputation

Using SNPtest Impute software,²⁵ it was possible to impute untyped SNPs of the MACS156 study subjects, absent from the the Affymetrix GeneChip Human Mapping 500K Array (Affymetrix, Santa Clara, CA) and present in the Illumina HumanHap300 BeadChips (Illumina). They were imputed using the HapMap release 21 phased data for the European population (Utah residents with ancestry from northern and western Europe [CEU]) as panel of reference (<http://www.hapmap.org>). Only the SNPs imputed with high reliability (imputation quality score²⁵ $P > 0.9$) were retained.

Internal Replication

After comparing the total SP group (combined from GRIV, ACS, and MACS 156) with the total control group (combined from DESIR, CTR-ACS, and CTR_Illumina), we also performed an individual analysis of each group GRIV, ACS, and MACS156 for the 4 SNPs passing the Bonferroni threshold. For GRIV, we checked the result obtained in our previous GWAS.⁶ For ACS, we tested the association between the four SNPs and times to time to AIDS93 by linear regression. For MACS156 group, we also tested the association between the SNPs and time to AIDS93 by linear regression.

Linkage Disequilibrium

For each SNP exhibiting a significant association, we looked for the other SNPs in linkage disequilibrium ($r^2 \geq 0.9$) in the HapMap population of Western European ancestry (Utah residents with ancestry from northern and western Europe [CEU], HapMap data Release 21a/phase II January 2007, on NCBI B35 assembly, dbSNP125, <http://www.hapmap.org>) to identify the genes possibly tracked by the SNP associations. A SNP was assigned to a gene if it was located within the gene or in the 2 kilobase flanking regions (potential regulatory sequence), otherwise it was considered intergenic.

Bioinformatics Exploration

To further explore the associations observed, we tried to identify putative modifications in mRNA expression as

shown in *Genevar*²⁶ and Dixon²⁷ databases, in splicing (FastSNP²⁸ http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp), in polyadenylation (polyAH, <http://linux1.softberry.com/berry.phtml?topic=polyah&group=programs&sub-group=promoter> and polyApred, <http://www.imtech.res.in/raghava/polyapred/submission.html>), or in transcription factor binding sites (SignalScan, <http://www-bimas.cit.nih.gov/molbio/signal/>, TESS, <http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME>, and TFSearch, <http://www.cbrc.jp/research/db/TFSEARCH.html>, derived from TRANSFAC database). We used the Genecards database to look for the tissues and organs expressing the proteins (GeneCards,²⁹ <http://www.genecards.org>).

RESULTS

For the 8584 SNPs with a MAF < 5%, we tested a total of 365 SP patients and compared them with a control cohort of 1394 seronegative individuals (from France, The Netherlands, and United States; Methods). Four signals passed the Bonferroni threshold in the dominant mode, 3 in chromosome 6 and 1 in chromosome 17 (Table 1). Not unexpectedly, the best result was obtained for the well-replicated *HCP5* rs2395029 ($P = 8.54 \times 10^{-15}$). The 2 other associations found in chromosome 6 were for rs9368699 in *C6orf48* ($P = 3.03 \times 10^{-10}$) and rs8192591 in *NOTCH4* ($P = 9.08 \times 10^{-7}$). These genes are in partial linkage disequilibrium with *HCP5*-rs2395029 (resp. $r^2 = 0.68$, $r^2 = 0.57$). The fourth association corresponded to the chromosome 17 SNP rs2072255 located in the *RICH2* gene ($P = 3.30 \times 10^{-6}$). This SNP is in full Linkage disequilibrium (LD) ($r^2 = 1$) with rs2072254 corresponding to a synonymous change (Gly188Gly) of *RICH2* (Fig. 1; see **Map, Supplemental Digital Content 1**, <http://links.lww.com/QAI/A114>). The 4 SNPs corresponded to the higher end of the MAF distribution of the SNP studied (see **Histogram, Supplemental Digital Content 2**, <http://links.lww.com/QAI/A115>), which is logical since larger numbers of subjects lead to stronger P values.

To evaluate the role of LD in these associations, we recomputed the P values using the *HCP5* SNP (rs2395029) as covariate. The 2 chromosome 6 SNPs were not significant in the adjusted analysis ($P = 0.78$ for rs9368699 in *C6orf48*, $P = 0.31$ for rs8192591 in *NOTCH4*), but the association remained statistically robust for the chromosome 17 SNP after

controlling for the *HCP5* SNP, $P = 1.82 \times 10^{-6}$ for rs2072255 in *RICH2*. In line with this computation, we found that the rs2072255 frequency was not significantly different between the SP elite controllers¹² and among the SP nonelite controllers¹² ($P = 0.7$).

The subjects carrying the rs2072255-A allele were 9.04% in the SP group (Table 1), 18.87% in the control group, and 17.1% in the RP group (this excludes the hypothesis of an effect on infection). Interestingly these frequencies were consistent within each of the 3 SP groups and within each of the 3 control groups (Fig. 2). Moreover, the positive signal for association of rs2072255 was confirmed in the individual cohorts GRIV, ACS, and MACS156 study: the comparison of the NP with DESIR controls in GRIV led to $P = 8.1 \times 10^{-5}$, the analysis of ACS by linear regression led to $P = 0.05$, and the analysis of the MACS156 group by linear regression led also to $P = 0.05$. A table summarizing the results in the different cohorts is provided in **Supplemental Digital Content 3** (<http://links.lww.com/QAI/A116>). Finally, the rs2072254 *RICH2* exonic SNP (in LD with the rs2072255) is located in a splicing site according to FastSNP²⁸ (Fig. 1).

When comparing the 10295 selected SNPs between the 147 RP patients with the 1394 seronegative control group, no signal passed the Bonferroni threshold.

DISCUSSION

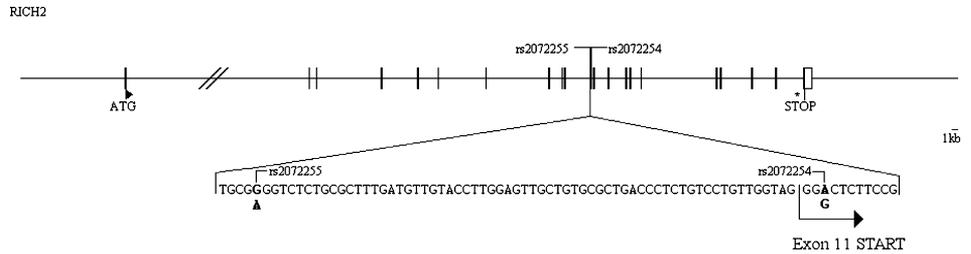
We decided to reanalyze previous GWAS data on AIDS cohorts by focusing specifically on low-frequency SNPs (MAF < 5%). For that, we combined rapid and slow progressors from 3 international cohorts from France (GRIV), Netherlands (ACS), and United States (MACS156 study) totalling 365 SP and 147 RP, who were compared with 1394 controls (seronegative individuals). No association was found when comparing the RP group with the CTR group. This was not a surprise since the RP group comprised only 147 individuals; and for a MAF of 5% in the CTR, one needed to get a quite strong biological effect ($OR > 2.8$) to pass the Bonferroni threshold. Four SNPs passed the Bonferroni threshold when comparing the SP group with the CTR group. Among them, 3 are in chromosome 6 and were previously found significant by several studies: rs2395029 in *HCP5*^{4-6,8,9,11}, rs9368699 in *C6orf48*^{5,6}, rs8192591 in

TABLE 1. Four SNPs Passed the Bonferroni Threshold After Comparing Slow Progressors Against Controls on Low-Frequency SNPs (MAF < 5%)

SNP	chr	A1	A2	MAF _{SP}	MAF _{CTR}	Frequency Dominant Mode: A1A1 and A1A2		P slow progressors versus controls (SP-CTR)	OR	Gene(s)/LD
						SP _{ALL}	CTR _{ALL}			
rs2395029	6	C	A	10.19%	3.41%	20.38%	6.81%	8.54×10^{-15}	3.41 CI 95%: 2.4 to 4.8	<i>HCP5</i> , intergenic, <i>MICB</i>
rs9368699	6	G	A	9.20%	3.48%	18.13%	6.96%	3.03×10^{-10}	2.9 CI 95%: 2.04 to 4.16	<i>C6orf48</i> , <i>RDBP</i> , <i>TNXB</i>
rs8192591	6	T	C	6.83%	3.15%	13.63%	6.30%	9.08×10^{-7}	2.32 CI 95%: 1.56 to 3.41	<i>NOTCH4</i> <i>GPSM3</i>
rs2072255	17	A	G	4.79%	10.04%	9.04%	18.87%	3.30×10^{-6}	0.43 CI 95%: 0.28 to 0.63	<i>RICH2</i>

For each SNP, the chromosome, the minor allele A1, the MAF, the frequency of the group in the dominant mode (used for association tests), the P value in the dominant mode, the OR, and the genes concerned (SNP and/or SNP(s) in LD with $r^2 \geq 0.9$ in Caucasians HapMap data) are provided. A SNP was assigned to a gene if it was located in the gene or within the 2kb flanking regions.

FIGURE 1. Genetic map of the *RICH2* gene region. 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 by an asterisk (*), respectively. The 2 SNPs rs2072255 (intronic) and rs2072254 (exonic) are represented ($r^2 = 1$) and a zoom with the genetic sequences is provided.



*NOTCH4*⁵. *NOTCH4* is an interesting candidate gene due to its role in immune regulation and the *NOTCH4* rs8192591 corresponds to a nonsynonymous Gly534Ser protein variant. This association was found independent from the *HCP5* rs2395029 by Fellay et al,⁵ however, the signal disappeared in our own study when using *HCP5* rs2395029 as covariate. A possible explanation for this discrepancy could be the use of viral load as an endpoint in the study by Fellay et al,⁵ whereas we used here a progression phenotype.

The fourth signal identified in the present study is new and corresponds to rs2072255 in the chromosome 17 *RICH2* gene. The *RICH2* gene encodes a Rho-type GTPase activator composed of 818 amino acid (89,247 kDa) with an intracellular localization. It is expressed highly in the brain and at a basal level in several tissues notably in the lymph nodes.²⁹ A recent study has shown that *RICH2* is a part of the physical link between BST-2 and the actin cytoskeleton and prevents the internalization of BST-2 which prevents HIV-1 virion budding and release.³¹ This is a counteraction of HIV-1 Vpu known to favor internalization and degradation of BST-2.³¹ The rs2072255-A *RICH2* allele favors progression to AIDS because 18.87% of the CTR carry the variant, whereas only 9.04% of the SP carry it (Table 1, Fig. 2). Interestingly, rs2072255 is in total LD with the SNP rs2072254 located in

a splicing site of *RICH2* as suggested by FastSNP (Fig. 2). If the rs2072254-G minor allele alters mRNA splicing, it could lead to a down-modulation of *RICH2*, and thus explain a diminished effect of BST-2 against HIV-1 production.

The identification of 3 chromosome 6 signals already confirmed by several studies shows the relevance of targeting specifically low-frequency SNPs in GWAS. The genetic and biological data regarding the *RICH2* signal are also quite compelling and provide a new relevant candidate gene to explore the molecular etiology of HIV-1 pathogenesis. Further genetic and experimental studies will be needed to confirm and understand the effect of *RICH2* in AIDS pathogenesis.

ACKNOWLEDGMENTS

The authors are grateful to all the patients and medical staff who have kindly collaborated with the GRIV project. Data in this article were collected by the Multicenter AIDS Cohort Study (MACS) with centers (Principal Investigators) at The Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa P. Jacobson), Howard Brown Health Center, Feinberg School of Medicine, Northwestern University, and Cook County Bureau of Health Services (John P. Phair, Steven M. Wolinsky), University of California, Los Angeles (Roger Detels), and University of Pittsburgh (Charles R. Rinaldo). Website located at <http://www.statepi.jhsph.edu/macs/mac.html>.

REFERENCES

1. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet.* 2010;42: 295–302.
2. O’Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet.* 2008;40:1053–1055.
3. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40:638–645.
4. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science.* 2007;317: 944–947.
5. Fellay J, Ge D, Shianna KV, et al. Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* 2009;5:e1000791.
6. 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; 199:419–426.
7. Le Clerc S, Limou S, Coulonges C, et al. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *J Infect Dis.* 2009;200:1194–1201.

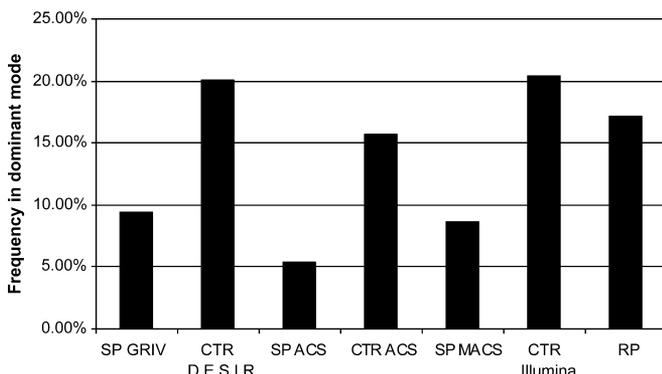


FIGURE 2. Frequency of individuals carrying the rs2072255-A variant (dominant mode) in the various groups: SP GRIV (n = 270), SP ACS (n = 31), SP MACS (n = 59), the D.E.S.I.R. control group (n = 694). CTR-ACS (n = 376), the Illumina CTR (n = 324), and the RP group (n = 147). The frequencies within each of the GRIV RP, ACS RP, and MACS156 study RP groups was similar.

8. 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;201:618–626.
9. Dalmasso 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. *PLoS One.* 2008;3:e3907.
10. Pelak K, Goldstein DB, Walley NM, et al. Host determinants of HIV-1 control in African Americans. *J Infect Dis.* 2010;201:1141–1149.
11. van Manen D, Kootstra NA, Boeser-Nunnink B, et al. Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS.* 2009;23:19–28.
12. Limou S, Coulonges C, Herbeck JT, et al. Multi-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in AIDS long-term non-progression. *J Infect Dis.* 2010;202:908–915.
13. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010;42:376–384.
14. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet.* 2007;80:727–739.
15. Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet.* 2008;82:100–112.
16. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet.* 2008;40:695–701.
17. Flores-Villanueva PO, Hendel H, Caillat-Zucman S, et al. Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development. *J Immunol.* 2003;170:1925–1929.
18. Hendel H, Caillat-Zucman S, Lebuane H, et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol.* 1999;162:6942–6946.
19. Rappaport J, Cho YY, Hendel H, et al. 32 bp CCR-5 gene deletion and resistance to fast progression in HIV-1 infected heterozygotes. *Lancet.* 1997;349:922–923.
20. Winkler CA, Hendel H, Carrington M, et al. Dominant effects of CCR2-CCR5 haplotypes in HIV-1 disease progression. *J Acquir Immune Defic Syndr.* 2004;37:1534–1538.
21. Kaslow RA, Ostrow DG, Detels R, et al. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol.* 1987;126:310–318.
22. Balkau B. An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome [in French]. *Rev Epidemiol Sante Publique.* 1996;44:373–375.
23. van Es MA, Veldink JH, Saris CG, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet.* 2009;41:1083–1087.
24. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–909.
25. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39:906–913.
26. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.* 2008;18:640–643.
27. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet.* 2007;39:1202–1207.
28. 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(Web Server issue):W635–W641.
29. Rebhan M, Chalifa-Caspi V, Prilusky J, et al. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* 1997;13:163.
30. Rollason R, Korolchuk V, Hamilton C, et al. A CD317/tetherin-RICH2 complex plays a critical role in the organization of the subapical actin cytoskeleton in polarized epithelial cells. *J Cell Biol.* 2009;184:721–736.
31. Tokarev A, Skasko M, Fitzpatrick K, et al. Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS Res Hum Retroviruses.* 2009;25:1197–1210.