

Exhaustive Genotyping of the *CEM15* (*APOBEC3G*) Gene and Absence of Association with AIDS Progression in a French Cohort

Hervé Do,^{1,3} Alexandre Vasilescu,³ Gora Diop,³ Thomas Hirtzig,³ Simon C. Heath,³ Cédric Coulonges,¹ Jay Rappaport,⁴ Amu Therwath,¹ Mark Lathrop,³ Fumihiko Matsuda,³ and Jean-François Zagury^{1,2,4}

¹Équipe Génomique, Bioinformatique et Pathologies du Système Immunitaire, INSERM EMI0355, and ²Chaire de Bioinformatique, Conservatoire National des Arts et Métiers, Paris, and ³Centre National de Génotypage, Evry, France; ⁴Center for Neurovirology, Temple University, Philadelphia, Pennsylvania

CEM15 (or APOBEC3G) has recently been identified as an inhibitor of human immunodeficiency virus type 1 (HIV-1) replication in vitro. To evaluate the impact of its genetic variations on the progression of acquired immunodeficiency syndrome (AIDS), we have performed an extensive genetic analysis of *CEM15*. We have sequenced *CEM15* in a cohort of 327 HIV-1-seropositive patients with extreme disease progression phenotypes—either slow progression or rapid progression—and in 446 healthy control subjects, all of white descent. We have identified 29 polymorphisms with allele frequencies >1%, 14 of which were newly characterized. There were no significant associations between the polymorphisms or haplotypes of *CEM15* and a disease progression phenotype in our cohort.

HIV-1, in addition to its structural and functional proteins (Gag, Pol, and Env), encodes several regulatory or auxiliary proteins—namely Tat, Rev, Nef, Vif, Vpr, and Vpu. The Vif-deficient virions can replicate in particular T cell lines (“permissive” cells) but not in others (“nonpermissive” cells) [1]. This observation was explained by the fact that Vif overcomes

an endogenous inhibitor of HIV-1 produced in nonpermissive cells [2]. Sheehy et al. identified this factor, CEM15 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G, or APOBEC3G; MIM 607113 in the Online Mendelian Inheritance in Man database [<http://www.ncbi.nlm.nih.gov/Omim/>]), and demonstrated that it has an inhibitory effect on the replication of Vif-deficient HIV-1 virions [3]. Like other members of the APOBEC family [4], CEM15 has a DNA mutator activity: it catalyzes the deamination of cytosine bases, transforming them into uracil bases, thus interfering with the viral replication cycle [5, 6]. The HIV-1 Vif protein counteracts the antiretroviral activity of CEM15 by inducing its ubiquitination and degradation [7, 8].

In the present study, we have performed a systematic genetic analysis of the *CEM15* gene by extensive nucleotide sequencing, to clarify whether genetic variations of this gene could influence disease onset and progression. We used the Genetics of Resistance to Immunodeficiency Virus (GRIV) cohort, which consists of 2 subpopulations of white French HIV-1-seropositive individuals with extreme phenotypes: 245 patients with slow progression, which are equivalent to a 1% extreme subset of patients in a cohort of 24,500 patients who experienced seroconversion [9], and 82 patients with rapid progression. We also included 446 healthy control subjects of similar ethnic origin. This cohort is, to our knowledge, the largest of its kind in the world and has made it possible to confirm and extend the major genetic associations described to date [10–12]. Indeed, transversal studies performed on patient cohorts with extreme phenotypes are complementary to longitudinal studies performed on seroconverter cohorts [9]. Single-nucleotide polymorphisms (SNPs) and other genetic variations identified through sequencing were evaluated for their association with disease susceptibility and progression.

Materials and methods. The GRIV cohort was established in 1995 in France to generate a large collection of DNAs for the identification and study of polymorphisms associated with rapid and slow progression to AIDS. Only white individuals of European descent living in France were recruited. These criteria limit the influence of the virogenetic and environmental factors (subjects are all infected by B strains and live in a similar environment) and put emphasis on the genetic background of

Table 1. Primers used to amplify the exons of *CEM15* by polymerase chain reaction (PCR).

The table is available in its entirety in the online edition of the *Journal of Infectious Diseases*.

Received 12 May 2004; accepted 4 August 2004; electronically published 15 December 2004.

Financial support: Agence Nationale de Recherche sur le SIDA, AIDS-Cancer Vaccine Development Foundation, and Neovacs SA (grants to the Genetics of Resistance to Immunodeficiency Virus project); Ministère de la Recherche et des Nouvelles Technologies (support to the Centre National de Génotypage).

Reprints or correspondence: Dr. Jean-François Zagury, Equipe Génomique, Bioinformatique et Pathologies du Système Immunitaire, INSERM EMI0355, 15 rue de l'École de Médecine, 75006 Paris, France (jz@ccr.jussieu.fr) or Fumihiko Matsuda, Centre National de Génotypage, 2 rue Gaston Crémieux, 91057 Evry Cedex, France (fumi@cng.fr).

The Journal of Infectious Diseases 2005;191:159–63

© 2004 by the Infectious Diseases Society of America. All rights reserved.
0022-1899/2005/19102-0003\$15.00

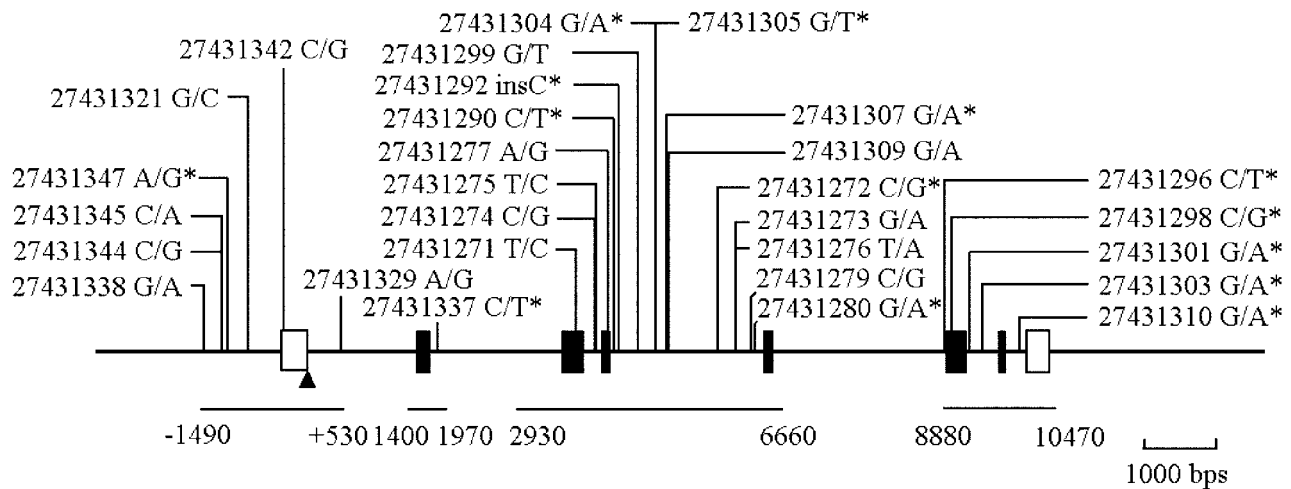


Figure 1. Genetic organization of the *CEM15* gene, located on chromosome 22 (22q13.1-q13.2). Coding and untranslated regions are indicated by black and white rectangles, respectively. The regions that have been sequenced are indicated by a horizontal line, with start and end positions numbered according to the first nucleotide of the initiation codon being considered as +1 (indicated by a black triangle). The polymorphism numbers are the attribution numbers from the Centre National de Génotypage database (the correspondence with the National Center for Biotechnology Information dbSNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) is given in table 2). The newly characterized polymorphisms are indicated by an asterisk. The genomic sequence used for alignment is GenBank sequence NT_011520.

each individual to determine the various patterns of progression. The GRIV cohort consists of 2 subpopulations of HIV-1-seropositive individuals with extreme phenotypes of disease progression: 245 patients with slow progression, defined as individuals who have been seropositive and asymptomatic for ≥ 8 years with a $CD4^+$ cell count >500 cells/mm³ in the absence of antiretroviral therapy, and 82 patients with rapid progression, defined as patients with a decrease in their $CD4^+$ cell count to <300 cells/mm³ in <3 years after the last seronegative test. The DNA was obtained from fresh peripheral blood mononuclear cells or from Epstein-Barr virus-transformed cell lines.

Polymerase chain reaction conditions and primers used to genotype the *CEM15* gene are presented in table 1. Alignment, SNP discovery, and genotyping were performed with the software Genalys (version 2.0b; M. Takahashi, Centre National de Génotypage) [13]. For all polymorphisms, we performed an initial screening of 150 patients with slow progression, 50 patients with rapid progression, and 155 control subjects. If the *P* value evaluating the association was of borderline significance (.05–.10), the genotyping was then extended to the rest of the cohort.

Statistical analyses were performed only on the polymor-

phisms with minor-allele frequencies $>1\%$ in the whole population. Differences in the allele frequencies of individual polymorphisms between the 3 groups were examined using a Fisher's exact test. Linkage disequilibrium (LD) was computed for each pair of polymorphisms. Haplotype estimates were obtained using the expectation maximization (EM) algorithm [14] and the Phase2 algorithm [15], either for all polymorphisms or for polymorphisms gathered in linkage groups (to determine haploblocks). For each haplotype (with a frequency $>1\%$ in the whole population), the expected numbers of individuals in each group with and without that haplotype were computed from the estimated marginal haplotype probability distribution for each individual. These numbers were rounded to the nearest integer, and a nominal *P* value was computed using Fisher's exact test.

Results. We systematically screened the *CEM15* gene for polymorphisms by resequencing the exons with their flanking regions as well as the 1-kb region upstream of the first exon, under the assumption that putative regulatory elements would be located within that region. Of these 42 polymorphisms identified in the *CEM15* gene, 29 had allele frequencies of $\geq 1\%$ in the whole population (figure 1). Among these 29 polymorphisms, 14 SNPs were newly identified in the present study. As shown in figure 1, 5 SNPs are located in the 1-kb region upstream of the first exon, and 4 SNPs are located in exons. The SNP *CEM15_27431271* leads to a synonymous change, whereas the SNPs *CEM15_27431277* and *CEM15_27431298* lead to nonsynonymous changes (histidine to arginine at position 186 and glutamine to glutamic acid at position 275, respectively, in

The figure is available in its entirety in the online edition of the *Journal of Infectious Diseases*.

Figure 2. Linkage disequilibrium (LD) plot of the *CEM15* gene. The color figure and legend are available in their entirety in the online edition of the *Journal of Infectious Diseases*.

Table 2. List of all polymorphisms with frequencies >1% in the whole population.

Polymorphism ^a	Allele ^b		Genotyped individuals, no.			Frequency of A1 allele, %			Nominal <i>P</i> value for test statistics ^c		NCBI dbSNP ID
	A1	A2	CTR	RP	SP	CTR	RP	SP	SP vs. RP	SP vs. RP vs. CTR	
CEM15_27431338	G	A	56	30	72	40.18	33.33	36.11	.75	.64	rs6519165
CEM15_27431344	C	G	64	26	53	39.84	40.38	42.45	.86	.91	rs7291971
CEM15_27431345	C	A	64	24	51	31.25	33.33	35.29	.86	.81	rs6519166
CEM15_27431347	A	G	130	37	109	95.38	94.59	94.50	1.00	.86	ss28474749 ^d
CEM15_27431321	G	C	93	27	70	4.30	3.70	4.29	1.00	1.00	rs5757463
CEM15_27431342	C	G	92	31	52	28.26	29.03	23.08	.46	.61	rs5750743
CEM15_27431329	A	G	68	36	102	91.18	95.83	93.63	.77	.46	rs12160242
CEM15_27431337	C	T	129	41	116	96.51	96.34	96.12	1.00	.95	ss28474750 ^d
CEM15_27431271	T	C	146	49	136	50.68	51.02	49.26	.81	.94	rs5757465
CEM15_27431274	C	G	129	31	111	96.51	96.77	98.20	.61	.64	rs6001417
CEM15_27431275	T	C	129	31	111	96.51	96.77	98.20	.61	.64	rs3736685
CEM15_27431277	A	G	155	46	141	96.77	96.74	97.87	.69	.64	rs8177832
CEM15_27431290	C	T	137	38	127	97.08	93.42	96.85	.19	.29	ss28474751 ^d
CEM15_27431292	–	C	149	44	133	93.29	94.32	93.23	1.00	.89	ss28474752 ^d
CEM15_27431299	G	T	75	39	110	97.33	96.15	97.73	.44	.68	rs10427620
CEM15_27431304	G	A	119	43	121	98.32	97.67	97.11	1.00	.70	ss28474753 ^d
CEM15_27431305	G	T	119	43	121	98.32	97.67	97.11	1.00	.70	ss28474754 ^d
CEM15_27431307	G	A	147	47	144	96.94	97.87	97.22	1.00	1.00	ss28474755 ^d
CEM15_27431309	G	A	147	47	144	96.94	96.81	97.92	.69	.76	rs6001418
CEM15_27431272	C	G	134	40	116	92.54	95.00	94.83	1.00	.54	ss28474756 ^d
CEM15_27431273	G	A	147	43	121	93.20	94.19	94.63	1.00	.82	rs12158985
CEM15_27431276	T	A	446	82	243	34.30	32.93	35.60	.57	.82	rs2294366
CEM15_27431279	C	G	134	73	220	41.79	43.84	42.27	.77	.91	rs2294367
CEM15_27431280	G	A	134	74	224	98.51	96.62	97.32	.78	.43	ss28474757 ^d
CEM15_27431296	C	T	148	81	234	93.92	93.83	94.23	.85	.95	ss28474758 ^d
CEM15_27431298	C	G	150	81	236	94.00	93.83	94.28	.85	.95	ss28474759 ^d
CEM15_27431301	G	A	138	49	124	93.84	92.86	93.15	1.00	.93	ss28474760 ^d
CEM15_27431303	G	A	104	39	58	88.46	88.46	89.66	.82	.95	ss28474761 ^d
CEM15_27431310	G	A	151	80	245	99.01	97.50	98.98	.23	.33	ss28474762 ^d

NOTE. CTR, control population; NCBI, National Center for Biotechnology Information; RP, population with rapid progression; SP, population with slow progression.

^a The genotyping of SNPs CEM15_27431276, 27431279, 27431280, 27431296, 27431298, and 27431310 has been extended, because they presented borderline *P* values in the initial screening of 150 patients with slow progression, 50 patients with rapid progression, and 155 control subjects.

^b A1 represents the allele that is identical to the reference sequence (GenBank accession no. NT_011520.10).

^c Fisher's exact test. Bonferroni corrections were not performed.

^d New variant.

protein sequence NP_068594.1). The SNP CEM15_27431342 is located in the 5' untranslated region. The polymorphism distributions in cases and controls were in Hardy-Weinberg equilibrium. Figure 2 shows pairwise LD measured between each pair of polymorphisms.

We performed a statistical analysis to test whether these polymorphisms or deduced haplotypes are associated with disease progression. We failed to obtain any significant association between polymorphisms and AIDS progression by use of Fisher's exact test (the lowest *P* value was 0.19, for CEM15_27431290, slow progression vs. rapid progression; table 2). Similarly, using estimated haplotypes, we did not find any significant association between particular haplotypes of *CEM15* and AIDS progression

(table 3). The 2 algorithms, Phase2 and EM, yielded similar haplotype estimates. The detailed list of haplotypes is given in figure 3. The lowest *P* value was .15, for haplotype 12, in the comparison between rapid progression and slow progression (table 3). No association was identified either in the dominant/recessive models or in the simple comparison of the diplotype distribution (data not shown).

To refine our haplotype analysis, we examined the distributions of haploblocks, which were computed by gathering the polymorphisms in linkage groups for various levels of LD (85%, 90%, and 95%). No association could be identified by use of this approach. Finally, we looked more specifically for associations based on the haplotypes/haploblocks derived solely from

Table 3. List of all haplotypes with frequencies >1% in the whole population.

Haplotype	Chromosomes, no.			Haplotype frequency, %			Nominal <i>P</i> value for test statistics ^a	
	CTR	RP	SP	CTR	RP	SP	SP vs. RP	SP vs. RP vs. CTR
HAPLO_1	113	36	109	40.07	43.90	41.92	.80	.80
HAPLO_2	70	19	57	24.82	23.17	21.92	.88	.73
HAPLO_3	25	5	30	8.87	6.10	11.54	.21	.29
HAPLO_4	18	5	13	6.38	6.10	5.00	.78	.78
HAPLO_5	12	4	11	4.26	4.88	4.23	.76	.97
HAPLO_6	9	2	6	3.19	2.44	2.31	1.00	.81
HAPLO_7	10	3	6	3.55	3.66	2.31	.45	.66
HAPLO_8	7	0	7	2.48	0.00	2.69	.20	.73
HAPLO_9	8	1	5	2.84	1.22	1.92	1.00	.62
HAPLO_10	4	3	6	1.42	3.66	2.31	.45	.43
HAPLO_11	3	1	7	1.06	1.22	2.69	.69	.33
HAPLO_12	3	3	3	1.06	3.66	1.15	.15	.19

NOTE. Estimation by the Phase2 and expectation maximization algorithms yielded similar results. The detailed description of the haplotypes and haploblocks (with linkage disequilibrium >95%) in relation to the polymorphisms is given in figure 3. CTR, control population; RP, population with rapid progression; SP, population with slow progression.

^a Fisher's exact test. Bonferroni corrections were not performed.

the SNPs located in the 1-kb region upstream of exon 1 (assumed to be the promoter region) or solely from the 2 non-synonymous SNPs (amino acid substitutions). Using these 2 approaches, we failed to obtain any significant association between AIDS progression and polymorphisms (data not shown).

Discussion. In this study, we have undertaken a systematic investigation of the association between genetic variations in the *CEM15* gene and susceptibility to disease progression in the GRIV cohort. We have identified 29 polymorphisms with minor-allele frequencies >1% in our whole population, 14 of which were newly characterized. Two SNPs of the 29 polymorphisms, which are located in exons, lead to amino acid substitutions. However, these substitutions (His186Arg for CEM15_27431277 and Gln275Glu for CEM15_27431298) involve amino acids with similar properties that are unlikely to modify the protein function. We failed to obtain any significant association between polymorphisms, haplotypes, or haploblocks in the *CEM15* gene and extreme AIDS progression phenotypes in the GRIV cohort. Therefore, it seems that there is no major genetic determinant in the *CEM15* gene influencing the various profiles of progression to AIDS. As for any AIDS genomic study, our results will need to be confirmed in other

cohorts, especially in seroconverter cohorts, which have a structure different from that of the GRIV cohort [9]. Since *CEM15* is known to play a relevant role in HIV-1 replication [3], it will also be of interest to investigate the other cellular proteins, such as Cullin-5 or Rbx-1, that participate with Vif in its ubiquitination and proteolysis [8]. The gene polymorphisms of these cellular proteins could potentially exhibit associations in combination with the *CEM15* polymorphisms, if not on their own. Finally, *CEM15*, because of its activity on single-stranded DNA, has been suspected to play a potential role in innate antiretroviral defense [6]. Therefore, the identification of 14 novel polymorphisms with relatively high frequencies (>1%) and haplotype information on this gene should prove useful in assessing the potential association of *CEM15* polymorphisms with other retroviral diseases.

Acknowledgments

The authors are grateful to all of the patients and medical staff who have generously collaborated with the Genetics of Resistance to Immunodeficiency Virus project. The authors also thank the Epidemiological study on the Genetics and Environment of Asthma (EGEA) cooperative group, for having given access to data on the EGEA study, which was partly supported by an INSERM/Merck Sharp & Dohme–Chibret convention.

References

- Gabuzda DH, Lawrence K, Langhoff E, et al. Role of vif in replication of human immunodeficiency virus type 1 in CD4⁺ T lymphocytes. *J Virol* **1992**; 66:6489–95.
- Simon JH, Gaddis NC, Fouchier RA, Malim MH. Evidence for a newly discovered cellular anti-HIV-1 phenotype. *Nat Med* **1998**; 4:1397–400.
- Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human

The figure is available in its entirety in the online edition of the *Journal of Infectious Diseases*.

Figure 3. List of all haplotypes with frequencies >1% in the whole population (similar results were obtained with the Phase2 and expectation maximization algorithms). The color figure and legend are available in its entirety in the online edition of the *Journal of Infectious Diseases*.

- gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **2002**;418:646–50.
4. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **2002**;10:1247–53.
 5. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **2003**;424:99–103.
 6. Harris RS, Bishop KN, Sheehy AM, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* **2003**;113:803–9.
 7. Sheehy AM, Gaddis NC, Malim MH. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **2003**;9:1404–7.
 8. Yu X, Yu Y, Liu B, et al. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **2003**;302:1056–60.
 9. 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.
 10. Rappaport J, Cho YY, Hendel H, Schwartz EJ, Schachter F, Zagury JE. 32 bp CCR-5 gene deletion and resistance to fast progression in HIV-1 infected heterozygotes. *Lancet* **1997**;349:922–3.
 11. Vasilescu A, Heath SC, Ivanova R, et al. Genomic analysis of Th1-Th2 cytokine genes in an AIDS cohort: identification of IL4 and IL10 haplotypes associated with the disease progression. *Genes Immun* **2003**;4:441–9.
 12. 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.
 13. Takahashi M, Matsuda F, Margetic N, Lathrop M. Automated identification of single nucleotide polymorphisms from sequencing data. *J Bioinform Comput Biol* **2003**;1:253–65.
 14. Laird N. The EM algorithm. In: Rao CR, ed. *Handbook of statistics, vol 9: computational statistics*. Amsterdam: Elsevier Science Publishers B.V., **1993**:509–20.
 15. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **2001**;68:978–89.