Article

A whole-genome association study of major determinants for host control of HIV-1

FELLAY, Jacques, *et al*.

Abstract

Understanding why some people establish and maintain effective control of HIV-1 and others do not is a priority in the effort to develop new treatments for HIV/AIDS. Using a whole-genome association strategy, we identified polymorphisms that explain nearly 15% of the variation among individuals in viral load during the asymptomatic set-point period of infection. One of these is found within an endogenous retroviral element and is associated with major histocompatibility allele human leukocyte antigen (HLA)-B*5701, whereas a second is located near the HLA-C gene. An additional analysis of the time to HIV disease progression implicated two genes, one of which encodes an RNA polymerase I subunit. These findings emphasize the importance of studying human genetic variation as a guide to combating infectious agents.

Reference

UNIVERSITÉ DE GENÈVE

A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1

Jacques Fellay[1], **Kevin V. Shianna**[2,*], **Dongliang Ge**[1,*], **Sara Colombo**[3,*], **Bruno Ledergerber**[4,*], **Mike Weale**[1,*], **Kunlin Zhang**[3], **Curtis Gumbs**[1], **Antonella Castagna**[5], **Andrea Cossarizza**[6], **Alessandro Cozzi-Lepri**[7], **Andrea De Luca**[8], **Philippa Easterbrook**[9], **Patrick Francioli**[10], **Simon Mallal**[11], **Javier Martinez-Picado**[12], **José M. Miro**[13], **Niels Obel**[14], **Jason P. Smith**[2], **Josiane Wyniger**[3], **Patrick Descombes**[15], **Stylianos E. Antonarakis**[16], **Norman L. Letvin**[17], **Andrew J. McMichael**[18], **Barton F. Haynes**[19], **Amalio Telenti**[3,†], and **David B. Goldstein**[1,†,‡]

[1]*Center for Population Genomics and Pharmacogenetics, Duke Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710, USA* [2]*Duke Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710, USA* [3]*Institute of Microbiology, University Hospital Center; and University of Lausanne, 1011 Lausanne, Switzerland* [4]*Division of Infectious Diseases, University Hospital, 8091 Zürich, Switzerland* [5]*Clinic of Infectious Diseases, Istituto di Ricovero e Cura a Carattere Scientifico, San Raffaele Hospital, 20127 Milan, Italy* [6]*Department of Biomedical Sciences, Section of General Pathology, University of Modena and Reggio Emilia, School of Medicine, 41100 Modena, Italy* [7]*Department of Primary Care and Population Sciences, Royal Free and University College Medical School, University College London, London NW3 2PF, UK* [8]*Institute of Clinical Infectious Diseases, Catholic University of the Sacred Heart, 00168 Rome, Italy* [9]*Academic Department of HIV and Genitourinary Medicine, Kings College London, at Guy's, King's, and St. Thomas' Hospitals, London SE5 9RJ, UK* [10]*Service of Infectious Diseases, Department of Medicine and Service of Hospital Preventive Medicine, University Hospital Center, 1011 Lausanne, Switzerland* [11]*Centre for Clinical Immunology and Biomedical Statistics, Royal Perth Hospital; and Murdoch University, Perth, WA 6000, Australia* [12]*irsiCaixa Foundation and Hospital Germans Trias i Pujol, 08916 Badalona, Spain; and Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain* [13]*Hospital Clinic–Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, 08036 Barcelona, Spain* [14]*Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, 2100 Copenhagen, Denmark* [15]*Genomics Platform, National Centre of Competence in Research "Frontiers in Genetics," University of Geneva, 1211 Geneva, Switzerland* [16]*Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland* [17]*Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA* [18]*Medical Research Council Human Immunology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK* [19]*Duke Human Vaccine Institute, Duke University, Durham, NC 27710, USA*

Abstract

Understanding why some people establish and maintain effective control of HIV-1 and others do not is a priority in the effort to develop new treatments for HIV/AIDS. Using a whole-genome association strategy, we identified polymorphisms that explain nearly 15% of the variation among individuals in viral load during the asymptomatic set-point period of infection. One of these is found within an endogenous retroviral element and is associated with major histocompatibility allele *human leukocyte antigen (HLA)–B*5701*, whereas a second is located near the *HLA-C* gene. An additional analysis

of the time to HIV disease progression implicated two genes, one of which encodes an RNA polymerase I subunit. These findings emphasize the importance of studying human genetic variation as a guide to combating infectious agents.

Humans show remarkable variation in vulnerability to infection by HIV-1 and especially in the clinical outcome after infection. One striking and largely unexplained difference is the level of circulating virus in the plasma during the nonsymptomatic phase preceding the progression to AIDS. This is known as the viral set point and can vary among individuals by as much as 4 to 5 logs (1–6). We aimed to identify human genetic differences that influence this variation.

To define a homogeneous phenotype for genetic analyses, a consortium of nine cohorts was formed [termed Euro-CHAVI (Center for HIV/AIDS Vaccine Immunology) (7)], and a total of 30,000 patients were screened to identify those most appropriate for analysis. All longitudinal viral-load (VL) data were assessed through a computerized algorithm to eliminate VL not reflecting the steady state and were individually inspected by an experienced infectious-disease clinician (Fellay) to exclude suspicious VL data and patients that do not show a clear set point, leaving 486 patients with a consistent and accurately measured phenotype (7). For patients with at least four CD4 cell-count results, we defined a progression phenotype as the time to treatment initiation or to the predicted or observed drop of the CD4 cell count below 350 (7,8).

All samples were genotyped with the use of Illumina's HumanHap550 BeadChip with 555,352 single-nucleotide polymorphisms (SNPs). A series of quality-control steps resulted in the elimination of 20,251 polymorphisms (7). We applied methods to identify deletions and targeted copy-number variations and to assess whether they influenced the phenotype (7). Our core association analyses focused on single-marker genotype-trend tests of the quality control–passed SNPs, using linear regression (7). To control for the possibility of spurious associations resulting from population stratification, we used a modified EIGENSTRAT method (7,9). We assessed significance with a Bonferroni correction (P cutoff = 9.3×10^{-8}). Analyses incorporating *human leukocyte antigen* (*HLA*) typing were carried out on a subgroup of 187 patients with available four-digit *HLA* class I allelic determination.

These analyses identified two independently acting groups of polymorphisms, associated with *HLA* loci *B* and *C*, that are estimated to explain 9.6 and 6.5% of the total variation in HIV-1 set point, respectively, and can thus be considered as major genetic determinants of viral set point. A third set located >1 Mb away in the major histocompatibility complex upstream of a gene that encodes an RNA polymerase I subunit explains 5.8% of the total variation in disease progression. Together, the three polymorphisms explain 14.1% of the variation in HIV-1 set point.

One polymorphism located in the *HLA* complex *P5* (*HCP5*) gene explains 9.6% of the total variation in set point, despite a minor-allele frequency of 0.05 (Single Nucleotide Polymorphism database number rs2395029, $P = 9.36 \times 10^{-12}$). A single copy of the controlling allele was found to result in a reduction in VL of >1 log (Fig. 1); at $P = 9.36 \times 10^{-12}$, this genome-wide association is significant.

The *HCP5* gene is located 100 kb centromeric from *HLA-B* on chromosome 6 (Fig. 2), and the associated variant is known to be in high linkage disequilibrium (LD) with the *HLA* allele *B*5701* (10) ($r^2 = 1$ in our data set). This allele itself has the strongest-described protective impact on HIV-1 disease progression (11) and has been associated with low VL (12).

Given the strong functional data supporting a role for *HLA-B*5701* in restricting HIV-1, our first hypothesis is that the association observed here is due to the effect of *HLA-B*5701*,

reflected in its tagging a SNP within *HCP5* (10). We emphasize, however, that genetics allows no resolution of whether this effect is exclusively due to *B*5701* or if *HCP5* variation also contributes to the control. In fact, as a human endogenous retroviral element (HERV) with sequence homology to retroviral *pol* genes (13) and confirmed expression in lymphocytes (14), *HCP5* is itself a good candidate to interact with HIV-1, possibly through an antisense mechanism (14). Moreover, *HCP5* is predicted to encode two proteins, and the associated polymorphism results in an amino acid substitution in one of these proteins.

A model in which *HCP5* and *HLA-B*5701* have a combined haplotypic effect on the HIV-1 set point is consistent with the observation that suppression of viremia can be maintained in *B*5701* patients with undetectable VL, even after HIV-1 undergoes mutations that allow escape from cytotoxic T lymphocyte (CTL)–mediated restriction (15). However, this observation has also been explained by a decrease in viral fitness associated with the escape variants (16). In addition, *B*5701* patients present less frequently with symptoms during acute HIV-1 infection (12), suggesting control before the time of a maximal CTL response (17).

The second most significant polymorphism we identified, rs9264942, is located in the 5′ region of the *HLA-C* gene, 35 kb away from transcription initiation (Fig. 2) and 156 kb telomeric of the *HCP5* gene. This SNP explains 6.5% of the variation in set point (Fig. 1) and shows a genome-wide significant association ($P = 3.77 \times 10^{-9}$). Despite minor LD between the *HCP5* and *HLA-C* SNPs ($r^2 = 0.05$, $D' = 0.84$), nested regression models clearly demonstrate an independent effect of each of these variants. In a model including the *HCP5* variant, the addition of the *HLA-C* variant results in a highly significant increase in explanatory power, as does the addition of the *HCP5* variant to a model already including the *HLA-C* variant [supporting online material (SOM) text].

This SNP also associates strongly with differences in *HLA-C* expression levels, both in the Sanger Institute Genevar expression database (18) (table S1) and in a replication group of 48 healthy volunteers established for this study (SOM text). The protective allele leads to a lower VL and is associated with higher expression of the *HLA-C* gene. This strong and independent association with *HLA-C* expression levels suggests that genetic control of expression levels of a classical *HLA* gene influences viral control. Other *HLA-C* 5′ variants also associate with *HLA-C* expression but do not contribute independently to viral control (SOM text).

Although these data make a strong case for a causal role for *HLA-C* expression levels, extensive LD throughout the MHC region makes it necessary to directly test for alternative causal variants. Specifically, we used nested regression models to assess whether the observed association could be determined by described functional *HLA* class I alleles. In fact, the *HLA-C* expression SNP shows association with certain alleles or group of alleles (Tables 1 and 2). In each case, however, although the *HLA-C* expression variant can explain the effect of these alleles on the HIV-1 set point, the reverse is not true. When a linear regression model includes known *HLA* alleles, the addition of rs9264942 results in a significant increase in the explained variation. On the other hand, none of the *HLA* alleles considered, with the exception of *HCP5/B*5701*, adds significantly to a model that already incorporates the *HLA-C* variant (Tables 1 and 2).

No other single marker reached genome significance, and none of the identified copy-number variations (7) showed any association with the HIV-1 set point. An analysis comparing the observed set of *P* values to that expected under the null hypotheses shows no overall inflation of *P* values (indicating little contribution from population stratification) but does show an excess of low *P* values, beginning with the 355th most associated SNP (fig. S1). This indicates that additional real effects are likely to be present among the most associated polymorphisms in this study (complete list available in table S2). Potentially interesting candidates with a lesser

association with the set point (listed in tables S3 to S5) were chosen on the basis of their ranking in the study or their link with HIV-1 biology.

We next identified polymorphisms that associate with progression rather than VL: The strongest association included a set of seven polymorphisms located in and near the *ring finger protein 39* (*RNF39*) and *zinc ribbon domain–containing 1* (*ZNRD1*) genes, respectively (rs9261174, rs3869068, rs2074480, rs7758512, rs9261129, rs2301753, and rs2074479). This group of polymorphisms explains 5.8% of the variation in progression, with a relative hazard of 0.64 (fig. S2), and approaches genome-wide significance ($P = 3.89 \times 10^{-7}$). It also associates with VL at the set point ($P = 7.11 \times 10^{-3}$). These variants are >1 Mb telomeric from the previous candidates (fig. S3), and their effect on both progression and set point is independent of *HCP5*- and *HLA-C*–related polymorphisms and *HLA* alleles or groups of alleles previously implicated in HIV-1 control (SOM text and table S6).

Using the Genevar database and our group of 48 healthy volunteers, we observed that *ZNRD1* expression is significantly associated with the identified SNPs (SOM text and table S1). Two of them (rs3869068 and rs9261174) are located in a putative regulatory 5′ region, 25 and 32 kb away from the gene, respectively. Because *ZNRD1* encodes an RNA polymerase I subunit, if this gene is responsible for the restriction of HIV-1, the mechanism could involve interference with the processing of HIV-1 transcripts by the HIV-1 regulatory protein Rev. Rev is known to be localized in the nucleolus (where RNA polymerase I transcribes ribosomal RNA), and the blockade of RNA polymerase I has been shown to influence the distribution of REV, causing a shift from the nucleolus to the cytoplasm (19,20). Efficiency in provirus transcription is highly variable among individuals; in one study, differences in transcription efficiency accounted for 64 to 83% of the total variance in virus production that was attributable to post-entry cellular factors (21).

The second gene, *RNF39*, is poorly characterized but cannot be ruled out as a candidate because two of the associated polymorphisms are located in its coding region and result in amino acid changes (rs2301753 and rs2074479). No other genome-wide significant association was observed in the analysis of the progression phenotype (SOM text).

We established an independent replication cohort of 140 Caucasian patients, drawn from the same participating cohorts. For this follow-up study, we relaxed the interval from a documented negative test to a positive test (for infection) from 2 years to 4 years to identify additional qualifying study participants. We genotyped representative polymorphisms for the associations reported above (*HCP5*, rs2395029; *HLA-C*, rs9264942; and *ZNRD1*, rs9261174). Each association was replicated with effects all in the same direction: *HCP5*, $P = 1.4 \times 10^{-2}$; *HLA-C*, $P = 2.8 \times 10^{-3}$; and *ZNRD1*, $P = 4.8 \times 10^{-2}$.

We have securely identified at least two mechanisms not previously known to restrict HIV-1: *HLA-C*, which has been suspected but never confirmed to contribute to HIV-1 control, and an RNA polymerase subunit that substantially changes the time course of HIV progression (fig. S1). We also suggest the possibility that a HERV-derived gene may contribute to the viral control attributed to the *HLA-B*5701* allele. Our findings confirm and emphasize the central role of the MHC region in HIV-1 restriction, estimate its contribution against all genome influences, and open up new perspectives in the understanding of its mode of action: It is necessary to expand *HLA* analysis to include high-density genotyping. It is also noteworthy that this genome-wide study of host determinants has three clear discoveries, implying that determinants of host response may often include gene variants with major effects. This suggests a degree of urgency in carrying out similar studies for other infectious diseases.

Our results suggest two possible directions for therapeutic intervention. First, if *HCP5* and *ZNRD1* contribute to the control associated with *HLA-B*5701*, they could lead to therapeutic

applications. On the other hand, the implication of *HLA-C* in HIV-1 control could present important opportunities, given that the HIV-1 accessory protein Nef selectively down-regulates the expression of *HLA-A* and *-B* but not that of *HLA-C* on the surface of infected cells (22). Originally, this strategy was considered advantageous for the virus because *HLA-A* and *-B* present foreign (notably viral) epitopes to CD8 T cells, resulting in cell destruction, whereas *HLA-C* binds self-peptides and interacts with natural killer (NK) cells to avoid NK attack. However, *HLA-C* also has the ability to present viral peptides to cytotoxic CD8[+] T cells and consequently to restrict HIV-1 (23,24). Our observations suggest that *HLA-C*–mediated restriction may be an important element of viral control in specific genetic backgrounds, and that the apparent immunity of *HLA-C* to Nef down-regulation could present an opportunity for vaccine strategies targeting *HLA-C*–mediated responses.

References and Notes

1. Telenti A, Goldstein DB. Nat Rev Microbiol 2006;4:865. [PubMed: 17041633]

2. O'Brien SJ, Nelson GW. Nat Genet 2004;36:565. [PubMed: 15167933]

3. Telenti A, Bleiber G. Future Virol 2006;1:55.

4. Carrington M, O'Brien SJ. Annu Rev Med 2003;54:535. [PubMed: 12525683]

5. Nelson GW, O'Brien SJ. J Acquired Immune Defic Syndr 2006;42:347. [PubMed: 16763524]

6. Bleiber G, et al. J Virol 2005;79:12674. [PubMed: 16188970]

7. Materials and methods are available as supporting material on *Science* Online.

8. Douek DC, Picker LJ, Koup RA. Annu Rev Immunol 2003;21:265. [PubMed: 12524385]

9. Price AL, et al. Nat Genet 2006;38:904. [PubMed: 16862161]

10. de Bakker PI, et al. Nat Genet 2006;38:1166. [PubMed: 16998491]

11. Migueles SA, et al. Proc Natl Acad Sci USA 2000;97:2709. [PubMed: 10694578]

12. Altfeld M, et al. AIDS 2003;17:2581. [PubMed: 14685052]

13. Kulski JK, Dawkins RL. Immunogenetics 1999;49:404. [PubMed: 10199916]

14. Vernet C, et al. Immunogenetics 1993;38:47. [PubMed: 8462994]

15. Bailey JR, et al. J Exp Med 2006;203:1357. [PubMed: 16682496]

16. Martinez-Picado J, et al. J Virol 2006;80:3617. [PubMed: 16537629]

17. McMichael AJ, Rowland-Jones SL. Nature 2001;410:980. [PubMed: 11309628]

18. Stranger BE, et al. PLoS Genet 2005;1:e78. [PubMed: 16362079]

19. Michienzi A, et al. Proc Natl Acad Sci USA 2000;97:8955. [PubMed: 10922055]

20. Dundr M, et al. J Cell Sci 1995;108:2811. [PubMed: 7593322]

21. Ciuffi A, et al. J Virol 2004;78:10747. [PubMed: 15367641]

22. Cohen GB, et al. Immunity 1999;10:661. [PubMed: 10403641]

23. Goulder PJ, et al. AIDS 1997;11:1884. [PubMed: 9412710]

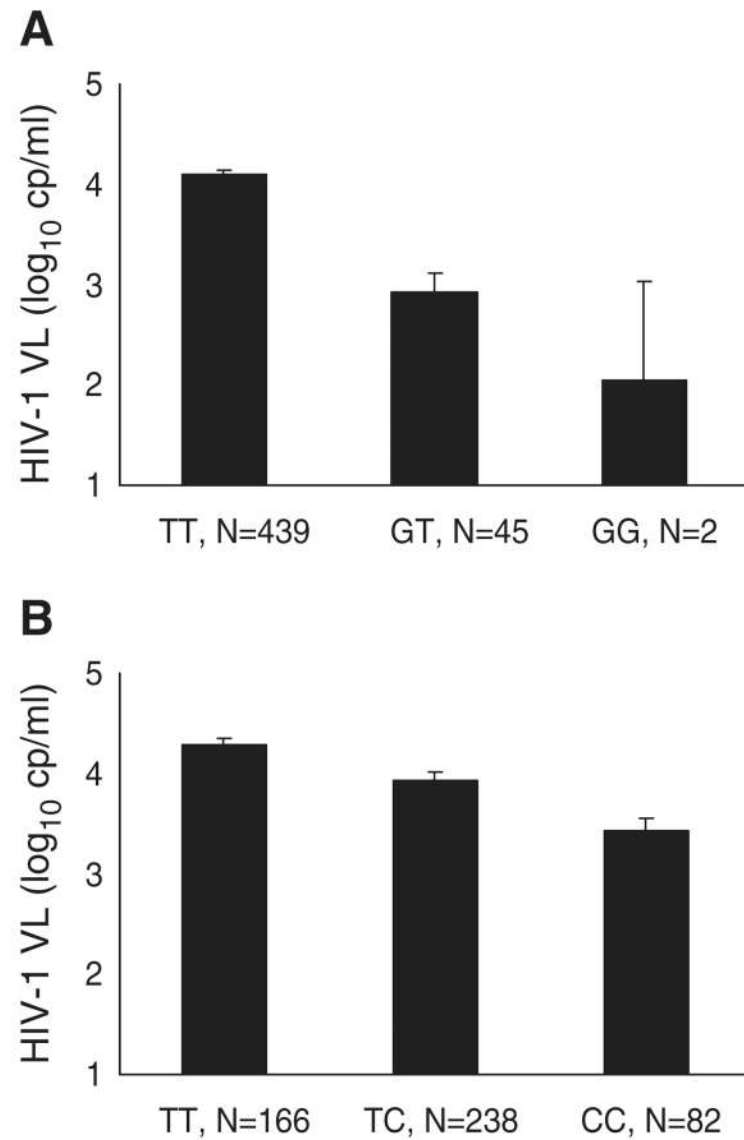24. Adnan S, et al. Blood 2006;108:3414. [PubMed: 16882705]

Fig 1.
HIV-1 VL at the set point is highly correlated with (**A**) the *HCP5* rs2395029 genotype, where T is the major allele and G is the minor allele, and with (**B**) the *HLA-C* 5′ region rs9264942 genotype, where T is the major allele and C is the minor allele. Mean and SEM (error bars) are represented for the respective genotypes.
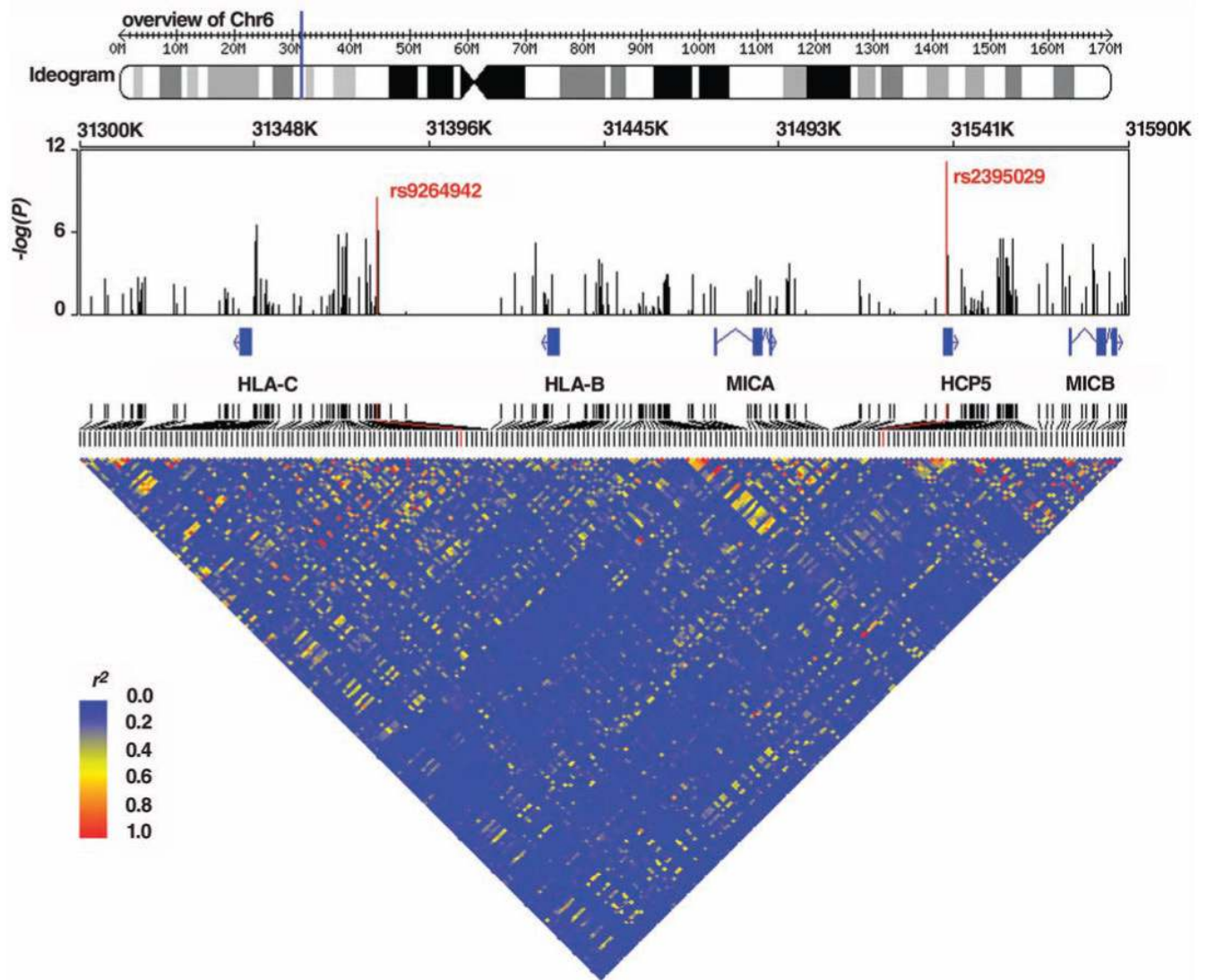
Fig 2.
Partial map of the *HLA* class I region (chromosome 6 p21.3). The *P* values [−log(*P*)] of all genotyped SNPs annotated with the gene structure are indicated. The two independent SNPs that show genome-wide significant association with HIV-1 VL at the set point are marked in red. The graph was drawn with WGAViewer software (www.genome.duke.edu/centers/pg2/index_html/downloads/AnnotationSoftware).

Table 1

The impact of *HLA-C 5′* expression polymorphism rs9264942 on the set point is independent of its association with *HLA* alleles and groups of alleles previously implicated in HIV-1 control. The addition of rs9264942 to the linear regression model significantly improves the fit for all *HLA* alleles or groups of alleles that have been suspected to have an influence on HIV disease. N.A., not applicable.

Allele	LD between rs9264942 and HLA alleles (r^2)	Models with HLA alleles (*P* value)	Addition of rs9264942 to models with HLA alleles (*P* value)	Addition of *HLA* alleles to a model with rs9264942 (*P* value)
*HLA-B*27*	0.07	0.19	1.3×10^{-5}	0.91
*HLA-B*5701*	0.05	2.6×10^{-5}	8.5×10^{-5}	4.1×10^{-4}
*HLA-B*35Px*	0.04	0.18	8.1×10^{-6}	0.73
*HLA-B*08*	0.09	0.042	3.6×10^{-5}	0.41
*HLA-B*51*	<0.01	0.44	5.3×10^{-6}	0.36
All 5 above	N.A.	N.A.	3.1×10^{-4}	4.4×10^{-4}
B22 serogroup	0.01	0.28	8.1×10^{-6}	0.59
B7 supertype	0.17	0.007	1.7×10^{-4}	0.35
Bw4 serotype	0.23	0.010	1.7×10^{-4}	0.65
Bw6 serotype	0.16	0.033	6.3×10^{-5}	0.76
*HLA-A*23*	<0.01	0.41	6.9×10^{-5}	0.38
A2 supertype	<0.01	0.72	6.4×10^{-5}	0.53
*HLA-Cw*4*	0.10	0.041	1.3×10^{-4}	0.56
*HLA-Cw*7*	0.25	0.065	1.0×10^{-4}	0.83

Table 2

In contrast to Table 1, only *HLA-B*5701* has an independent impact after taking into account the effect of rs9264942. The independence of *HLA-C* is also clearly seen in the mean values of the HIV-1 set point for each rs9264942 genotype: The minor allele C is associated with a decrease in VL, independent of all considered alleles and groups of alleles. Numbers refer to a subgroup of 187 patients with available four-digit *HLA* class I allelic results.

rs9264942 genotype	Number of patients	Mean	SD
	All patients		
TT	67	4.37	0.85
TC	87	3.84	1.19
CC	33	3.24	1.28
	*Patients without HLA-B*27*		
TT	66	4.39	0.83
TC	78	3.86	1.21
CC	25	3.13	1.23
	*Patients without HLA-B*5701*		
TT	67	4.37	0.85
TC	76	3.97	1.08
CC	27	3.43	1.25
	*Patients without HLA-B*35Px*		
TT	56	4.32	0.88
TC	84	3.86	1.17
CC	32	3.23	1.30
	*Patients without HLA-B*08*		
TT	50	4.26	0.87
TC	81	3.83	1.22
CC	33	3.24	1.28
	*Patients without HLA-B*51*		
TT	56	4.38	0.85
TC	73	3.77	1.17
CC	28	3.21	1.38