Sciencexpress Report

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]† 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, IRCCS 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, UCL London NW3 2PF, UK. [8]Institute of Clinical Infectious Diseases, Catholic University of the Sacred Heart, 00168 Rome, Italy. [9]Academic Department of HIV/GUM, 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 - IDIBAPS, University of Barcelona, 08036 Barcelona, Spain. [14]Department of Infectious Diseases, Copenhagen University Hospital, Rigshospitalet, 2100 Copenhagen, Denmark. [15]Genomics Platform, NCCR "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]MRC 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.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: amalio.telenti@chuv.ch (A.T.); d.goldstein@duke.edu (D.B.G.)

‡On behalf of the Center for HIV/AIDS Vaccine Immunology (CHAVI) and the Euro-CHAVI consortia.

Understanding why some people establish and maintain effective control of HIV-1 and others do 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 *HLA-B*5701*, while a second is located near the *HLA-C* gene. An additional analysis of the time to HIV disease progression implicated a third locus encoding a RNA polymerase 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 following infection. One striking and largely unexplained difference is the level of circulating virus in the plasma during the non-symptomatic phase preceding 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*). Here we aimed to identify human genetic differences that influence this variation.

In order to define a homogeneous phenotype for genetic analyses, a consortium of 9 cohorts was formed [termed Euro-CHAVI (*7*)] and a total of 30,000 patients 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 individually inspected by an experienced infectious disease

clinician (JF) 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 4 CD4 cell count results, we defined a progression phenotype as the time to treatment initiation, or the time to the predicted or observed drop of CD4 cells below 350 (*7*, *8*).

All samples were genotyped using Illumina's HumanHap550 genotyping 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 (CNV) and to assess whether they influence the phenotype (*7*). Our core association analyses focused on single-marker genotype trend tests of the QC-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 *HLA* typing were carried out on a subgroup of 187 patients with available 4-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 MHC upstream of a gene that encodes an RNA polymerase 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 (dbSNP rs2395029, $P = 9.36 \times 10^{-12}$). A single copy of the controlling allele was found to result in a greater than one log reduction in VL (Fig. 1) and 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 dataset). 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 SNP within *HCP5* (*10*). We emphasize however that genetics allows no resolution on 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.

A model in which *HCP5* and *HLA-B*5701* have a combined haplotypic effect on 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-lymphocytes (CTL) mediated restriction (*15*). This observation has also been explained however 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, addition of the *HLA-C* variant results in a highly significant increase in explanatory power as does 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 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 does show association with certain alleles or group of alleles (Table 1). In each case, however, while the *HLA-C* expression variant can explain the

effect of these alleles on HIV-1 set point the reverse is not true. When a linear regression model includes known *HLA* alleles, 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 (Table 1).

No other single marker reached genome significance, and none of the identified copy number variations (*7*) showed any association with 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 amongst the most associated polymorphisms in this study (complete list available in table S2). Potentially interesting candidates with a lesser association with set point are listed in tables S3 to S5, chosen on the basis of their ranking in the study or of 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 the 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 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 of *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). Notably, two of them are located in a putative regulatory 5′ region 25 and 32 kb away from the gene (rs3869068 and rs9261174, respectively). As *ZNRD1* encodes an RNA polymerase subunit, a possible interaction with HIV-1 during transcription is the most plausible causal mechanism if this gene indeed restricts HIV-1. 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 (*19*).

The second gene, *RNF39*, is poorly characterized, but cannot be ruled out as a candidate as 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 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 to positive test (for infection) from two years to four years to identify additional qualifying subjects. We genotyped representative polymorphisms for the associations reported above (*HCP5* rs2395029; *HLA-C* rs9264942; *ZNRD1* rs9261174). Each association was replicated with effects all in the same direction: *HCP5*, $P = 1.4 \times 10^{-2}$; *HLA-C*, $P = 2.8 \times 10^{-3}$; *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 *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 major effect gene variants. This suggests a degree of urgency in carrying out similar studies for other infectious diseases.

Finally, we note that 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 HIV-1 accessory protein nef selectively down regulates the expression of *HLA-A* and *-B* but not of *HLA-C* on the surface of infected cells (*20*). 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 cells (NK) in order to avoid NK attack. However, *HLA-C* also has the ability to present viral peptides to cytotoxic CD8+ T cells and consequently restrict HIV-1 (*21*, *22*). 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 in vaccine strategies targeting *HLA-C* mediated responses.

References and Notes

1. A. Telenti, D. B. Goldstein, *Nat. Rev. Microbiol.* **4**, 865 (2006).
2. S. J. O'Brien, G. W. Nelson, *Nat. Genet.* **36**, 565 (2004).
3. A. Telenti, G. Bleiber, *Future Virology* **1**, 55 (2006).
4. M. Carrington, S. J. O'Brien, *Annu. Rev. Med.* **54**, 535 (2003).
5. G. W. Nelson, S. J. O'Brien, *J. Acquir. Immune Defic. Syndr.* **42**, 347 (2006).
6. G. Bleiber *et al.*, *J. Virol.* **79**, 12674 (2005).
7. Materials and methods are available as supporting material on *Science* Online.
8. D. C. Douek, L. J. Picker, R. A. Koup, *Annu. Rev. Immunol.* **21**, 265 (2003).
9. A. L. Price *et al.*, *Nat. Genet.* **38**, 904 (2006).
10. P. I. de Bakker *et al.*, *Nat. Genet.* **38**, 1166 (2006).
11. S. A. Migueles *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2709 (2000).
12. M. Altfeld *et al.*, *Aids* **17**, 2581 (2003).
13. J. K. Kulski, R. L. Dawkins, *Immunogenetics* **49**, 404 (1999).
14. C. Vernet *et al.*, *Immunogenetics* **38**, 47 (1993).
15. J. R. Bailey *et al., J. Exp. Med.* **203**, 1357 (2006).
16. J. Martinez-Picado *et al., J. Virol.* **80**, 3617 (2006)
17. A. J. McMichael, S. L. Rowland-Jones, *Nature* **410**, 980 (2001).
18. B. E. Stranger *et al.*, *PLoS Genet.* **1**, e78 (2005).
19. A. Ciuffi *et al.*, *J. Virol.* **78**, 10747 (2004).
20. G. B. Cohen *et al.*, *Immunity* **10**, 661 (1999).
21. P. J. Goulder *et al., Aids* **11**, 1884 (1997).
22. S. Adnan *et al. Blood* **108**, 3414 (2006)
23. We thank R. Sahli (University of Lausanne, Switzerland). Funding was provided by the NIH funded Center for HIV/AIDS Vaccine Immunology. J.F. is supported by the Swiss Foundation for Grants in Biology and Medicine. A.T. is supported by Infectigen and the Swiss National Science Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1143767/DC1
Material and Methods
SOM Text
Figs. S1 to S3
Tables S1 to S6
References

13 April 2007; accepted 2 July 2007
Published online 19 July 2007; 10.1126/science.1143767
Include this information when citing this paper.

Fig. 1. HIV-1 VL at set point is highly correlated with *HCP5* rs2395029 genotype (**A**) and with *HLA-C* 5′ region rs9264942 genotype (**B**). Mean and SEM are represented for the respective genotypes.

Fig. 2. Partial map of the *HLA* Class I region (chromosome 6 p21.3). Indicated are the *P*-values [–log(*P*)] of all genotyped SNPs annotated with the gene structure. The two independent SNPs that show genome-wide significant association with HIV-1 VL at set point are displayed and marked in red. Graph was drawn from WGAViewer software (http://www.genome.duke.edu/centers/pg2/index_html/downloads/AnnotationSoftware).

Table 1. The impact of *HLA-C 5′* expression polymorphism rs9264942 on 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 improves fit significantly for all *HLA* alleles or groups of alleles that have been suspected to have an influence on HIV disease (**A**). In contrast, only *HLA-B*5701* has an independent impact after taking into account rs9264942 effect. The independence of *HLA-C* is also clearly seen in the mean values of HIV-1 set point for each rs9264942 genotype (**B**): The minor allele C is associated with a decrease in VL independently of all considered alleles and groups of alleles. Numbers refer to a subgroup of 187 patients with available 4-digit *HLA* Class I allelic results.

A

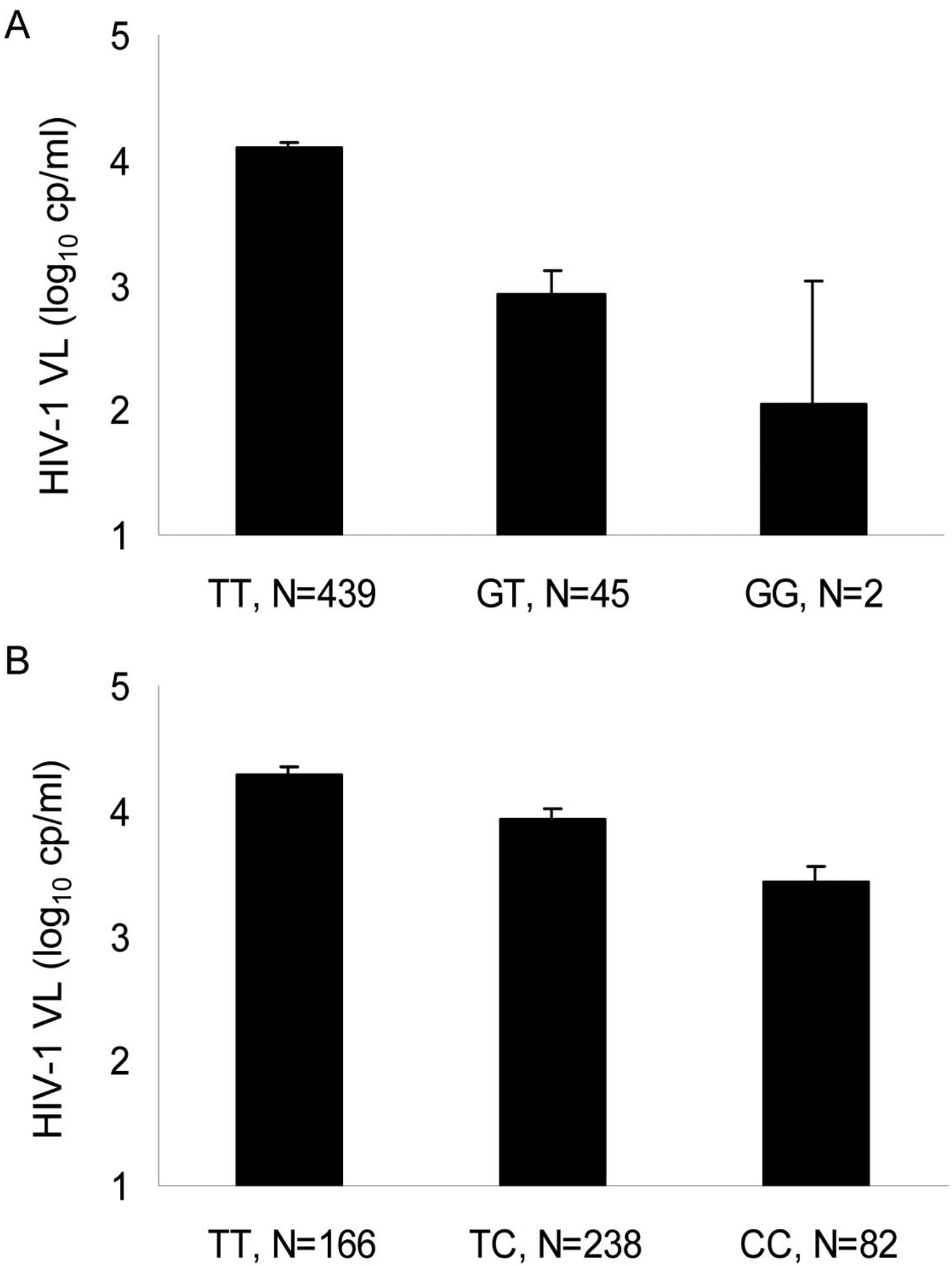
	LD between rs9264942 and *HLA* alleles, r^2	Models with *HLA* alleles, *P*-value	Addition of rs9264942 to models with *HLA* alleles, *P*-value	Addition of *HLA* alleles to a model with rs9264942, *P*-value
*HLA-B*27*	0.07	0.19	1.3×10^{-5}	0.91
*HLA-B*5701*	0.05	2.6×10^{-5}	8.5×10^{-5}	4.1×10^{-4}
*HLA-B*35Px*	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

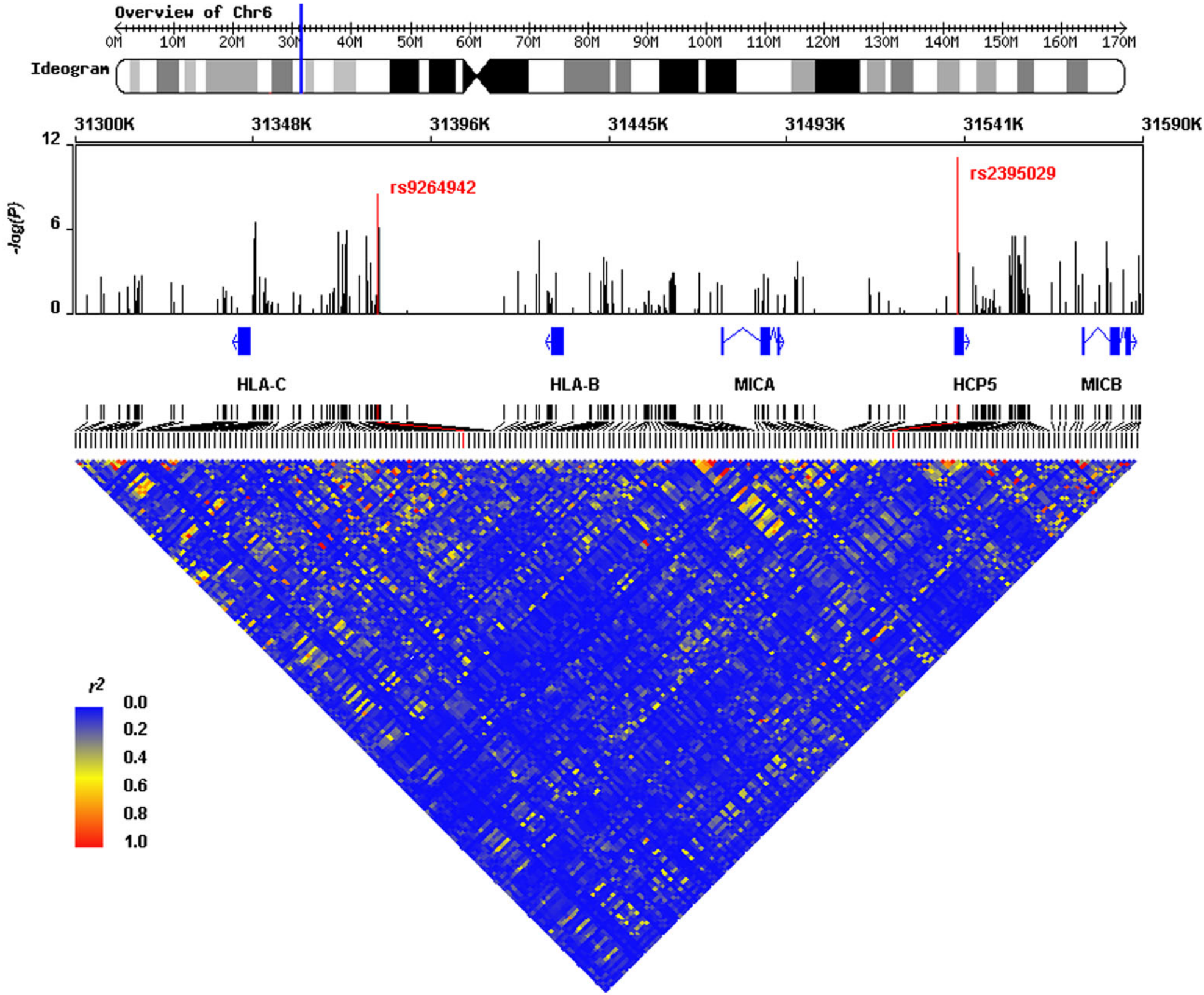
N.A., not applicable; LD, linkage disequilibrium.

B

	rs9264942 genotype	N	*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

Patients without any of the above alleles	TT	29	4.21	0.89
	TC	47	3.94	1.11
	CC	15	3.25	1.45
Patients without *B22* serogroup alleles	TT	64	4.35	0.86
	TC	85	3.81	1.19
	CC	33	3.24	1.28
Patients without *B7* supertype alleles	TT	18	4.33	0.78
	TC	47	3.70	1.21
	CC	27	3.20	1.40
	TT	42	4.48	0.85
Patients without *Bw4* serotype	TC	20	4.19	0.89
	CC	5	2.52	0.64
	TT	6	3.98	0.85
Patients without *Bw6* serotype	TC	19	4.05	1.17
	CC	18	3.31	1.46
	TT	59	4.34	0.85
Patients without *HLA-A*23*	TC	82	3.82	1.19
	CC	29	3.34	1.29
	TT	30	4.25	0.84
Patients without *A2* supertype	TC	41	4.02	1.10
	CC	10	2.58	1.38
	TT	40	4.32	0.87
Patients without *HLA-Cw*4*	TC	70	3.83	1.19
	CC	31	3.23	1.32
Patients without *HLA-Cw*7*	TT	16	4.28	0.85
	TC	47	3.92	1.19
	CC	30	3.15	1.27


A
5
4
3
2
1
HIV-1 VL (log10 cp/ml)
TT, N=439
GT, N=45
GG, N=2
B
5
4
3
2
1
HIV-1 VL (log10 cp/ml)
TT, N=166
TC, N=238
CC, N=82


Overview of Chr6
0M
10M
20M
30M
40M
50M
60M
70M
80M
90M
100M
110M
120M
130M
140M
150M
160M
170M
Ideogram
31300K
31348K
31396K
31445K
31493K
31541K
31590K
12
6
0
-log(P)
rs9264942
rs2395029
HLA-C
HLA-B
MICA
HCP5
MICB
r2
0.0
0.2
0.4
0.6
0.8
1.0