MAJOR ARTICLE

Genomewide Association Study of a Rapid Progression Cohort Identifies New Susceptibility Alleles for AIDS (ANRS Genomewide Association Study 03)

Sigrid Le Clerc,[1,2,4,a] Sophie Limou,[1,2,4,5,a] Cédric Coulonges,[1,2] Wassila Carpentier,[2] Christian Dina,[6] Lieng Taing,[1] Olivier Delaneau,[1] Taoufik Labib,[1,4] Rob Sladek,[8] ANRS Genomic Group,[2,b] Christiane Deveau,[2] Hélène Guillemain,[1] Rojo Ratsimandresey,[1] Matthieu Montes,[1] Jean-Louis Spadoni,[1] Amu Therwath,[3] François Schächter,[1] Fumihiko Matsuda,[9] Ivo Gut,[5] Jean-Daniel Lelièvre,[4] Yves Lévy,[4] Philippe Froguel,[6,10] Jean-François Delfraissy,[2] Serge Hercberg,[7] and Jean-François Zagury[1,2,4]

[1]Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, [2]Agence Nationale de Recherche sur le SIDA (ANRS) Genomic Group, and [3]Laboratoire d'Oncologie Moléculaire, Paris, and [4]Université Paris 12, Institut National de la Santé et de la Recherche Médicale (INSERM) U955, Créteil, [5]Commissariat à L'Energie Atomique/Institut de Génomique, Centre National de Génotypage, Evry, [6]Unité Mixte de Recherche (UMR) Centre National de la Recherche Scientifique 8090, Institut Pasteur de Lille, Lille Cedex, and [7]UMR U557 INSERM/U1125 Institut National de la Recherche Agronomique/Conservatoire National des Arts et Métiers/Université Paris 13, and Centre de Recherche en Nutrition Humaine Ile-de-France, Santé-Médecine-Biologie Humaine (SMBH) Paris 13, Bobigny, France; [8]Department of Human Genetics, Faculty of Medicine, McGill University, and Génome Québec Innovation Centre, Montreal, Canada; [9]INSERM U852, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; [10]Genomic Medicine, Hammersmith Hospital, Imperial College London, London, United Kingdom

Background. Previous genomewide association studies (GWASs) of AIDS have targeted end points based on the control of viral load and disease nonprogression. The discovery of genetic factors that predispose individuals to rapid progression to AIDS should also reveal new insights into the molecular etiology of the pathology.

Methods. We undertook a case-control GWAS of a unique cohort of 85 human immunodeficiency virus type 1 (HIV-1)–infected patients who experienced rapid disease progression, using Illumina HumanHap300 BeadChips. The case group was compared with a control group of 1352 individuals for the 291,119 autosomal single-nucleotide polymorphisms (SNPs) passing the quality control tests, using the false-discovery rate (FDR) statistical method for multitest correction.

Results. Novel associations with rapid progression (FDR, $\leqslant$25%) were identified for *PRMT6* ($P = 6.1 \times 10^{-7}$; odds ratio [OR], 0.24), *SOX5* ($P = 1.8 \times 10^{-6}$; OR, 0.45), *RXRG* ($P = 3.9 \times 10^{-6}$; OR, 3.29), and *TGFBRAP1* ($P = 7 \times 10^{-6}$; OR, 0.34). The haplotype analysis identified exonic and promoter SNPs potentially important for *PRMT6* and *TGFBRAP1* function.

Conclusions. The statistical and biological relevance of these associations and their high ORs underscore the power of extreme phenotypes for GWASs, even with a modest sample size. These genetic results emphasize the role of the transforming growth factor β pathway in the pathogenesis of HIV-1 disease. Finally, the wealth of information provided by this study should help unravel new diagnostic and therapeutic targets.

Genomewide association studies (GWASs) may provide new insights into the molecular etiology of complex diseases by discovering unsuspected genetic risk factors and, as a consequence, identify new diagnostic or therapeutic targets [1]. Reports of 3 GWASs of AIDS have already been published [2–4] and described mainly

Received 8 April 2009; accepted 28 May 2009; electronically published 15 September 2009.

Reprints or correspondence: Dr Jean-François Zagury, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75003 Paris, France (zagury@cnam.fr).

The Journal of Infectious Diseases 2009; 200:000–000

0022-1899/2009/20008-00XX$15.00
DOI: 10.1086/605892

Potential conflicts of interest: none reported.

Financial support: Agence Nationale de Recherche sur le SIDA (ANRS); Innovation 2007 program of Conservatoire National des Arts et Métiers; AIDS Cancer Vaccine Development Foundation; Neovacs SA; Vaxconsulting. S.L. benefits from a fellowship from the French Ministry of Education, Technology, and Research, and S.L.C. benefits from a fellowship from the ANRS.

[a] S.L.C. and S.L. contributed equally to this work.

[b] The ANRS Genomic Group oversees the AIDS genomic projects of the ANRS; study group members are listed at the end of the text.

genes involved in the control of viral load. Indeed, these 3 GWASs identified the *HCP5* rs2395029 polymorphism, which is in linkage disequilibrium with HLA-B*57 and other immunity genes, such as *MICB*, *TNF*, *LTB*, *BAT1*, and *HLA-C*. An association with the HIV DNA level (reservoir) was also depicted by the PRIMO GWAS [2] for *SDC2* (chromosome 8), whose encoded protein is required for Tat internalization. Finally, the *ZNRD1* locus of chromosome 6 was associated with the control of disease progression but not of viral load in both the Euro-CHAVI (Center for HIV/AIDS Vaccine ImmunologyCenter for HIV/AIDS Vaccine Immunology) [3] and the nonprogressor Genomics of Resistance to Immunodeficiency Virus (GRIV) [4] GWASs.

To identify genetic loci predisposing a person to rapid progression of AIDS rather than disease control, we undertook a case-control GWAS involving a unique cohort of human immunodeficiency virus type 1 (HIV-1)–infected patients who experienced rapid progression, using Illumina HumanHap300 BeadChips. In a manner symmetric to the published nonprogressor GRIV GWAS [4], the use of this extreme phenotype should lead to an enrichment of our knowledge of the genetic factors involved in rapid disease progression. The power of this extreme design has indeed been demonstrated by previous candidate gene studies [5–7].

METHODS

The GRIV cohort. The GRIV cohort was established in France in 1995 to generate a large collection of DNA for genetic studies to identify host genes associated with either rapid progression or nonprogression to AIDS [5, 7]. Only white people of European descent living in France were eligible for enrollment to reduce confounding by population substructure. These criteria limit the influence of ethnic and environmental factors (all subjects live in a similar environment and are infected by B strains) and emphasize how the genetic makeup of each individual determines the various patterns of progression. Rapid progressors were included on the basis of the main clinical outcomes, CD4 T cell count and time to disease progression, and were defined as those who had 2 or more CD4 T cell counts below 300 cells/mm^3 within 3 years after the last seronegative test result. DNA was obtained from fresh peripheral blood mononuclear cells or from Epstein-Barr virus–transformed cell lines. The rapid progression group ($n = 85$) was composed of 73 men and 12 women aged 21–55 years (median, 32 years) at inclusion. At inclusion, the median CD4 T cell count was 230 cells/mm^3 (minimum and maximum values, 20 and 297 cells/mm^3). All patients provided written informed consent before enrollment in the GRIV GWAS.

The seropositive control population. To discriminate between positive signals corresponding to either an association with rapid progression or an association with HIV-1 infection, we needed a group of seropositive control subjects who were not rapid progressors. For that, we used 275 white French subjects who qualified as nonprogressors to AIDS (ie, those who had an asymptomatic HIV-1 infection for >8 years, no receipt of treatment, and a CD4 T cell count consistently remaining at >500 cells/mm^3). This control group was composed of 201 men and 74 women aged 19–62 years (median, 35 years) at inclusion. The median CD4 T cell count of this seropositive control population was 706 cells/mm^3 (minimum and maximum values, 501 and 2298 cells/mm^3).

The SU.VI.MAX seronegative control group. The SU.VI.MAX (Supplémentation en Vitamines et Minéraux Antioxydants) study was a randomized, double-blind, placebo-controlled, primary-prevention trial designed to test the efficacy of daily supplements of antioxidant vitamins and minerals at nutrition-level doses in reducing several major health problems in industrialized countries, especially the main causes of premature death, cancers and cardiovascular diseases. This cohort study was started in 1994 in France and included 12,735 subjects [8]. The control group genotyped in the present study comprised 1352 representative SU.VI.MAX participants, all of them white persons living in France who were HIV-1 seronegative. This control cohort was composed of 525 men and 827 women, with a mean age of 53.1 and 48.5 years, respectively.

The second seronegative control group. The D.E.S.I.R. (Data from an Epidemiological Study on Insulin Resistance Syndrome) program was a 9-year follow-up study designed to clarify the development of 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 [9]. This second control group comprised 697 participants who were both not obese and normoglycemic from the D.E.S.I.R. trial, all French and HIV-1 seronegative. It was composed of 281 men and 416 women aged 30–64 years.

Genotyping method. The GRIV cohort and the 2 seronegative groups were genotyped using Illumina Infinium II HumanHap300 BeadChips (Illumina). Genomic DNA (750 ng) was whole-genome amplified, fragmented, denatured, and hybridized on prepared HumanHap300 BeadChips for a minimum of 16 h at 48°C. Non–specifically hybridized fragments were removed by washing, and the remaining specifically hybridized DNA was fluorescently labeled by a single base extension reaction and detected using a BeadArray scanner (Illumina). Normalized bead-intensity data obtained for each sample were loaded into BeadStudio software (version 3.1; Illumina), which converted fluorescence intensities into single-nucleotide polymorphism (SNP) genotypes.

Quality control. Using the BeadStudio software, we analyzed the crude genotyping data, and SNPs were filtered according to the following parameters. First, samples with a call

Table 1. Best Results Obtained for the Comparison between Rapid Progressors and Control Subjects

SNP	Gene	Chr	Chr position	A1	A2	Allelic frequency (A1), %				OR (95% CI)	Fisher *P* Value, RP-$CTR_{SU.VI.MAX}$	FDR *q* value
						RP	$CTR_{SU.VI.MAX}$	$CTR_{D.E.S.I.R.}$	SCP			
rs4118325	Intergenic[a]	1	107379355	A	G	5.2	19.0	18.7	15.8	0.24 (0.12–0.46)	6.09×10^{-7}	0.17
rs1522232	*SOX5*	12	24285639	T	C	29.1	47.7	48.2	50.7	0.45 (0.32–0.63)	1.80×10^{-6}	0.20
rs1360517	Intergenic	9	12997129	A	G	16.3	5.9	7.0	5.8	3.09 (2.00–4.78)	3.27×10^{-6}	0.20
rs3108919	Intergenic	8	101910722	C	T	43.6	26.6	27.8	27.5	2.13 (1.56–2.91)	3.86×10^{-6}	0.20
rs10800098	*RXRG*	1	163675719	A	G	14.5	4.9	4.7	5.6	3.29 (2.08–5.20)	3.86×10^{-6}	0.20
rs10494056	Intergenic[b]	1	107349442	A	C	5.8	18.6	17.9	16.1	0.27 (0.14–0.52)	4.29×10^{-6}	0.20
rs12351740	Intergenic[c]	9	13010010	T	C	12.8	4.1	4.7	3.8	3.46 (2.12–5.62)	6.63×10^{-6}	0.25
rs1020064	*TGFBRAP1*[d]	2	105264172	T	G	9.3	23.2	25.6	25.2	0.34 (0.20–0.57)	7.04×10^{-6}	0.25

NOTE. *P* values were computed by the Fisher exact test in the allelic frequency mode and are presented with their corresponding allelic frequencies in the various populations (rapid progressors [RP], seronegative control subjects [$CTR_{SU.VI.MAX}$], and the seropositive control population [SCP]), chromosome (Chr) positions, odds ratios (ORs) with 95% confidence intervals (95% CIs), and false-discovery rate (FDR) *q* values. The frequencies in the D.E.S.I.R. second seronegative control group ($CTR_{D.E.S.I.R.}$) are also indicated and are similar to those in the SU.VI.MAX cohort. SNP, single-nucleotide polymorphism.

[a] This intergenic SNP is in linkage disequilibrium with *PRMT6* and intergenic SNPs.

[b] This intergenic SNP is in linkage disequilibrium with *PRMT6* and intergenic SNPs ($r^2 = 0.92$ with rs4118325).

[c] This intergenic SNP is in partial linkage disequilibrium with the intergenic SNP rs1360517 ($r^2 = 0.68$).

[d] This *TGFBRAP1* SNP is in linkage disequilibrium with an intergenic SNP.

rate (percentage of SNPs genotyped by sample) <95% in the Illumina clusters were deleted. Second, the SNPs with a call frequency (percentage of samples genotyped by SNP) <99% were reclustered. Third, after reclustering, samples with a call rate below 97% were deleted. The clustering step can create SNP genotyping errors, which can be prevented by following the Illumina procedure (see http://www.illumina.com/downloads/GTDataAnalysis_TechNote.pdf). This method evaluates the quality of the newly created clusters according to several criteria, which can be manually checked and corrected as necessary. In total, 1300 SNPs were excluded by this Illumina quality control procedure. Finally, after all the quality control steps were performed, the 15,731 SNPs with a call frequency <98% (>2% of missing data) were excluded. This procedure ensures reliable genotyping data with little missing data.

Hardy-Weinberg equilibrium analysis was performed for each SNP in each group by using an exact statistical test implemented in PLINK software ([10]; available at: http://pngu.mgh.harvard.edu/~purcell/plink/). Deviation from Hardy-Weinberg equilibrium in a group of patients suggests that the SNP has a biological effect, while deviation in the control group or all groups suggests a systematic error in genotyping. The 1475 SNPs that were not in the Hardy-Weinberg equilibrium in the SU.VI.MAX control group ($P < 1.0 \times 10^{-3}$) were rejected in this way.

In total, 235 SNPs with low minor allelic frequency (<1%) in the global population were also filtered.

Haplotype inference. Haplotype inference was obtained using the rapid and accurate Shape-IT algorithm [11].

Linkage disequilibrium. For each SNP exhibiting a significant association, we looked for the other SNPs in linkage disequilibrium ($r^2 \geq 0.8$) in the HapMap population of Western European ancestry (CEU, HapMap data Release 21a/phase II, January 2007, on NCBI B35 assembly, dbSNP125; available at http://www.hapmap.org) to identify the genes possibly concerned by the associations. A SNP was assigned to a gene if it was located in the gene or in the 2 kb flanking regions (potential regulatory sequence); otherwise, it was considered intergenic.

Statistical analysis. For each SNP, we performed a standard case-control analysis by using the Fisher exact test (with PLINK software) to compare allelic distributions between the rapid progressors and the control subjects.

To take into account the multiple tests while controlling for the risk of false discovery, we computed for each *P* value a false-discovery rate (FDR) under the form of a *q* value: the *q* value is an estimate of the proportion of false-positive signals below a threshold *P* value [12]. The FDR computation is more complex but more powerful than the standard Bonferroni corrections [13–15], because it allows the identification of more true-positive signals. For polyfactorial diseases in which several genes are at stake, it thus provides a more adapted outlook on the GWAS results than the "all or nothing" Bonferroni cutoff. We thus used an FDR approach called local base estimating, with a 25% threshold for our case-control study.

For all the SNPs meeting this statistical threshold (Table 1), the quality of genotyping was individually rechecked with the BeadStudio software. We also checked that the allelic frequen-

This figure is available in its entirety in the online version of the *Journal of Infectious Diseases.*

Figure 1. Number of independent ($r^2 < 0.5$) single-nucleotide polymorphisms (SNPs) meeting the false-discovery rate threshold of 25%.

This figure is available in its entirety in the online version of the *Journal of Infectious Diseases*.

Figure 2. Quantile-quantile plot for expected (*red*) versus observed (*black*) *P* values from the comparison of rapid progressors with control subjects.

cies in the seropositive control population were similar to those in the seronegative SU.VI.MAX control population for those SNPs of interest, confirming that the observed associations were indeed linked to rapid progression.

Statistical reliability of the rapid progression results. The GRIV rapid progression cohort is unique, and no other independent cohort was readily available for replication. To check in an independent fashion the statistical relevance of the associations with rapid progression, we decided to evaluate the odds of obtaining just as many associations with the same statistical method (ie, the Fisher exact test and FDR cutoff) by comparing the genotypes of 1000 control subgroups (randomly extracted from the D.E.S.I.R. control cohort and composed of 85 subjects each) with the genotypes of the entire SU.VI.MAX control group. For each simulation, we counted the number of independent ($r^2 < 0.5$) SNPs meeting the statistical threshold of an FDR of 25% (Figure 1).

Identification of population stratification. To correct for possible population stratification at the intercontinental level, case and control genotypes were analyzed using STRUCTURE software (version 2.2; see [16] and http://pritch.bsd.uchicago.edu/software.html). We selected a set of 328 SNPs informative for ancestral origin (*F* statistics fixation index, >0.2) based on the Perlegen data set and separated by 5 Mb to avoid linkage disequilibrium. We also included genotypes obtained from unrelated individuals representing the 3 populations studied by the HapMap project to better separate our rapid progressors and control individuals according to their continent of origin. While nearly all case and control subjects were within the range of the white individuals from HapMap, one individual was outside the white subjects cluster in the rapid progressors (decreasing the rapid progression group from 86 to 85 subjects).

To avoid spurious associations resulting from possible population stratification or genotyping errors, a quantile-quantile plot was also produced by plotting the ranked values of the test statistics against the approximated expected order statistic (Figure 2). We also computed the genomic inflation factor λ [17]. The result ($\lambda = 1.038$), along with the quantile-quantile plot, suggested little overall effect of stratification.

RESULTS

After the quality control tests, a total of 291,119 autosomal SNPs were tested for association with rapid progression. Figure 3 depicts the distribution of the *P* values along the chromosomes, and Table 1 presents the best signals (for an extended list, see Table 2). Eight associations with FDRs ⩽25% were identified, corresponding to 6 independent ($r^2 < 0.5$) SNPs. Since no replication was readily available, we performed simulations on independent control populations to evaluate the reliability of these results (see Methods and Figure 1). Overall, less than 1% of these tests yielded 6 or more independent signals with a FDR ⩽25% (mean ± standard deviation = 0.4 ± 1.29), suggesting that most of the rapid progression associations found in this study are likely to be true positives. This result underscores the statistical relevance of the rapid progression associations, even with a modest sample size.

The best rapid progression signals were observed for SNPs on chromosome 1 (rs4118325 [$P = 6.09 \times 10^{-7}$] and rs10494056 [$P = 4.29 \times 10^{-6}$]) (Figure 3), which are in linkage disequilibrium themselves ($r^2 = 0.92$) and are also in linkage disequilibrium with SNPs of the *PRMT6* gene. The rs4118325-A allele

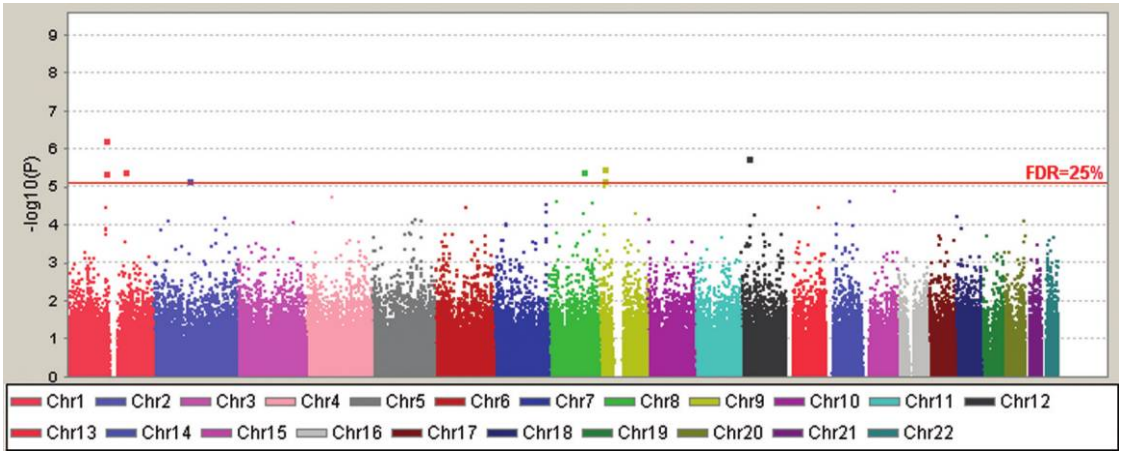

Figure 3. Distributions along the human autosomes of $-\log_{10}$(*P* values) obtained for the comparison of rapid progressors with control subjects. The red line marks the false-discovery rate threshold of 25%. Chr, chromosome.

Table 2. Fifty Best *P* Values Obtained for the Comparison between Rapid Progressors and Control Subjects

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*.

and/or the alleles in linkage disequilibrium are associated with prevention of rapid progression, with an odds ratio (OR) of 0.24 [95% confidence interval [CI], 0.12–0.46] (Table 1). Another association was identified on chromosome 1 in *RXRG* gene ($P = 3.86 \times 10^{-6}$, OR, 3.29 [95% CI, 2.08–5.20]). Finally, SNPs modulating rapid progression were also found in *SOX5* ($P = 1.80 \times 10^{-6}$; OR, 0.45 [95% CI, 0.32–0.63]) and *TGFBRAP1* ($P = 7.04 \times 10^{-6}$; OR, 0.34 [95% CI, 0.20–0.57]). Two other independent SNPs, rs1360517 and rs3108919, met the FDR threshold of 25% (Table 1) but were not close to any known gene (distance, <20 kb).

To complete our analysis, we explored the influence of covariables—such as *CCR5*-Δ32 and *CCR5-P1* haplotypes [7], the HIV-1 infection mode (mucosal or parenteral), age at seroconversion, and sex—on all the associations presented in Table 1. None of them was found to affect the observed signals.

To deepen our genomic analysis, we also tried to combine our results with previously published data from the Euro-CHAVI cohort, which assessed the viral load end point [3]. We combined their *P* values with ours using the classical Fisher method. Unfortunately, no combined *P* values met the FDR threshold of 25% (data not shown; best combined *P* value, 6.23×10^{-5}). The lack of significant common signals between the 2 studies may likely stem from the difference in inclusion criteria, in particular the use of viral load versus the use of CD4 T cell count, and also from the elimination of very rapid progressors in the Euro-CHAVI study, since such patients could not exhibit a sufficiently prolonged viral load set point [3].

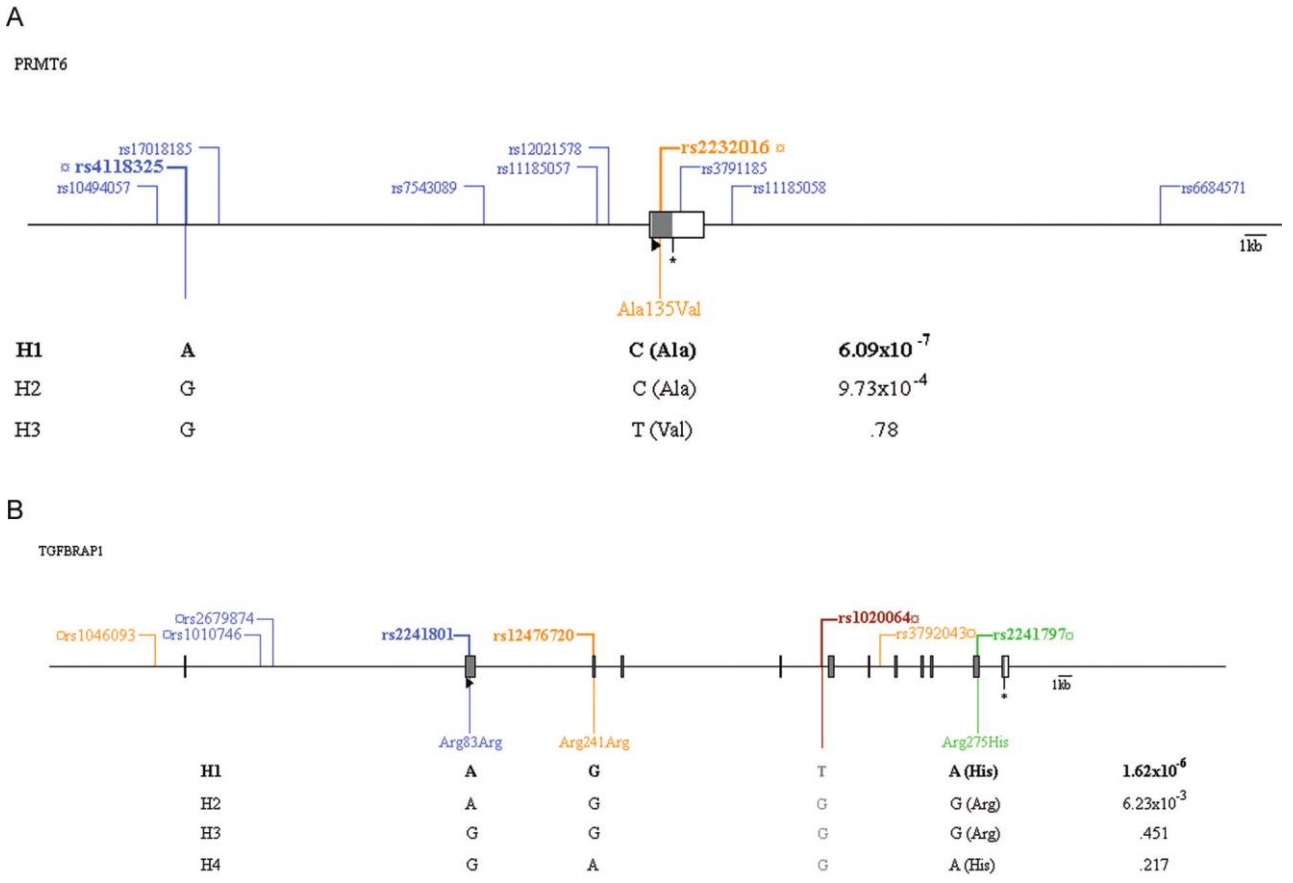

Figure 4. Haplotype maps for *PRMT6* (*A*) and *TGFBRAP1* (*B*). Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. Single-nucleotide polymorphisms (SNPs) in high linkage disequilibrium ($r^2 > 0.9$) are represented with the same color, and the genotyped tagSNPs are marked with the symbol ¤. For greater clarity, only the exonic polymorphisms in linkage disequilibrium with the genotyped tagSNPs ($r^2 > 0.9$) are specified in panel B. The numbers in the right column correspond to the *P* values obtained when comparing the allelic frequency of each haplotype between rapid progressors and control subjects.

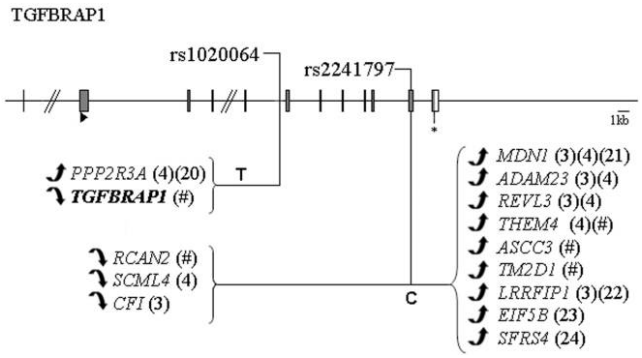

Figure 5. Correlation between some *TGFBRAP1* polymorphisms and differential gene expression according to the Genevar database [18] and the Dixon database [19]. Exons and untranslated regions are symbolized by shaded and unshaded boxes, respectively. The positions of the ATG and stop codons are indicated by a triangle (▶) and by an asterisk (*), respectively. The modulation of gene expression is indicated by the arrow direction: increased (↑) or decreased (↓) expression. The numbers correspond to the bibliographic references of previous works linking these genes to AIDS. The present genomewide association study of rapid progression is indicated by a pound sign (#).

In the past, we have often observed that haplotypes may be more informative than individual SNPs; this was notably shown for *CCR5* [7], *CXCR1* [6], and *HLA* [5]. We thus decided to explore the haplotype patterns for the genes exhibiting the best signals in our rapid progression GWAS, limiting the investigation to exonic and promoter SNPs (Figure 4). For *PRMT6*, we demonstrated that the effect of the *PRMT6* SNP rs4118325 could be tracking a haplotype (composed of rs4118325 and rs2232016 [Ala135Val]) through linkage disequilibrium ($r^2 = 0.12$, $D' = 1$) (Figure 4*A*). No differential messenger RNA (mRNA) expression could be significantly associated with either of these SNPs using the Genevar database [18] or the Dixon database [19].

For *TGFBRAP1*, the rs1020064 SNP was in high linkage disequilibrium ($r^2 = 0.97$) with a haplotype containing 3 exonic SNPs (rs2241801 [Arg83Arg], rs12476720 [Arg241Arg], and rs2241797 [Arg275His]) (Figure 4*B*). The SNP rs2241797 (Arg275His) appeared to be essential, since the signal disappeared when that SNP was removed from the haplotype ($P > 5.0 \times 10^{-2}$) and remained identical when either of the 2 synonymous SNPs were removed ($P = 1.62 \times 10^{-6}$). Interestingly, rs1020064 has been associated with differential expression of *TGFBRAP1* in the Genevar database and of *PPP2R3A* [20] in the Dixon database (Figure 5). More strikingly, in the Dixon database the nonsynonymous SNP rs2241797 was significantly associated with the differential expression of several proteins (Figure 5), among which 4 have been independently described to interact with HIV-1 (*MDN1* [21], *LRRFIP1* [22], *EIF5B* [23], and *SFRS4* [24]), and several have exhibited a positive association in an AIDS GWAS. Of note, no differential expression could be found for the 2 synonymous SNPs rs2241801 and rs12476720. For the *SOX5* rs1522232 and *RXRG* rs10800098 SNPs, no haplotype involving exonic or promoter SNPs and no differential mRNA expression could be found.

DISCUSSION

For the first time, the extreme HIV-1 rapid progression phenotype was specifically examined in a GWAS. The power of using the GRIV extreme design has been demonstrated by previous candidate gene studies [5–7, 25] and by the previous GWAS of nonprogressors [4]. Here we have identified 6 novel associations with rapid progression with ORs as high as 4, emphasizing the power of this extreme design in spite of a relatively modest sample size. As a comparison, there was only 1 independent ($r^2 < 0.5$) signal passing the FDR threshold of 25% in the GWAS of the nonprogressor GRIV cohort [4]. No replication was readily available, since this specific rapid progression design is rather unique in the world. The biological relevance of the rapid progression associations and their statistical validation through simulations with a second control group (<1% odds of finding as many independent signals) are nevertheless compelling. The use of an extreme population may provide a strong contrast and indeed help unravel new genetic factors, behaving as a magnifying glass [25, 26].

Four of the 6 SNPs associated with rapid progression to AIDS were clearly linked to a gene (distance, <2 kb). *SOX5* (OR, 0.45) encodes a transcription regulator notably expressed in lymph nodes and lymphoid tissues [27] and is also known to be involved in the transforming growth factor β (TGF-β)/SMAD chondrogenesis signaling pathway [28, 29]. There is no other experimental evidence linking this protein to the pathogenesis of HIV-1 infection. *RXRG* (OR, 3.29) encodes a retinoic acid nuclear receptor mediating the antiproliferative effects of retinoic acid and is known to repress HIV-1 transcription and replication [30, 31]. Several studies have also associated vitamin A deficiency with a bad prognosis in patients with AIDS, but the benefit of vitamin A supplementation in AIDS patients is still controversial [32, 33]. The genetic association for *PRMT6* (OR, 0.24) points toward a direct interaction between the host and the virus. Indeed, *PRMT6* codes for an arginine *N*-methyltransferase previously reported to methylate HIV-1 Tat and Rev, impairing their function [34, 35]. The product of *PRMT6* also methylates the high mobility group protein HMGA1, thereby modifying HMGA1 interactions with DNA [36], which could alter HIV-1 integration into the human genome [37]. The modulation of *PRMT6* transcription could thus affect HIV-1 integration or replication and prevent rapid progression. Finally, an association with rapid progression was discovered for *TGFBRAP1* (OR, 0.34). *TGFBRAP1* is expressed in most lymphoid tissues and is involved in the TGF-β signaling pathway [38]. TGF-β is a pleiotropic immunosuppressive cy-

tokine involved in immune homeostasis and in the differentiation of and balance between type 17 T helper (Th17) cells (T cells protecting the mucosal barrier integrity [39]) and regulatory T (T_{reg}) cells (T cells essential for immune suppression [40]). Interestingly, recent works have also supported a combined role for TGF-β and retinoic acid for T_{reg} cell differentiation [41, 42]. T_{reg} cells are rapidly induced after simian immunodeficiency virus (SIV) and HIV-1 infection [43, 44] and may have a deleterious effect during the chronic phase of infection [45, 46]. Moreover, an impairment in the balance between Th17 and T_{reg} cells during HIV-1 infection was recently described [40, 47, 48]. Finally, TGF-β can stimulate HIV-1 Tat transcription [49] and, reciprocally, be induced by Tat early during infection [50], supporting a key role for the intermingled effects of TGF-β and Tat in the early development of HIV-1 infection. Overall, our results support the central role played by the TGF-β pathway and the balance between Th17 and T_{reg} cells in AIDS progression. Interestingly, *PRMT6* and *TGFBRAP1* haplotypes involving promoter and exonic SNPs were also associated with disease progression; moreover, some of the haplotype SNPs could be associated with downstream differential mRNA expression (Figure 5).

In conclusion, as for all genetic studies, our results—including SNPs with higher *P* values, which were not discussed—will need to be confirmed by replication in other cohorts and by other investigations, such as fine gene mapping and biological experiments.

ANRS GENOMIC GROUP

The ANRS Genomic Group is composed of Prof Jean-François Delfraissy (Agence Nationale de Recherche sur le SIDA, Paris), Dr Laurence Meyer (Hôpital Kremin-Bicêtre, France), Prof Philippe Broët (Hôpital Kremlin-Bicêtre, France), Dr Cyril Dalmasso (Hôpital Kremlin-Bicêtre, France), Prof Patrice Debré (Hôpital La Salpétrière, Paris), Dr Ioannis Théodorou (Hôpital La Salpétrière, Paris), Prof Christine Rouzioux (Hôpital Necker, Paris), Cédric Coulonges (Conservatoire National des Arts et Métiers, Paris), Sigrid Le Clerc (Conservatoire National des Arts et Métiers, Paris), Sophie Limou (Conservatoire National des Arts et Métiers, Paris), and Prof Jean-François Zagury (Conservatoire National des Arts et Métiers, Paris).

Acknowledgments

We are grateful to all the patients and medical staff who have kindly collaborated with the GRIV project.

References

1. Kingsmore SF, Lindquist IE, Mudge J, Gessler DD, Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov **2008**; 7:221–30.
2. 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.
3. 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.
4. 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.
5. 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**; 170:1925–9.
6. Vasilescu A, Terashima Y, Enomoto M, et al. A haplotype of the human CXCR1 gene protective against rapid disease progression in HIV-1+ patients. Proc Natl Acad Sci U S A **2007**; 104:3354–9.
7. 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**; 37:1534–8.
8. Hercberg S, Galan P, Preziosi P, et al. Background and rationale behind the SU.VI.MAX Study, a prevention trial using nutritional doses of a combination of antioxidant vitamins and minerals to reduce cardiovascular diseases and cancers: SUpplementation en VItamines et Mineraux AntioXydants Study. Int J Vitam Nutr Res **1998**; 68:3–20.
9. 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. Rev Epidemiol Sante Publique **1996**; 44:373–5.
10. 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.
11. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics **2008**; 9:540.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B **1995**; 57:289–300.
13. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stati Med **1990**; 9:811–8.
14. Perneger TV. What's wrong with Bonferroni adjustments. BMJ **1998**; 316:1236–8.
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS **2003**; 100:9440–5.
16. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics **2000**; 155:945–59.
17. Devlin B, Roeder K. Genomic control for association studies. Biometrics **1999**; 55:997–1004.
18. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. Genome Res **2008**; 18:640–3.
19. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet **2007**; 39:1202–7.
20. Ammosova T, Washington K, Debebe Z, Brady J, Nekhai S. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-1 in Tat-activated HIV-1 transcription. Retrovirology **2005**; 2:47.
21. Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science **2008**; 319:921–6.
22. Wilson SA, Brown EC, Kingsman AJ, Kingsman SM. TRIP: a novel double stranded RNA binding protein which interacts with the leucine rich repeat of flightless I. Nucleic Acids Res **1998**; 26:3460–7.
23. Wilson SA, Sieiro-Vazquez C, Edwards NJ, et al. Cloning and characterization of hIF2, a human homologue of bacterial translation initiation factor 2, and its interaction with HIV-1 matrix. Biochem J **1999**; 342:97–103.
24. Exline CM, Feng Z, Stoltzfus CM. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol **2008**; 82:3921–31.
25. Huber C, Pons O, Hendel H, et al. Genomic studies in AIDS: problems

and answers—development of a statistical model integrating both longitudinal cohort studies and transversal observations of extreme cases. Biomed Pharmacother **2003**; 57:25–33.
26. Froguel P, Blakemore AI. The power of the extreme in elucidating obesity. N Engl J Med **2008**; 359:891–3.
27. Su AI, Cooke MP, Ching KA, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A **2002**; 99:4465–70.
28. Furumatsu T, Tsuda M, Taniguchi N, Tajima Y, Asahara H. Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. J Biol Chem **2005**; 280:8343–50.
29. Ikeda T, Kawaguchi H, Kamekura S, et al. Distinct roles of Sox5, Sox6, and Sox9 in different stages of chondrogenic differentiation. J Bone Miner Metab **2005**; 23:337–40.
30. Kiefer HL, Hanley TM, Marcello JE, Karthik AG, Viglianti GA. Retinoic acid inhibition of chromatin remodeling at the human immunodeficiency virus type 1 promoter: uncoupling of histone acetylation and chromatin remodeling. J Biol Chem **2004**; 279:43604–13.
31. Maeda Y, Yamaguchi T, Hijikata Y, et al. All-trans retinoic acid attacks reverse transcriptase resulting in inhibition of HIV-1 replication. Hematology **2007**; 12:263–6.
32. Austin J, Singhal N, Voigt R, et al. A community randomized controlled clinical trial of mixed carotenoids and micronutrient supplementation of patients with acquired immunodeficiency syndrome. Eur J Clin Nutr **2006**; 60:1266–76.
33. Mehta S, Fawzi W. Effects of vitamins, including vitamin A, on HIV/AIDS patients. Vitam Horm **2007**; 75:355–83.
34. Invernizzi CF, Xie B, Richard S, Wainberg MA. PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. Retrovirology **2006**; 3:93.
35. Xie B, Invernizzi CF, Richard S, Wainberg MA. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat interactions with both cyclin T1 and the Tat transactivation region. J Virol **2007**; 81:4226–34.
36. Sgarra R, Lee J, Tessari MA, et al. The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. J Biol Chem **2006**; 281: 3764–72.
37. Li L, Yoder K, Hansen MS, Olvera J, Miller MD, Bushman FD. Retroviral cDNA integration: stimulation by HMG I family proteins. J Virol **2000**; 74:10965–74.
38. Charng MJ, Zhang D, Kinnunen P, Schneider MD. A novel protein distinguishes between quiescent and activated forms of the type I transforming growth factor beta receptor. J Biol Chem **1998**; 273:9365–8.
39. Manel N, Unutmaz D, Littman DR. The differentiation of human T_H-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. Nat Immunol **2008**; 9:641–9.
40. de St Groth BF, Landay AL. Regulatory T cells in HIV infection: pathogenic or protective participants in the immune response? AIDS **2008**; 22:671–83.
41. Benson MJ, Pino-Lagos K, Rosemblatt M, Noelle RJ. All-trans retinoic acid mediates enhanced T reg cell growth, differentiation, and gut homing in the face of high levels of co-stimulation. J Exp Med **2007**; 204:1765–74.
42. Mucida D, Park Y, Kim G, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. Science **2007**; 317:256–60.
43. Estes JD, Wietgrefe S, Schacker T, et al. Simian immunodeficiency virus–induced lymphatic tissue fibrosis is mediated by transforming growth factor beta 1-positive regulatory T cells and begins in early infection. J Infect Dis **2007**; 195:551–61.
44. Kekow J, Wachsman W, McCutchan JA, Cronin M, Carson DA, Lotz M. Transforming growth factor beta and noncytopathic mechanisms of immunodeficiency in human immunodeficiency virus infection. Proc Natl Acad Sci U S A **1990**; 87:8321–5.
45. Kared H, Lelievre JD, Donkova-Petrini V, et al. HIV-specific regulatory T cells are associated with higher CD4 cell counts in primary infection. AIDS **2008**; 22:2451–60.
46. Weiss L, Donkova-Petrini V, Caccavelli L, Balbo M, Carbonneil C, Levy Y. Human immunodeficiency virus-driven expansion of CD4+CD25+ regulatory T cells, which suppress HIV-specific CD4 T-cell responses in HIV-infected patients. Blood **2004**; 104:3249–56.
47. Brenchley JM, Paiardini M, Knox KS, et al. Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections. Blood **2008**; 112:2826–35.
48. Favre D, Lederer S, Kanwar B, et al. Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. PLoS Pathog **2009**; 5:e1000295.
49. Li JM, Shen X, Hu PP, Wang XF. Transforming growth factor beta stimulates the human immunodeficiency virus 1 enhancer and requires NF-kappaB activity. Mol Cell Biol **1998**; 18:110–21.
50. Zocchi MR, Contini P, Alfano M, Poggi A. Pertussis toxin (PTX) B subunit and the nontoxic PTX mutant PT9K/129G inhibit Tat-induced TGF-beta production by NK cells and TGF-beta–mediated NK cell apoptosis. J Immunol **2005**; 174:6054–61.