Title: The major genetic determinants of HIV-1 control affect HLA class I peptide presentation.

The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation

The International HIV Controllers Study*,†, Writing team, Florencia Pereyra1,2,‡, Xiaoming Jia3,‡, Paul J. McLaren4,5,‡, Amalio Telenti6, Paul I.W. de Bakker4,5,7,8,†[co-chair], Bruce D. Walker1,8,†[co-chair], Analysis team, Xiaoming Jia3, Paul J. McLaren4,5[project leaders], Stephan Ripke4,10, Chanson J. Brumme1, Sara L. Pulit4,5, Amalio Telenti6, Mary Carrington1,11, Carl M. Kadie12, Jonathan M. Carlson13, David Heckerman13, Paul I.W. de Bakker4,5,7,8,†[chair]. Study design, Florencia Pereyra1,2, Paul I.W. de Bakker4,5,7,8,†, Robert R. Graham14, Robert M. Plenge1,15, Steven G. Deeks16, Bruce D. Walker1,8,†[chair]. SNP genotyping, HLA typing, and sample management, Lauren Gianniny4, Gabriel Crawford4, Jordan Sullivan4, Elena Gonzalez2, Leela Davies4, Amy Camargo4, Jamie M. Moore4, Nicole Beattie8, Supriya Gupta8, Andrew Crenshaw4, Noël P. Burnett6, Candace Guiducci4, Namrata Gupta1, Mary Carrington1,11, Xiaojiang Gao11, Ying Qi11, Yuko Yukit11, HIV controllers recruitment and sample management, Florencia Pereyra1,2[project leader], Alicia Piechocka-Trocha1, Emily Currell1, Rachel Rosenberg1, Kristin L. Moss1, Paul Lemay1, Jessica O’Leary1, Todd Schaefer1, Pranshu Verma1, Ildiko Toth1, Brian Block1, Brett Baker1, Alissa Rothschild1, Jeffrey Lian1, Jacqueline Proudfoot1, Donna Marie L. Alvino1, Seanna Vine1, Maryln M. Addo1, Todd M. Allen1, Marcus Altfeld1, Matthew R. Henn4, Sylvie Le Gall1, Hendrik Streeck1, Bruce D. Walker1,8,†[chair], AIDS Clinical Trials Group, David W. Haas17, Daniel R. Kurtzkes2, Gregory K. Robbins18, Robert W. Shafer19, Roy M. Gulick20, Cecilia M. Shikuma21, Richard Haubrich22, Sharon Riddler23, Paul E. Sax2, Eric S. Daar24, Heather J. Ribaudo25, HIV controllers referral team, Brian Agan26, Shana Agarwal27, Richard L. Ahern18, Brady L. Allen28, Sherly Alldor29, Eric L. Altschuler30, Sujata Ambardar31, Kathryn Anastos32, Ben Anderson33, Val Anderson34, Usha Andrade34, Diana Antoniskis35, David Bangsberg1,18, Daniel Barbaro36, William Barrie37, J. Bartczak38, Simon Barton39, Patricia Basden40, Nesli Basgoz18, Suzane Bazner1, Nicholas C. Bellois41, Anne M. Benson10, Judith Berger42, Nicole F. Bernard43, Annette M. Bernard44, Christopher Birch1, Stanley J. Bodner45, Robert K. Bolan46, Emilie T. Boudreaux47, Meg Bradley1, James F. Braun48, Jon E. Brndjar49, Stephen J. Brown50, Katherine Brown51, Sheldon T. Brown52, Jedidiah Burack53, Larry M. Bush54, Virginia Cafaro55, Obombola Campbell18, John Campbell56, Robert H. Carlson57, J. Kevin Carmichael58, Kathleen K. Casey59, Chris Cavacuti60, Gregory Celestin61, Steven T. Chambers62, Nancy Chez63, Lisa M. Chirch64, Paul J. Cimoch65, Daniel Cohen66, Lillian E. Cohn67, Brian Conway68, David A. Cooper69, Brian Cornelson60, David T. Cox70, Michael V. Cristofano71, George Cuchural Jr. 72, Julie L. Czartoski73, Joseph M. Dahanman74, Jennifer S. Daly75, Benjamin T. Davis18, Kristine Davis76, Sheila M. Davud18, Steven G. Deeks18, Edwin DeJesus77, Craig A. Dietz78, Eleanor Dunham59, Michael E. Dunn79, Todd B. Ellerin80, Joseph J. Eron81, John J.W. Fangman82, Claire E. Fare18, Helen Ferlazzo83, Sarah Fieder84, Anita Fleenor-Ford85, Renee Franke86, Kenneth A. Freedberg18, Neel K. French87, Jonathan D. Fuchs88, Jon D. Fuller89, Jonna Garberman90, Joel E. Gallant91, Rajesh T. Gandhi18, Efrain Garcia92, Donald

Copyright 2010 by the American Association for the Advancement of Science; all rights reserved.

*To whom correspondence should be addressed. bwalker@partners.org (B.D.W.); pdebakker@rics.bwh.harvard.edu (P.I.W.d.B.).

†These authors contributed equally to this work.

‡All authors with their contributions and affiliations appear at the end of this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1195271/DC1
Lausanne, Switzerland 7Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center Utrecht, Netherlands 8Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands 9Howard Hughes Medical Institute, Chevy Chase, MD, USA 10Department of Medicine, Center for Human Genetic Research, MGH, Harvard Medical School, Boston, MA, USA 11Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, NCI-Frederick, Frederick, MD, USA 12Microsoft Research, Redmond, WA, USA 13Microsoft Research, Los Angeles, CA, USA 14Genentech, South San Francisco, CA, USA 15Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA 16University of California San Francisco, San Francisco, CA, USA 17Vanderbilt University School of Medicine, Nashville, TN, USA 18MGH, Harvard Medical School, Boston, MA, USA 19Stanford University, Palo Alto, CA, USA 20Weill Medical College of Cornell University, New York, NY, USA 21Hawaii Center for AIDS, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA 22University of California San Diego, San Diego, CA, USA 23University of Pittsburgh, Pittsburgh, PA, USA 24University of California Los Angeles, Los Angeles, CA, USA 25Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA 26Infectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD, USA 27Summa Health System, Akron, OH, USA 28Uptown Physicians Group, Dallas, TX, USA 29St. Luke’s Roosevelt Hospital, New York, NY, USA 30New Jersey Medical School, University Hospital, Newark, NJ, USA 31Infectious Disease Physicians, Annandale, VA, USA 32Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY, USA 33St. Leonards Medical Centre, St. Leonards, Australia 34Ysbyty Gwynedd Hospital, Gwynedd, UK 35Kaiser Permanente, Portland, OR, USA 36Tarrant County Infectious Disease Associates, Fort Worth, TX, USA 37Private Practice of William Barrie, Toronto, Canada 38Rowan Tree Medical, Wilton Manors, FL, USA 39Chelsea and Westminster Hospital, St. Stephen’s Centre, London, UK 40Beaver Street Family Practice, Flagstaff, AZ, USA 41Southwest Infectious Disease Associates, Dallas, TX, USA 42St. Barnabas Hospital, Bronx, NY, USA 43Research Institute, McGill University Health Centre, Montreal General Hospital, Montreal, Canada 44Thacker, Thompson and Bernard, Atlanta, GA, USA 45Vanderbilt University School of Medicine, Hermitage, TN, USA 46LA Gay and Lesbian Center, Los Angeles, CA, USA 47Louisiana State University Health Sciences Center, University Medical Center East Clinic, Lafayette, LA, USA 48Physicians’ Research Network, Callen-Lorde Community Health Center, New York, NY, USA 49Brndjar Medical Associates, Allentown, PA, USA 50AIDS Research Alliance, Los Angeles, CA, USA 51David Powell Community Health Center, Austin, TX, USA 52James J. Peters VA Medical Center, Bronx, NY, USA 53Sunrise Medical Group, Brooklyn, NY, USA 54University of Miami-Miller School of Medicine, Lake Worth, FL, USA 55WellSpring Medical Group, San Francisco, CA, USA 56Moses Cone Health System, Greensboro, NC, USA 57HealthPartners Infectious Disease, St Paul, MN, USA 58El Rio Special Immunology Associates, Tuscon, AZ, USA 59Jersey Shore University Medical Center, Neptune, NJ, USA 60St. Michaels Hospital, Toronto, Canada 61The Brooklyn Hospital Center, PATH Center, Brooklyn, NY, USA 62University of Otago, Christchurch, New Zealand 63H.E.L.P./Project Samaritan, Bronx, NY, USA 64David E. Rogers Center for HIV/AIDS Care, Southhampton, NY, USA 65Center for Special Immunology, Fountain Valley, CA, USA 66Fenway Community Health, Boston, MA, USA 679th Street Internal Medicine Associates, Philadelphia, PA, USA 68University of British Columbia, Vancouver, Canada 69National Centre in HIV Epidemiology and Clinical Research, Sydney, Australia 70Metro Infectious Disease Consultants, Indianapolis, IN, USA 71John H. Stroger Hospital of Cook County, Chicago, IL, USA 72New England Quality Care Alliance, Braintree, MA, USA 73Fred Hutchinson Cancer Research Center, Seattle, WA, USA 74Desert AIDS Project, Palm Springs, CA, USA 75University of Massachusetts Memorial Medical Center, Worcester, MA, USA 76University of Iowa Hospitals and Clinics, Iowa City, IA, USA 77Orlando Immunology Center, Orlando, FL, USA 78The Kansas City Free Health Clinic, Kansas City, MO, USA 79Private Practice of Michael E. Dunn, M.D., Tampa,
et al. Page 4

Science. Author manuscript; available in PMC 2011 December 11.

FL, USA 80South Shore Hospital, Weymouth, MA, USA 81University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 82AIDS Resource Center of Wisconsin, Milwaukee, WI, USA 83Visiting Nurse Association of Central New Jersey, Community Health Center, Asbury Park, NJ, USA 84Imperial College, London, UK 85Heartland Clinic, Paducah, KY, USA 86Morristown Memorial Hospital, Morristown, NJ, USA 87Private Practice of Neel K. French, M.D., Chicago, IL, USA 88San Francisco Department of Public Health, San Francisco, CA, USA 89Boston University Medical Center, Boston, MA, USA 90Baystate Medical Center, Springfield, MA, USA 91Johns Hopkins University School of Medicine, Baltimore, MD, USA 92Private Practice of Efrain Garcia, M.D., Miami, FL, USA 93The Rockefeller University, New York, NY, USA 94Private Practice of Joseph C. Gathe Jr., M.D., Houston, TX, USA 95Tower Infectious Disease, Los Angeles, CA, USA 96Nassau University Medical Center, East Meadow, NY, USA 97Sharp Rees Stealy Medical Center, San Diego, CA, USA 98Medical College of Wisconsin, Milwaukee, WI, USA 99University of Alabama, Birmingham, Birmingham, AL, USA 100Synergy Hematology and Oncology, Los Angeles, CA, USA 101University of Utah, Salt Lake City, UT, USA 102South Dayton Acute Care Consultants, Dayton, OH, USA 103T. Douglas Gurley, M.D., Atlanta, GA, USA 104St. Vincent’s Hospital, New York, NY, USA 105University of Texas Health Science Center, Houston, TX, USA 106British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada 107Southwest C.A.R.E. Center, Santa Fe, NM, USA 108Hennepin County Medical Center, Minneapolis, MN, USA 109The Catholic University of America, School of Nursing, Washington, DC, USA 110Mercy Medical Center, Springfield, MA, USA 111CMC Myers Park Medical Center, Charlotte, NC, USA 112New York University Medical Center, New York, NY, USA 113The Ruth M. Rothshon Care Center, Chicago, IL, USA 114Feldman Medical Group, San Francisco, CA, USA 115HIV Research and Clinical Care Centre, Munich, Germany 116Albany Medical College, Albany, NY, USA 117Murdoch University, Murdoch, Australia 118University of Cincinnati, Cincinnati, OH, USA 119OASIS Clinic, Los Angeles, CA, USA 120VA Ann Arbor Healthcare System, Ann Arbor, MI, USA 121AIDS Healthcare Foundation, Los Angeles, CA, USA 122Capitol Hill Medical, Seattle, WA, USA 123Astor Medical Group, New York, NY, USA 124The Kinder Medical Group, Miami, FL, USA 125Lancaster General Hospital, Lancaster, PA, USA 126Tufts Medical Center, Boston, MA, USA 127Alpert Medical School of Brown University, Providence, RI, USA 128Oregon Health and Science University, Portland, OR, USA 129G.B. Rossi Hospital, Verona, Italy 130Case Western Reserve University, Cleveland, OH, USA 131LifeWay, Fort Lauderdale, FL, USA 132Saint Claire Medical Associate, Toronto, Canada 133Greater Lawrence Family Health Center, Lawrence, MA, USA 134Hospital Universitari Germans Trias i Pujol, Barcelona, Spain 135Infectious Disease Consultants, Tucker, GA, USA 136University of Toronto, Toronto, Canada 137Infectious Disease Associates, Sarasota, FL, USA 138Emory University, School of Medicine, Atlanta, GA, USA 139Aaron Diamond AIDS Research Center, Rockefeller University, New York, NY, USA 140Deruico Doencas Infecciosas, Porto, Portugal 141Park Nicollet Clinic, St. Louis, MN, USA 142Absolute Care, Atlanta, GA, USA 143College of Physicians and Surgeons, Columbia University, New York, NY, USA 144Highland Medical Associates, Hendersonville, NC, USA 145Medizinische Hochschule, Abteilung Klinische Immunologie, Hannover, Germany 146Hospital for Special Surgery, New York, NY, USA 147Family Practice Specialists, Phoenix, AZ, USA 148Philadelphia FIGHT, Philadelphia, PA, USA 149Lower East Side Service Center, New York, NY, USA 150Infectious Diseases Care, Hillsborough, NJ, USA 151University of Colorado, Denver, Aurora, CO, USA 152Sutter Medical Group, Sacramento, CA, USA 153University of Wisconsin in Madison, Madison, WI, USA 154Brooke Army Medical Center, San Antonio, TX, USA 155Eastern Virginia Medical School, Norfolk, VA, USA 156Columbia University Medical Center, New York, NY, USA 157CA Pacific Medical Center, San Francisco, CA, USA 158Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA 159St. Joseph Hospital, Chicago, IL, USA 160Aaron M. Perslmutter, M.D., Beverly Hills, CA, USA 161Codman Square Health Center, Dorchester, MA, USA 162CARE Resource, Miami, FL, USA 163Castro-Mission Health Center, San Francisco, CA, USA 164Rush Medical College, Chicago, IL, USA 165Boulder Community Hospital, Boulder, CO, USA 166Midway Immunology and
Infectious and inflammatory diseases have repeatedly shown strong genetic associations within the major histocompatibility complex (MHC); however, the basis for these associations remains elusive. To define host genetic effects on the outcome of a chronic viral infection, we performed genome-wide association analysis in a multiethnic cohort of HIV-1 controllers and progressors, and we analyzed the effects of individual amino acids within the classical human leukocyte antigen (HLA) proteins. We identified >300 genome-wide significant single-nucleotide polymorphisms (SNPs) within the MHC and none elsewhere. Specific amino acids in the HLA-B peptide binding groove, as well as an independent HLA-C effect, explain the SNP associations and reconcile both protective and risk HLA alleles. These results implicate the nature of the HLA–viral peptide interaction as the major factor modulating durable control of HIV infection.

Hiv infection is characterized by acute viremia, often in excess of 5 million viral particles per milliliter of plasma, followed by an average 100-fold or greater decline to a relatively stable plasma virus load set point (1). In the absence of antiretroviral therapy, the level of
viremia is associated with the rate of CD4+ T cell decline and progression to AIDS. There is substantial interperson variability in the virus load set point, with most individuals having stable levels exceeding 10,000 RNA copies/ml. Yet a small number of people demonstrate sustained ability to control HIV replication without therapy. Such individuals, referred to as HIV controllers, typically maintain stable CD4+ cell counts, do not develop clinical disease, and are less likely to transmit HIV to others (2).

To determine the genetic basis for this rare phenomenon, we established a multinational consortium (www.hivcontrollers.org) to recruit HIV-1 controllers, who are defined by at least three measurements of plasma virus load (VL) < 2000 RNA copies/ml over at least a 12-month period in the absence of antiviral therapy. We performed a genome-wide association study (GWAS) in the HIV controllers (median VL, CD4 count, and disease duration of 241 copies/ml, 699 cells/mm3, and 10 years, respectively) and treatment-naïve chronically infected individuals with advanced disease (median VL and CD4 count of 61,698 copies/ml and 224 cells/mm3, respectively) enrolled in antiviral treatment studies led by the AIDS Clinical Trials Group. After quality control and imputation on the basis of HapMap Phase 3 (3), we obtained data on 1,384,048 single-nucleotide polymorphisms (SNPs) in 974 controllers (cases) and 2648 progressors (controls) from multiple populations (table S1).

After stratification into European, African American, and Hispanic ethnic groups (fig. S1), we tested each SNP for association using logistic regression, including the major principal components as covariates to correct for population substructure (4). In the largest group, comprising 1712 individuals of European ancestry, we identified 313 SNPs with genome-wide significance, defined by \( P < 5 \times 10^{-8} \) due to correction for multiple comparisons (table S2). All SNPs that reached genome-wide significance were located in the major histocompatibility complex (MHC) region on chromosome 6 (Fig. 1A). We obtained similar results for the other two ethnic groups and in a meta-analysis of all participants (fig. S2). We also performed a genome-wide analysis to test the influence of local chromosomal ancestry in the African American sample (4), but we detected no signal outside the MHC (figs. S3 and S4). The impact of the MHC was further underscored when we specifically tested published associations related to HIV disease progression outside the MHC. Only variants in the CCR5-CCR2 locus—namely, CCR5Δ32 deletion polymorphism (5), C927T in CCR5 (6), and Val64→Ile64 in CCR2 (7)—replicate with nominal statistical significance in our study (Fig. 1B and table S3).

Closer examination of the significant SNPs within the MHC showed that they are located within a 3-Mb region concentrated around class I human leukocyte antigen (HLA) genes (fig. S5), but extensive linkage disequilibrium (LD) makes precise assignment of causal variants challenging (8). Therefore, we used stepwise regression to define independent markers associated with host control. From the initial set of 313 SNPs that reached genome-wide significance in the European sample, for which the greatest numbers of participants were available, we found only four independent markers of association (Table 1). rs9264942, located 35 kb upstream of HLA-C and a putative variant associated with HLA-C expression levels [odds ratio (OR) = 2.9, \( P = 2.8 \times 10^{-35} \), where an OR > 1 indicates a protective effect], and rs2395029, a proxy for HLA-B*57:01 (OR = 5.3, \( P = 9.7 \times 10^{-26} \), had been previously reported to be associated with virus load set point after acute infection (9). We also defined rs4418214, a noncoding SNP near MICA (OR = 4.4, \( P = 1.4 \times 10^{-34} \)), and rs1313018 in PSORS1C3, a gene implicated in psoriasis (OR = 2.1, \( P = 4.2 \times 10^{-16} \)). These four SNPs explain 19% of the observed variance of host control in the European sample; together with those in CCR5, these SNPs explain 23%, using Nagelkerke’s approximation (Fig. 1C) (10).
In the smaller African American sample, we observed 33 SNPs with genome-wide significance, four of which were identified as independent markers, but all differed from those in the European sample (Table 1). This suggests that shared causal variants are tagged by different SNPs in these two populations or that the mechanism of control differs with ethnicity. Only rs2523608 was previously identified, in a recent study of virus load set point in African Americans (11). Despite no evidence for historical recombination ($D' = 1$), this SNP is only weakly correlated ($r^2 < 0.1$) with $HLA-B^*57:03$, the class I allele most strongly associated with durable control of HIV in populations of African ancestry (11-13). In the Hispanic sample, which was much smaller, the most significant SNP was rs2523590, 2 kb upstream of $HLA-B$, also identified in the African American sample described here.

Given the localization of significant SNPs entirely to the HLA class I region, as well as previous studies showing $HLA$ alleles to affect disease progression (13-20), we next sought to evaluate whether these SNP and HLA associations might be due to specific amino acids within HLA. Because $HLA$ types were available for only a portion of the entire cohort, we developed a method to impute classical $HLA$ alleles and their corresponding amino acid sequences (4) on the basis of haplotype patterns in an independent data set collected by the Type 1 Diabetes Genetics Consortium (T1DGC) (21). This data set contains genotype data for 639 SNPs in the MHC that overlap with genotyped SNPs in our GWAS and classical $HLA$ types for class I and II loci at four-digit resolution in 2767 unrelated individuals of European descent.

We imputed $HLA$ types in the European sample of our study and validated the imputations by comparing to empirical four-digit $HLA$ typing data collected for class I loci in a subset ($n = 371$) of the HIV controllers. The quality of the imputations was such that the imputed and true frequencies for all $HLA$ alleles in this subset were in near-perfect agreement (Fig. 2A) ($r^2 = 0.99$). Furthermore, the positive predictive value was 95.2% and the sensitivity was 95.2% at two-digit resolution (92.7 and 95.6%, respectively, at four-digit resolution) for $HLA$ alleles with frequency >2% (Fig. 2B). This indicates that the performance of the imputation was generally excellent for common alleles, consistent with previous work (22). We used $HLA$ allele imputations in all participants (even those with $HLA$ types defined by sequencing) for association analyses to avoid systematic bias between cases and controls. Lower imputation quality would only decrease power, not increase the false-positive rate, because cases and controls would be equally affected.

We tested all $HLA$ alleles for association via logistic regression, adjusting for the same covariates used in SNP analysis (tables S4 and S5). The most significant HLA association is $B^*57:01$ (OR = 5.5, $P = 1.4 \times 10^{-26}$), which explains the proxy association of rs2395029 in $HCP5$. With the use of stepwise regression modeling in the European sample of controllers and progressors, we were able to implicate $B^*57:01, B^*27:05, B^*14/Cw^*08:02, B^*52,$ and $A^*25$ as protective alleles and $B^*35$ and $Cw^*07$ as risk alleles. These associations are consistent with earlier studies that highlighted a role for $HLA$ class I loci (13-20), and particularly $HLA-B$ alleles in control of HIV, which indicated that the imputations are robust. Collectively explaining 19% of the variance of host control, these $HLA$ allele associations are consistent with the effects of the four independent SNPs.

Virus-infected cells are recognized by CD8$^+$ T cells after presentation of short viral peptides within the binding groove of HLA class I, and HIV-specific CD8$^+$ T cells are strongly associated with control (23). We thus evaluated whether the SNP associations identified in the GWAS, and the HLA associations derived from imputation, might be due to specific amino acid positions within the HLA molecules, particularly those involved in the interaction between the viral peptide and the HLA class I molecule. Using the official DNA sequences defined for known HLA alleles (24), we encoded all variable amino acid positions.
within the coding regions of the HLA genes in each of the previously HLA-typed 2767 individuals in the T1DGC reference panel, and we used this data set to impute the amino acids in the cases and controls (4). Among a total of 372 polymorphic amino acid positions in class I and II HLA proteins, 286 are biallelic like a typical nonsynonymous coding SNP. The remaining 86 positions accommodate more than two amino acids; position 97 is the most diverse in HLA-B with six possible amino acids observed in European populations.

After imputing these amino acids in the European sample, we used logistic regression to test all positions for association with host control (fig. S6 and table S6). Notably, position 97 in HLA-B was more significant (omnibus $P = 4 \times 10^{-45}$) than any single SNP in the GWAS, and three amino acid positions (67, 70, and 97), all in HLA-B, showed much stronger associations than any single classical HLA allele, including $B^*57:01$ (Fig. 3A). Moreover, allelic variants at these positions were associated with substantial frequency differences between cases and controls (Fig. 3B). These results indicate that the effect of HLA-B on disease outcome could be mediated, at least in part, by these positions. These three amino acid positions are located in the peptide binding groove, which suggests that conformational differences in peptide presentation at these sites contribute to the protective or susceptible nature of the various HLA-B allotypes. Although both innate and adaptive mechanisms could be at play, the hypothesis that HLA affects peptide presentation and subsequent T cell functionality is supported by experimental data showing substantial functional differences between CTL targeting identical epitopes but restricted by different HLA alleles (25).

We next performed stepwise regression modeling and identified six residues as independent markers associated with durable control of HIV. These include Arg$^97$, Cys$^{67}$, Gly$^{62}$, and Glu$^{63}$, all in HLA-B; Ser$^{77}$ in HLA-A; and Met$^{304}$ in HLA-C, which collectively explain 20% of the observed variance (similar to the variance explained by the seven classical HLA alleles described above). With the exception of Met$^{304}$ in the transmembrane domain of HLA-C, these residues are all located in the MHC class I peptide binding groove, again suggesting that the binding pocket—and, by inference, the conformational presentation of class I-restricted epitopes—plays a key role in host control.

Having identified these amino acid positions as strong candidates to account for the SNP and HLA association signals in this study, we next investigated their effects on protection or risk, revealing allelic variants at these positions linked to both extremes (Table 2). HLA-B position 97 (omnibus $P = 4 \times 10^{-45}$), located at the base of the C pocket, has important conformational properties for peptide binding (26). Position 97 has six allelic variants: Protective haplotypes $B^*57:01$, $B^*27:05$, and $B^*14$ are uniquely defined by Val$^97$ (3% frequency in controls), Asn$^97$ (4%) and Trp$^97$ (3%), respectively; the other amino acids at this position (Ser, Thr, Arg) segregate on a diverse set of haplotypes. Ser$^97$ (27% frequency) lies on risk haplotypes $Cw^*07$, $B^*07$, and others, whereas Thr$^97$ (11%) lies on protective $B^*52$ (and others). Arg$^97$ is the most common amino acid (51%) and is carried by risk allele $B^*35$, among others. The importance of this amino acid position to host control is underscored by conditional analyses revealing significance when we adjust incrementally for Val$^97$ (omnibus test for position 97, $P = 3 \times 10^{-20}$), Asn$^97$ ($P = 2 \times 10^{-9}$), and Trp$^97$ ($P = 7 \times 10^{-5}$). Thus, at a single position within the peptide binding groove (position 97, C-pocket), discrete amino acids are associated with opposite disease outcomes, even after controlling for $B^*57$ and $B^*27$, alleles associated with host control.

We also found similar discordant associations for alleles at positions 67, 63, and 62 (Table 2), all of which line the α1 helix along the peptide binding groove and help shape the B-pocket (Fig. 4). At position 67 (omnibus $P = 2 \times 10^{-42}$), risk haplotypes $B^*35$ and $B^*07$ carry aromatic residues Phe$^{67}$ and Tyr$^{67}$, respectively, whereas protective $B^*57:01$, $B^*27:05$, and $B^*14$ alleles carry sulfur-containing residues Met$^{67}$ or Cys$^{67}$. Position 62 ($P = \ldots$)
5 × 10^{−27}) is biallelic (Arg/Gly) with the Gly^{62} allele segregating with protective alleles B^{*57:01} and B^{*58} (1% frequency, OR = 1.7, P = 0.2). Adjacent position 63 (P = 9 × 10^{−16}) is also biallelic (Glu/Asn) with Glu^{65} appearing in complete LD (D = 1) with B^{*57:01}, B^{*27:05}, and B^{*52}. In contrast, at this position the risk alleles B^{*07} (14% frequency, OR = 0.5, P = 1 × 10^{−7}) and B^{*35} both carry Asn^{63}. Position 70 (omnibus P = 3 × 10^{−39}) accommodates four alleles that are tightly coupled with positions 67 and 97: Ser^{70} appears exclusively with Met^{67} (which defines B^{*57} and B^{*58}), Gln^{70} with Tyr^{67}, and Lys^{70} with Asn^{97} (B^{*27}). Hence, these data create a consistent and parsimonious model that can explain the associations of classical HLA-B alleles by specific amino acids lining the binding groove (and residues tightly coupled to them), which are expected to have an impact on the three-dimensional structure of the peptide-MHC complex.

To further investigate the role of individual amino acid positions in HLA-B, we implemented a permutation procedure to assess how consistent the above observations are with a null model in which there is no relation between amino acids at a particular position and host control (4). The results of this procedure provided evidence that multiple amino acid positions in the peptide binding groove are indeed associated with host control (Table S7), including positions 62, 63, 67, 70, and 97, thus providing a structural basis for the effect of HLA-B on host control (Fig. 4).

Within HLA-A position 77, which lies on the α helix contributing to the F-pocket, we identified a weaker but still significant association (omnibus P = 3 × 10^{−6}). Ser^{77} (6% frequency, OR = 2.0, P = 2 × 10^{−6}) is carried by only two HLA-A alleles (joint r^2 = 1): A^{*25} (2.4% frequency, OR = 2.6, P = 1 × 10^{−5}) and A^{*32} (3.2%, OR = 1.6, P = 0.02). Given its location and earlier association evidence for the A10 supertype (27), HLA-A could play a role in host control, although the evidence is not as strong as for HLA-B.

The signals within HLA-C are less straight-forward to interpret. Position 304 is a biallelic variant (Val/Met) located in the transmembrane domain (Met^{304}, 28% frequency, OR = 2.3, P = 7 × 10^{−23}). Met^{304} is in moderate LD (r^2 = 0.5) with rs9264942, which is known to be associated with HLA-C expression levels (28). Addition of this SNP to a multivariate model of all six amino acids is marginally significant (P = 0.013) but eliminates the effect of Met^{304} (P = 0.06). Similarly, addition of rs9264942 to a multivariate model of all seven independent classical HLA alleles is also significant (P = 2 × 10^{−4}) but eliminates the effect of Cw^{*07} (P = 0.08). These observations make it difficult to determine the extent to which epitope presentation in the HLA-C peptide binding pocket is important for host control. Thus, rs9264942 could be a proxy for not only many protective and risk HLA alleles (predominantly at HLA-B), but also for an independent effect on HLA-C gene expression, differentially affecting the response to HIV (29).

We next evaluated associations for the SNPs in the MHC, classical HLA alleles, and amino acids in a second independent cohort of untreated HIV-infected persons from Switzerland (Fig. S7 and tables S8 and S9) (4), in whom virus load set point was measured as a quantitative trait. Allelic variants at positions 67, 70, and 97 were also associated with highly significant differences in virus load set point in this second cohort (Fig. 3C). The effect estimates of all variable amino acids in HLA-B (r^2 > 0.9) and, to a lesser degree, those in HLA-C (r^2 > 0.8) in that cohort are in excellent agreement (Figs. S8 and S9). As before, position 97 in HLA-B is the most significant association (omnibus P = 1 × 10^{−13}). The HLA-A associations (A^{*25} or Ser^{77}) did not replicate, which reduces the likelihood that HLA-A plays a major role in host control.

In the African American sample (Fig. S10), the most significant HLA allele association was observed for two-digit B^{*57} (OR = 5.1, P = 1.7 × 10^{−21}) and four-digit B^{*57:03} (OR = 5.1,
$P = 2.8 \times 10^{-17}$; tables S10 and S11), consistent with previous studies (11-13). Position 97 in HLA-B (omnibus $P = 2 \times 10^{-25}$) is again the most significant amino acid (table S12). The consistency of these results demonstrates that imputation and association testing at amino acid resolution in multiple ethnicities can resolve disparate SNP associations in the MHC and help with fine-mapping of classical HLA associations.

Altogether, these results link the major genetic impact of host control of HIV-1 to specific amino acids involved in the presentation of viral peptides on infected cells. Moreover, they reconcile previously reported SNP and HLA associations with host control and lack of control to specific amino acid positions within the MHC class I peptide binding groove. Although variation in the entire HLA protein is involved in the differential response to HIV across HLA allotypes, the major genetic effects are condensed to the positions highlighted in this study, indicating a structural basis for the HLA association with disease progression that is probably mediated by the conformation of the peptide within the class I binding groove. The most significant residue, position 97 in the floor of the peptide binding groove of HLA-B, is associated with the extremes of viral load, depending on the expressed amino acid. This residue has been shown to have important conformational properties that affect epitope-contacting residues within the binding groove (26, 30) and has also been implicated in HLA protein folding and cell-surface expression (31).

Although the main focus of this study was on common sequence variation, it remains an open question as to the role of variants outside the MHC and the contribution of epistatic effects and epigenetic regulation. Additional factors also contribute to immune control of HIV, including fitness-altering mutations, immuno-regulatory networks, T cell help, thymic selection, and innate effector mechanisms such as killer cell immunoglobulin-like receptor recognition (23), some of which are influenced by the peptide-HLA class I complex. However, the combination and location of the significant amino acids defined here are most consistent with the genetic associations observed being modulated by HLA class I restricted CD8+ T cells. These results implicate the nature of the HLA-viral peptide interaction as the major genetic factor modulating durable control of HIV infection and provide the basis for future studies of the impact of HLA-peptide conformation on immune cell induction and function.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

This work was made possible through a generous donation from the Mark and Lisa Schwartz Foundation and a subsequent award from the Collaboration for AIDS Vaccine Discovery of the Bill and Melinda Gates Foundation. This work was also supported in part by the Harvard University Center for AIDS Research (grant P-30-AI06354); University of California San Francisco (UCSF) Center for AIDS Research (grant P-30 AI27763); UCSF Clinical and Translational Science Institute (grant UL1 RR024131); Center for AIDS Research Network of Integrated Clinical Systems (grant R24 AI067039); and NIH grants AI28568 and AI030914 (B.D.W.); AI087145 and K24AI069994 (S.G.D.); AI069513, AI34835, AI069432, AI069423, AI069477, AI069501, AI069474, AI069428, AI69467, AI069415, AI22782, AI27661, AI25859, AI28568, AI30914, AI069495, AI069471, AI069532, AI069452, AI069450, AI069556, AI069484, AI069472, AI34853, AI069465, AI069511, AI38844, AI069424, AI069434, AI46370, AI68634, AI069502, AI069419, AI068636, and RR024975 (AIDS Clinical Trials Group); and AI077505 and MF071205 (D.W.H.). The Swiss HIV Cohort Study is supported by the Swiss National Science Foundation (SNF grants 33CSC0-108787 and 310000-110012). S. Ripke acknowledges support from NIH/National Institute of Mental Health (grant MH085520). This project has been funded in whole or in part with funds from National Cancer Institute/NIH (grant HHSN261200800001E to M. Carrington). The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.
References and Notes

4. See supporting online material on Science Online for detailed background on the analyses that we performed.
33. Single-letter abbreviations for the amino acid residues referred to in Table 2 are as follows: Glu E, Phe F, Gly G, Lys K, Met M, Asn N, Gln Q, Arg R, Ser S, Val V, Trp W, Tyr Y.
Fig. 1.
Genome-wide association results in the European sample. (A) Manhattan plot of 1.3 million autosomal SNPs. Only SNPs in the MHC on chromosome 6 reach genome-wide significance, indicated by the horizontal dotted line ($P < 5 \times 10^{-8}$). Red and blue colors alternate between chromosomes. (B) Quantile-quantile plot of the association results with (black) and without (blue) SNPs in the extended MHC and the CCR5-CCR2 locus, indicating that the detectable effect is entirely attributable to these two loci. The red line denotes the expected distribution under the null hypothesis of no effect. (C) Distribution of the genotype protective score, defined as the total number of alleles associated with host control at the four independent SNPs in the MHC and the variants at CCR5-CCR2, showing marked differences in controllers (orange) and progressors (blue). In aggregate, these variants explain 23% of the observed variance of durable host control.
Fig. 2.
Imputation quality of classical HLA alleles in the European sample. (A) Concordance between imputed (y-axis) and observed (x-axis) frequencies of classical HLA types in 371 HIV-1 controllers with four-digit HLA types obtained through Sanger sequencing. (B) Positive predictive value, sensitivity, and genotype correlation ($r^2$) with typed alleles as a function of the observed frequency.
Fig. 3.
Associations at amino acids in HLA-B in the European sample. (A) Association results for all variable amino acid positions, as calculated by the omnibus test. Colors denote conventional pocket positions. P values for significant classical HLA-B alleles are shown for comparison. (B) Marked allele frequency differences between controllers and progressors for amino acids at positions 67, 70, and 97. Numbers above the bars indicate odds ratios (values >1 indicate a protective effect). (C) Associations between allelic variants at amino acid positions 67, 70, and 97 and quantitative virus load set point in the independent Swiss HIV cohort study. Effect estimates (beta coefficients from a linear-regression model) are given in log_{10} units of virus load set point. P values refer to the omnibus test for association at each position. Error bars indicate the standard error of the beta coefficient.
Fig. 4.
Three-dimensional ribbon representation of the HLA-B protein based on Protein Data Bank entry 2bvp (30), highlighting amino acid positions 62, 63, 67, 70, and 97 lining the peptide binding pocket. The peptide backbone of the epitope is also displayed. This figure was prepared with UCSF Chimera (32).
Table 1

Association results for the independent SNPs in the MHC identified with stepwise regression in the European and African American samples. The odds ratio and frequency is given for the A1 allele, where OR > 1 indicates a protective effect. Odds ratios and P values were computed for univariate and multivariate regression models. C, cytosine; G, guanine; T, thymine; A, adenine.

<table>
<thead>
<tr>
<th>SNP</th>
<th>A1</th>
<th>A2</th>
<th>Frequency in controllers</th>
<th>Frequency in progressors</th>
<th>Univariate OR</th>
<th>P value</th>
<th>Multivariate OR</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9264942</td>
<td>C</td>
<td>T</td>
<td>0.595</td>
<td>0.336</td>
<td>2.9</td>
<td>2.8 x 10^{-35}</td>
<td>2.1</td>
<td>6.3 x 10^{-16}</td>
</tr>
<tr>
<td>rs4418214</td>
<td>C</td>
<td>T</td>
<td>0.240</td>
<td>0.075</td>
<td>4.4</td>
<td>1.4 x 10^{-34}</td>
<td>1.8</td>
<td>4.9 x 10^{-4}</td>
</tr>
<tr>
<td>rs2395029</td>
<td>G</td>
<td>T</td>
<td>0.139</td>
<td>0.032</td>
<td>5.3</td>
<td>9.7 x 10^{-26}</td>
<td>2.1</td>
<td>3.5 x 10^{-4}</td>
</tr>
<tr>
<td>rs3131018</td>
<td>C</td>
<td>A</td>
<td>0.777</td>
<td>0.625</td>
<td>2.1</td>
<td>4.2 x 10^{-16}</td>
<td>1.5</td>
<td>1.2 x 10^{-5}</td>
</tr>
<tr>
<td>African American</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2523608</td>
<td>G</td>
<td>A</td>
<td>0.522</td>
<td>0.326</td>
<td>2.6</td>
<td>8.9 x 10^{-20}</td>
<td>2.3</td>
<td>3.7 x 10^{-15}</td>
</tr>
<tr>
<td>rs2555221</td>
<td>T</td>
<td>G</td>
<td>0.264</td>
<td>0.137</td>
<td>2.7</td>
<td>3.5 x 10^{-14}</td>
<td>1.9</td>
<td>2.1 x 10^{-6}</td>
</tr>
<tr>
<td>rs2523590</td>
<td>C</td>
<td>T</td>
<td>0.300</td>
<td>0.164</td>
<td>2.4</td>
<td>1.7 x 10^{-13}</td>
<td>2.3</td>
<td>1.2 x 10^{-12}</td>
</tr>
<tr>
<td>rs9262632</td>
<td>G</td>
<td>A</td>
<td>0.097</td>
<td>0.034</td>
<td>3.1</td>
<td>1.0 x 10^{-8}</td>
<td>2.2</td>
<td>2.8 x 10^{-4}</td>
</tr>
</tbody>
</table>
Table 2

Haplotypes defined by the four independent SNPs, classical HLA alleles, and amino acids associated with host control in the European sample. Haplotypes are ordered by the estimated odds ratio, where the most common haplotype was taken as reference (OR = 1). P values are for each haplotype tested against all other haplotypes. Only haplotypes with >1% frequency are listed, accounting for >85% of haplotype diversity. HLA-A alleles were excluded to limit the number of haplotypes. See (33).

<table>
<thead>
<tr>
<th>rs3131018</th>
<th>HLA-C</th>
<th>rs9264942</th>
<th>HLA-B</th>
<th>rs4418214</th>
<th>rs2395029</th>
<th>Frequency</th>
<th>OR</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classical</td>
<td>304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td>B*57:01</td>
<td>G</td>
<td>E</td>
<td>M</td>
<td>S</td>
<td>V</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td>B*52:01</td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>C</td>
<td>B*27:05</td>
<td>R</td>
<td>E</td>
<td>C</td>
<td>K</td>
<td>N</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>Cu*08:02</td>
<td>M</td>
<td>C</td>
<td>B*14:02</td>
<td>R</td>
<td>N</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>N</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>C</td>
<td></td>
<td>R</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>N</td>
<td>C</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>C</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>S</td>
</tr>
<tr>
<td>A</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>T</td>
<td></td>
<td>R</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>N</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>T</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>C</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>A</td>
<td>V</td>
<td>T</td>
<td></td>
<td>R</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>C</td>
<td>Cu*07:01</td>
<td>V</td>
<td>T</td>
<td>B*08:01</td>
<td>R</td>
<td>N</td>
<td>F</td>
<td>N</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>T</td>
<td></td>
<td>R</td>
<td>N</td>
<td>Y</td>
<td>Q</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>Cu*07:02</td>
<td>V</td>
<td>T</td>
<td>B*07:02</td>
<td>R</td>
<td>N</td>
<td>Y</td>
<td>Q</td>
</tr>
<tr>
<td>A</td>
<td>V</td>
<td>T</td>
<td></td>
<td>R</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>A</td>
<td>V</td>
<td>T</td>
<td></td>
<td>R</td>
<td>N</td>
<td>F</td>
<td>N</td>
<td>R</td>
</tr>
</tbody>
</table>