IMMUNOLOGY

Impact of HLA class I functional divergence on HIV control

Mathias Viard[1]†, Colm O'hUigin[1]†, Yuko Yuki[1], Arman A. Bashirova[1], David R. Collins[2], Jonathan M. Urbach[2], Steven Wolinsky[3], Susan Buchbinder[4,5], Gregory D. Kirk[6], James J. Goedert[7], Nelson L. Michael[8], David W. Haas[9], Steven G. Deeks[10], Bruce D. Walker[2,11], Xu Yu[2], Mary Carrington[1,2]*

Heterozygosity of *Human leukocyte antigen* (*HLA*) class I genes is linked to beneficial outcomes after HIV infection, presumably through greater breadth of HIV epitope presentation and cytotoxic T cell response. Distinct allotype pairs, however, differ in the extent to which they bind shared sets of peptides. We developed a functional divergence metric that measures pairwise complementarity of allotype-associated peptide binding profiles. Greater functional divergence for pairs of HLA-A and/or HLA-B allotypes was associated with slower AIDS progression and independently with enhanced viral load control. The metric predicts immune breadth at the peptide level rather than gene level and redefines *HLA* heterozygosity as a continuum differentially affecting disease outcome. Functional divergence may affect response to additional infections, vaccination, immunotherapy, and other diseases where *HLA* heterozygote advantage occurs.

It has been over two decades since a robust demonstration of *Human leukocyte antigen* (*HLA*) class I heterozygous advantage was documented for people living with HIV (PLWH) (*1*), in which progression to AIDS after seroconversion was lengthened in proportion to the number of heterozygous class I gene loci. The blunt assignment of heterozygosity, however, does not account for the degree of functional similarity between pairs of allotypes encoded by a given *HLA* genotype. A larger repertoire of HIV peptides presented by two allotypes that bind very distinct types of peptides may improve immune surveillance relative to a smaller repertoire contributed by functionally similar allotypes. In support of this model, *HLA* sequence divergence caused by radical substitutions has been shown to be selectively favored (*2*), and indeed, has been proposed as a source of fitness differences among heterozygous individuals (*3*), potentially accounting for maintenance of allelic lineages with up to 10% sequence difference. Metrics estimating genetic distance between *HLA* allelic pairs of a given genotype as a proxy for distinctiveness of peptide repertoires (*4*) have been used as an outcome predictor in infectious disease (*5*), transplantation (*6*, *7*), and certain cancer treatment studies (*8*). However, a more accurate measurement of the impact of a larger peptide repertoire on disease risk or outcome would employ metrics that are directly peptide-based.

We describe a novel metric to measure the proportion of overlap between HLA allotypic pairs based on empirically determined peptide repertoires for each allotype, which we term "functional divergence". Analyses of functional divergence in a large cohort of well-defined untreated HIV-1 seroconverters in which heterozygous advantage was previously observed shows that greater functional divergence between pairs of allotypes at *HLA-A* and/or *HLA-B* correspondingly associates with slower disease progression. The association was far stronger than either the dichotomous assignment of genetic zygosity or metrics of allelic genetic distances. The impact of functional divergence on HIV disease was robustly and consistently reproduced in an independent cohort consisting of HIV-1 elite controllers and noncontrollers, as well as in a cohort of >3500 individuals where longitudinal viral loads were available for analysis. The peptide-centric algorithm described herein holds an advantage over genetic-based approaches and demonstrates the sizeable impact of functional divergence at *HLA* class I on outcome to HIV-1 infection.

Effect of allelic homozygosity on HIV disease progression

Our analysis focused initially on a cohort of up to 870 seroconverters, 60% of whom were included in the prior study that established the protective effects of *HLA* class I heterozygosity (*1*). Reanalysis of the enlarged cohort strengthens the observation that homozygosity at *HLA* class I results in faster progression to AIDS, as indicated by progression to CD4 counts below 200 cells per ml (CD4 < 200) [hazard ratio (HR) = 1.53 and *P*-value = 6×10^{-4}] (Fig. 1).

Effects of individual *HLA* class I loci are strongest for homozygosity at *HLA-B* (HR = 1.82, *P* = 0.004), intermediate for *HLA-A* (HR = 1.56, *P* = 0.003), and lowest for *HLA-C* (HR = 1.43, *P* = 0.036). The tight linkage disequilibrium between the *HLA-B* and *-C* loci leads to difficulty in resolving effects for each locus independently, but neither homozygosity at *HLA-C* nor *HLA-B* alone showed a significant association with progression to CD4 < 200 (table S1), the latter trending in the same direction (HR = 1.36) in a particularly small number of individuals (N = 15).

Effect of genetic distance on HIV progression

Dichotomously categorizing individuals as genetic homozygotes versus heterozygotes does not account for distinctions between the two alleles (or their products) present at each heterozygous locus. Interallelic genetic distances at *HLA* class I loci [*HLA* evolutionary divergence (HED)] have been used as a proxy for differential peptide binding across pairs of allotypes encoded by the corresponding genotype (*4*). A negative correlation between *HLA-B* genetic distances and HIV viral load (VL) was observed in a large cohort of PLWH but the observed correlation was low (Kendall t coefficient < 0.08) (*5*). No correlation was observed between HIV VL and *HLA-A* genetic distance and a positive correlation was found with *HLA-C* distances (*5*). We find no effect of the HED genetic distance metric for *HLA-A*, *-B*, or *-C* individually or combined on progression to low CD4 counts in our seroconverter cohort in analyses including or excluding genetic homozygotes (fig. S1). Analyses of other survival outcomes, including time to an AIDS-defining illness or death after seroconversion, also failed to show significant effects of HED. These data suggest that the relatively weak effects of HED on HIV VL cannot be mirrored in analyses of HIV disease progression.

Impact of HLA class I supertypes on HIV progression

Global analyses of various peptide pools eluted from class I allotypes or binding assays of peptides exogenously-bound to specific HLA allotypes have identified broad peptide motifs associating with groups of *HLA-A* and *HLA-B* alleles, known as supertypes (*9*). Similarity of preferred amino acids at position 2 or the C terminus of the peptide, the residues anchoring the peptide into the groove of the class I molecule, constitutes the primary basis for broadly grouping the alleles into 6 supertypes for *HLA-A* and

[1]Basic Science Program, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA and Laboratory of Integrative Cancer Immunology, Center for Cancer Research, National Cancer Institute Bethesda, MD, USA. [2]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [3]Division of Infectious Diseases, Department of Medicine, The Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [4]Bridge HIV, San Francisco Department of Public Health, San Francisco, CA, USA. [5]Department of Medicine, Epidemiology and Biostatistics, University of California, San Francisco, CA, USA. [6]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [7]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [8]US Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, MD, USA. [9]Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. [10]Department of Medicine, University of California, San Francisco, CA, USA. [11]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.
***Corresponding author. Email: carringm@mail.nih.gov**
†These authors contributed equally to this work.

6 supertypes for *HLA-B* (*9*). As the supertypes are based on peptide binding data (of various types and sources) rather than genetic distances of allotypic pairs, we tested for an effect of supertype zygosity on progression to CD4 < 200 but observed no significant effect in a comparison of heterozygotes (across supertypes) versus homozygotes (within supertypes) at *HLA-A* (HR = 1.09, $P = 0.47$) or *HLA-B* (HR = 1.21, $P = 0.14$) on progression to CD4 < 200, even though these analyses included *HLA* allelic homozygotes. We next generated further subdivisions of HLA-A and HLA-B supertypes based on physico-chemical properties (size, charge, and hydrophobicity) of the motif-defining amino acids, which resulted in 19 distinct supertypes for HLA-B and 10 for HLA-A (see Materials and Methods). The increased specificity of the novel supertype definition indicated significant heterozygosity advantage across supertypes at both *HLA-A* (HR = 1.30, $P = 0.048$) and *HLA-B* (HR = 1.76, $P = 0.0028$). This significance is not sustained, however, upon the exclusion of allelic homozygotes at *HLA-A* (HR = 0.88, $P = 0.60$) or *HLA-B* (HR = 1.52, $P = 0.32$). Overall, these data suggest that accounting for greater scrutiny of peptide properties may be necessary in order to detect differential effects of specific pairs of HLA allotypes on HIV disease outcomes.

Determination of functional divergence through submotif overlap

A more in-depth analysis of complementarity between pairs of allotypes encoded by genetic heterozygotes at *HLA* class I was attempted using the well-standardized HLAthena dataset of peptides eluted from class I allotypes that were individually expressed on the otherwise *HLA* class I negative .221 cell line (*10*). Eluted peptides derived from each of 95 HLA allotypes ranged in length from 7 to 42 amino acids and peptide repertoire sizes varied from 692 to 4033 per allotype. To simplify analyses, we focused solely on 9-mer peptides, which account for about 60% of the characterized HLA class I bound peptides. Sarkizova *et al.* (*10*) performed a sequence-based clustering analysis of peptides eluted from each class I allotype, which we used to derive a specific logogram for each identified cluster (Fig. 2A). This approach allowed us to focus on aggregate properties of the peptides (i.e., those defining a cluster) presented by a given allotype rather than minor differences across peptides (i.e., variants within a cluster).

Only those amino acids with a positional entropy lower than 0.1 (i.e., high selectivity for one amino acid at a given position) were retained in the analysis in order to reduce noise contributed at positions with high entropy (Fig. 2A). The resulting simplified logograms, termed submotifs, allow compact representation of the different families of peptides that a given allotype can present. This approach condensed a universe of 111,898 nonamer peptides presented by 95 different HLA class I allotypes into 382 distinct submotifs, with a range of 2 to 33 9-mer clusters per HLA allotype. We compared submotifs of the allotypic pairs encoded by a given genotype at each class I locus and calculated the proportion of submotifs that differ between that pair of allotypes, allowing quantification of the functional divergence and redundancy between pairs of allotypes (Fig. 2B). Peptides behaving as singletons (shown in black in Fig. 2A) were included as a separate subgroup in order to avoid a complete lack of overlap between any given allotypic pair. By including this subgroup of singletons, the metric accounts for the total number of submotifs associated with each pair of allotypes.

Individuals in the cohort were then ranked by the extent of submotif redundancy between pairs of allotypes encoded at each of their *HLA* loci, and disease outcome was compared between the top- and bottom-ranked groupings of individuals (defined by appropriate quantiles) for each locus (Fig. 2C and Materials and Methods). The comparison of individuals from top and bottom quantiles is advantageous for delineating effects of functional divergence from noise. Functional divergence and its effects are continuous, varying according to the allotype combinations involved and not by effects of individual allotypes. That is, a single allotype may exist at both extremes of the functional divergence spectrum depending on the alternative allotype with which it is paired. Effects of intermediate levels of functional divergence can be masked by numerous other genetic, viral, or environmental factors that affect HIV outcome. Consequently, in order to determine whether functional divergence has an impact on disease outcome, we initially compared groups that differ substantially in divergence. Correction for the well-documented (including the cohorts used in this study) effects on progression of *HLA-B*57*, *-B*27*, certain *-B*35* alleles (*11*), as well as for racial background, was consistently applied in all analyses described.

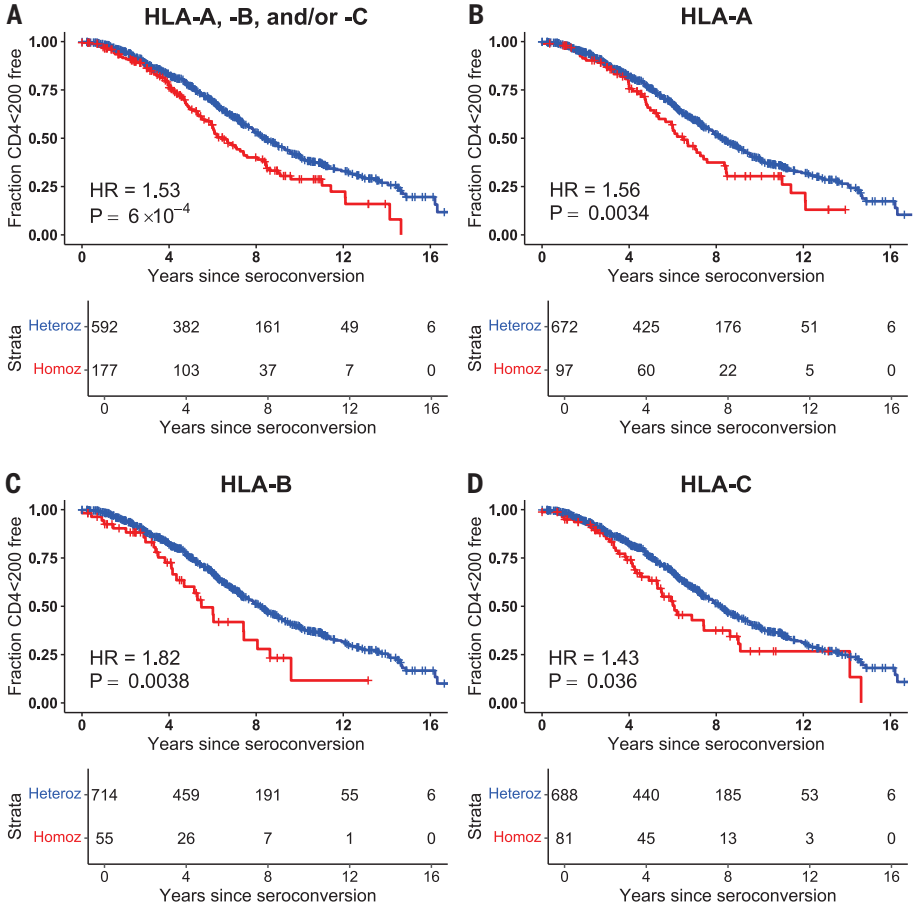
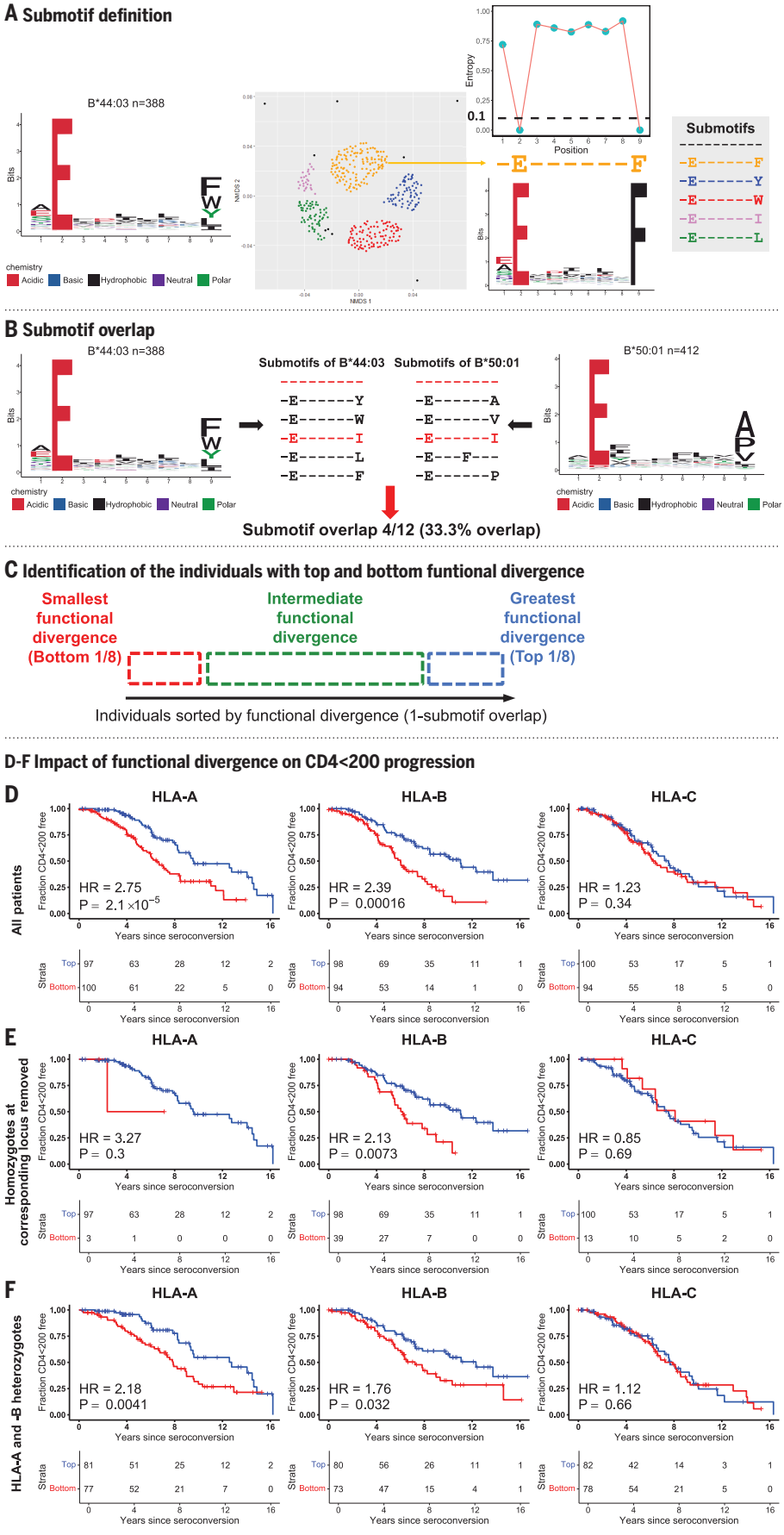


Fig. 1. Effect of genetic homozygosity on progression to CD4 < 200 after HIV infection. Progression to CD4 < 200 among (**A**) individuals homozygous at any *HLA-A*, *-B*, and/or *-C* locus/loci (red curve) compared with heterozygotes at all three loci (blue curve). (**B**) *HLA-A* homozygotes (red) compared with *HLA-A* heterozygotes (blue). (**C**) *HLA-B* homozygotes (red) compared with *HLA-B* heterozygotes (blue). (**D**) *HLA-C* homozygotes (red) compared with *HLA-C* heterozygotes (blue). All analyses were corrected for *HLA-B*57*, *B*27*, *B*35Px*, and racial background.

Fig. 2. Submotif overlap and its effect on progression to CD4 < 200 after HIV infection. (**A**) Representative motif logogram and nonmetric ordination plot as generated for each of 95 HLA allotypes studied previously (*10*). Submotifs corresponding to each cluster were determined by retaining only the amino acids at each position with an entropy lower then 0.1 (shown for B*44:03). (**B**) Determination of submotif overlap corresponding to the two HLA allotypes encoded by each genotype. Red indicates identical whereas black indicates distinct submotifs across the two allotypes. (**C**) Determination of the range in functional divergence from smallest to greatest at each *HLA* locus across individuals. Comparisons were made between individuals falling into the top-ranked versus bottom-ranked quantiles (shown in octiles). (**D** to **F**) Progression to CD4 < 200 was compared between individuals in the bottom octile (red curves) versus top octile (blue curves) of functional divergence at each HLA locus (D) among all individuals, (E) after removing genetic homozygotes corresponding to each individual locus, but keeping the same octile definition as in (D) (diminishing the number of individuals in the bottom octile), or (F) after removing genetic homozygotes at *HLA-A* and/or *HLA-B* prior to selection of the bottom and top octiles. Analyses were corrected for *HLA-B*57*, *B*27*, *B*35Px*, and racial background.



Effect of functional divergence on HIV disease progression

Analyses measuring the impact of functional divergence on HIV disease outcome were performed comparing extreme octiles as this quantile size was determined to represent a balance between the magnitude and significance of resultant hazard ratios (Materials and Methods, table S2). Low functional divergence (i.e., high submotif redundancy) associated with faster progression to low CD4 counts (Fig. 2D), where significant differences between the top versus bottom ranked octiles for HLA-A (HR = 2.75, $P = 2.1 \times 10^{-5}$) and HLA-B (HR = 2.39, $P = 1.5 \times 10^{-4}$) were observed. The lack of a significant effect of functional divergence for HLA-C (HR = 1.23, $P = 0.34$) was consistent with prior findings of reduced association between genetic homozygosity at *HLA-C* (Fig. 1) and disease progression, which likely reflected homozygosity at the closely linked *HLA-B* locus.

To eliminate the possibility that genetic homozygosity is predominantly responsible for the more rapid disease progression among individuals in the bottom octile (red curves in Fig. 2D), individuals homozygous at each respective locus were removed from the top and bottom octiles defined in Fig. 2D, diminishing the power of the analysis (Fig. 2E). Only three *HLA-A* genetic heterozygotes remained in the bottom functional divergence octile after removal of *HLA-A* homozygotes, resulting in a lack of significance (HR = 3.27, $P = 0.30$), but both the trend and significance were well conserved for HLA-B (HR = 2.13, $P = 0.0073$) after exclusion of *HLA-B* genetic homozygotes. Once again, functional divergence at HLA-C had no significant association with HIV disease progression (HR = 0.85, $P = 0.69$) after removal of *HLA-C* genetic homozygotes. Genetic homozygotes at *HLA-C* were not removed from further analyses given the apparent lack of an effect of HLA-C functional divergence on HIV disease progression. Analyses in which the top and bottom octiles were fully repopulated after exclusion of homozygotes at *HLA-A* and/or *HLA-B* (Fig. 2F) restored numbers and power, and significant protective effects of greater functional divergence were observed for HLA-A (HR = 2.18, $P = 0.0041$) and HLA-B (HR = 1.76, $P = 0.032$), but not for HLA-C (HR = 1.12, $P = 0.66$). We conclude that greater functional divergence at both HLA-A and HLA-B independently confer substantial protection against HIV disease progression.

Combination of HLA-A and HLA-B functional divergence on HIV disease progression

The linked *HLA-A* and *-B* loci are separated by roughly 1Mb and sufficient levels of recombination between the two loci has occurred (*12*) such that independent effects of each locus on HIV disease can be detected. Figure 2, D to F, indicated that functional divergence levels at both HLA-A and -B individually have an impact on HIV disease progression, implying that their combined effects might be more influential than considering each separately. We tested the combined effects of functional divergence at HLA-A and -B by comparing individuals with the greatest functional divergence (top octile) to those with the least functional divergence (bottom octile) at HLA-A and/or HLA-B based on genotypes for each person at these loci. Besides being biologically relevant, accounting for combined effects of these two loci enhances analytical power as a result of the enlarged sizes of the two comparison groups. In the first set of analyses, individuals having HLA-A and -B allotypic pairs in which one ranked in the top octile and the other ranked in the bottom octile were excluded. HLA-A and HLA-B combined functional divergence based on the following groups were then compared: the combined grouping of those whose HLA-A divergence was in the top octile and HLA-B was intermediate; plus those whose HLA-B divergence was in the top octile and HLA-A was intermediate; plus those whose HLA-A and HLA-B divergences were both in the top octile versus the combined grouping of those whose HLA-A divergence was in the bottom octile and HLA-B was intermediate; plus those whose HLA-B divergence was in the bottom octile and HLA-A was intermediate; plus those whose HLA-A and HLA-B divergences were both in the bottom octile. A particularly robust protective effect of greater functional divergence was observed when considering the combination of both loci (HR = 2.87, $P = 2.4 \times 10^{-8}$) (Fig. 3A), as opposed to HLA-A or HLA-B individually (Fig. 2D) and is among the strongest documented HLA effects on HIV disease progression. A second approach in which HLA-A and HLA-B functional divergence values were first summed followed by selection of the top and bottom octiles of divergence (as opposed to first selecting individuals from each of the extreme octiles for HLA-A and HLA-B separately and then combining them into the respective top/bottom divergence groupings) produced similar results but with less statistical support (fig. S2).

The impact of functional divergence on AIDS progression was strongest and most significant for early outcomes, including time to CD4 < 200 and time to AIDS93 (defined by the CDC as CD4 < 200 or an AIDS-defining illness, whichever is diagnosed first) (*13*), but showed less significant effects for the later outcomes, AIDS87 (AIDS-defining illness) (*14*), or death (table S3). These data reflect the same pattern observed for genetic zygosity. Notably, the effect size (HR) was always greater for functional divergence than for genetic zygosity across each outcome (table S3).

To ascertain whether the protective effect of greater HLA functional divergence between pairs of HLA allotypes tends to be continuous, we subdivided the cohort into further categories. Inclusion of individuals with intermediate functional divergence scores at both HLA-A and HLA-B indicates that this grouping progresses to loss of CD4 counts at a rate intermediate to the two extreme quantiles, as expected (mean HR with each stepwise category = 1.57, $P_{trend} = 2.2 \times 10^{-7}$; Fig. 3B). Further subdivisions splitting both the top and bottom groupings according to whether only one or both HLA-A and HLA-B were responsible for their respective rankings resulted in a total of 5 distinct categories. These groups consisted of individuals ranking as follows: (A) the top group (most divergent) at both HLA-A and HLA-B, (B) the top group at only one and intermediate at the other, (C) intermediate at both; (D) the bottom group (least divergent) at only one and intermediate at the other, and (E) the bottom group at both. These progressive rankings revealed a perfectly continuous effect of functional divergence on HIV disease progression (mean HR with each decreasing stepwise category = 1.49, $P_{trend} = 7.6 \times 10^{-8}$, Fig. 3C), despite low numbers in each of the two extreme groupings (i.e., greatest divergence at both HLA-A and HLA-B and least divergence at both HLA-A and HLA-B). Individuals having HLA-A and -B genotypes in which one ranked in the top octile and the other ranked in the bottom octile exhibit a mixed phenotype whereby upon an initial slow progression to CD4 < 200, subsequently, the rate of progression substantially increased (purple curve, fig. S3).

Significance of disease associations with level of functional divergence was maintained after removal of genetic homozygotes at HLA-A and -B from each of the respective analyses (Fig. 3, D to F), illustrating the impact of this property beyond genetic zygosity status alone. The strength of the effects upon exclusion of genetic homozygotes was equal to or slightly stronger than those when genetic homozygotes were included in the analyses (compare HR of Fig. 3A, B, and C to 3D, E, and F, respectively).

Effect of functional divergence on HIV elite controller status and mean VL in untreated individuals

An independent replication cohort consisting of HIV elite controllers (those who maintain undetectable VL for at least one year; $n = 308$) and noncontrollers (those with viral VLs >10,000 copies per ml; $n = 1895$) was used to confirm findings observed in the seroconverter cohort. Functional divergence based on the five progressive categories shown in Fig. 3, C and F, was strongly predictive of outcomes in comparisons of elite controllers versus noncontrollers, either in the presence [mean odds ratio with each stepwise category (OR_{trend}) = 1.53, $P = 6.6 \times 10^{-6}$ or absence ($OR_{trend} = 1.43$, $P = 6.3 \times 10^{-6}$) of genetic homozygotes (Table 1). VL provides an additional means to assess functional divergence

by measuring the strength and significance of the regression estimate (change in mean VL per unit increase in functional divergence). An analysis of the five progressive categories assigned to 3563 individuals with longitudinal VL measurements—2156 of whom overlapped with the elite controller versus noncontroller cohort—showed that greater functional divergence significantly associated with reduction in VLs in both the presence (estimate $P = 2.5 \times 10^{-7}$) or absence (estimate $P = 4.2 \times 10^{-4}$) of homozygotes at *HLA-A* and/or *HLA-B* loci (Table 1). These data soundly confirm the substantial impact of functional divergence on HIV control.

Functional divergence based only on submotifs to which HIV peptides correspond

The 382 submotifs used in our analyses were derived from self peptides eluted from .221 cells expressing single allotypes of HLA class I (*10*), and some of these submotifs may not be relevant when considering HIV-derived peptides. We took several approaches to test this possibility, stringently restricting the analyses to individuals fully heterozygous at both HLA-A and -B. We tested whether functional divergence based only on the set of submotifs corresponding to the HIV peptidome derived from clade B and clade C HIV 2004 consensus sequences

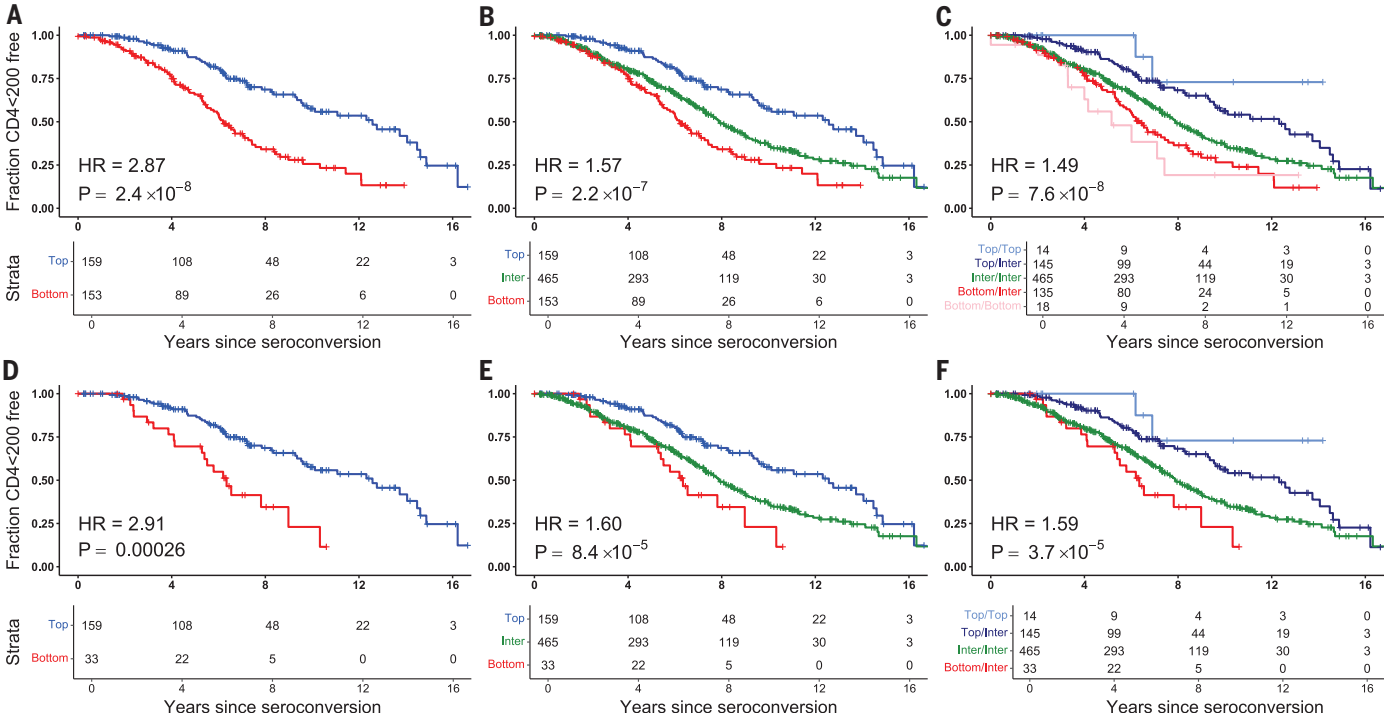


Fig. 3. Combined effect of HLA-A and HLA-B submotif overlap on progression to CD4 < 200. (**A**) Progression to CD4 < 200 among individuals in the bottom (red curve) versus top (blue curve) octile(s) of functional divergence at HLA-A and/or HLA-B. For individuals in whom only one of the two (HLA-A or HLA-B) was in the top octile, the other was necessarily in the intermediate range. (**B**) Individuals carrying intermediate levels of functional divergence at both *HLA-A* and *HLA-B* (green curve) were added to the analysis shown in (A). *P* = test for trend. (**C**) Individuals were further stratified as follows: bottom octiles at both HLA-A and HLA-B (pink curve); bottom octile at HLA-A or HLA-B and intermediate at the other (red curve); intermediate at both (green curve); bottom octile for only one and intermediate at the other (dark blue curve); top octiles at both (light blue curve). *P* = test for trend. The HR represents an average of the effect with each decreasing category. The analyses were performed among all individuals (A to C) or following removal of homozygotes at *HLA-A* and *HLA-B* (i.e., heterozygous at both *HLA-A* and *HLA-B*) (**D** to **F**). Analyses were corrected for *HLA-B*57*, *B*27*, *B*35Px*, and racial background.

Table 1. Effect of HLA-A and/or HLA-B submotif overlap on controller status and mean VL after HIV infection. Logistic regression of controller status (elite controller versus noncontroller) and linear regression of mean VL were used to estimate the effect of functional divergence at HLA-A and/or HLA-B structured in 5 categories as defined for Fig. 3C.

Study	Study Participants	Outcome	Modeling	N	Effect estimate†	*P*
Elite controllers (301) versus Noncontrollers (1844)	All	Controller status	Logistic regression	2145	1.53	6.6×10^{-6}
	HLA-A and *-B* genetic heterozygotes only			1759	1.43	6.3×10^{-4}
Individuals with mean VL (3462)	All	Mean VL	Linear regression	3462	0.14	2.5×10^{-7}
	HLA-A and *-B* genetic heterozygotes only			2865	0.11	4.2×10^{-4}

†The effect estimate corresponds to the OR_{trend} for the logistic regression and the regression coefficient for the linear regression. Analyses were adjusted for *HLA-B*57*, *B*27*, *B*35Px*, and racial background.

present in the Los Alamos National Laboratory (LANL) database (www.hiv.lanl.gov/content/sequence/HIV/CONSENSUS/Consensus.html) would produce similar results to those observed when the full set of 382 submotifs were employed. All possible overlapping nonamers derived from HIV consensus sequences (n = 5042 from the two clades combined) represented 289 of the 382 submotifs derived from self. Thus, 93 submotifs to which none of the HIV nonamer peptides corresponded were excluded from the analysis. Calculation of functional divergence using only the 289 submotifs corresponding to the HIV peptidome slightly improved the effect of functional divergence on HIV disease progression for HLA-B, and somewhat diminished the effect for HLA-A (fig. S4A). Thus, both a broad and more restricted calculation of HLA functional divergence reveals an impact of this property on HIV disease.

HIV sequences are highly variable, and the number of distinct nonamers greatly exceeds those derived from consensus HIV sequences. The 2021 LANL compendium of HIV sequences (www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2021compendium.html) contained a total of 1,933,474 unique nonamers, which corresponded to all but one of the 382 submotifs derived from self. We ranked submotifs based on the number of HIV nonamers corresponding to each submotif (ranging from 3 to 163,533). An iterative estimation of the effect on disease progression with stepwise elimination of submotifs represented by the fewest number of peptides was performed for each locus. Upon elimination of submotifs poorly represented among HIV nonamers at each HLA locus, functional divergence trended toward higher HR values (fig. S4B). This pattern may indicate improved accuracy of the estimation of functional divergence following elimination of redundant or irrelevant submotifs. Further elimination of submotifs over the course of 30 elimination steps eventually resulted in reduced HR values, presumably due to elimination of relevant submotifs (fig. S4B).

The inclusion of all potential HIV nonamers in these analyses is likely to somewhat diminish the accuracy in determining the strength of the effect of functional divergence on disease outcome as many of the nonamers included in the analyses do not represent true epitopes (i.e., cannot be processed and presented in vivo). To address this issue, we have used 2483 experimentally verified HIV nonamer epitopes recorded in the LANL database (www.hiv.lanl.gov/content/immunology/variants/ctl_variant.html), which represent 270 submotifs out of the 382 submotifs corresponding to self peptides. Although the LANL epitope data are incomplete and biased toward alleles known to have effects on natural history outcomes of HIV infection, our analyses of functional divergence based on the 270 epitope-relevant submotifs showed consistent differences in HR between the bottom and top octiles for both HLA-A and HLA-B (fig. S4C).

Overall, the use of pertinent, disease-specific submotifs may somewhat improve the detection of functional divergence on disease outcome, but when disease-specific submotifs are not defined or attainable, submotifs based on self peptides are nearly as effective.

Discussion

Protection against viral disease conferred by *HLA* class I heterozygosity is likely mediated by a broader peptide repertoire presented by distinct HLA allotypic combinations relative to the more limited repertoire presented by homozygotes. Previous reports have studied *HLA* class I zygosity dichotomously by comparing heterozygotes composed of any given pair of distinct alleles to homozygotes, providing evidence to support heterozygote advantage (*1*, *8*, *15*–*19*). This approach, however, ignores differences in the breadth of the peptide repertoire among heterozygotes, and any effect these differences may have on disease outcome, response to vaccination, or other clinically relevant phenotypes. We probed more deeply into the impact of complementarity across peptide binding repertoires between pairs of HLA class I allotypes, developing a metric termed functional divergence that takes into account HLA allotype-specific peptide binding data directly. We find that untreated PLWH who carry any given pair of HLA allotypes characterized by greater functional divergence in their peptide binding profiles (i) progress more slowly to AIDS, (ii) are more likely to be an elite controller, and (iii) sustain lower VL levels over time, even when restricting analyses to genetic heterozygotes only.

Properties of HLA class I that affect their ability to present a broader range of antigenic epitopes have been shown to associate with HIV control. Dependence on tapasin for peptide loading varies markedly across HLA allotypes (*20*–*26*). Those HLA class I allotypes that are less dependent on tapasin have been shown to elicit T cell responses with broader epitope specificity and to associate with protection against HIV disease (*25*). Likewise, greater functional divergence, by definition, extends the number of distinct peptides capable of being presented by allotypes encoded by the specific *HLA* genotypes carried in each individual. Properties of HLA allotypes/allotypic pairs, such as tapasin independence and functional divergence, that enhance the repertoire of antigen presentation decrease the chances of viral immune escape, likely explaining their observed associations with control of HIV in natural history cohorts. We predict that these properties will also influence efficacy of vaccination.

Previous studies have employed metrics such as Grantham distance to estimate genetic distance between *HLA* allelic pairs of each given genotype (*4*), an approach that has been a useful initial step in the study of HLA class I functional divergence. Studies of a large HIV cohort (>6000 individuals) indicated that greater genetic divergence between *HLA-B* allelic pairs of any given genotype significantly associated with decreased HIV VLs, but this same trend was not observed for *HLA-A* or *HLA-C* (*5*). A method based on standard peptide binding prediction algorithms, which are indirect and limited in that they ignore antigen processing effects on peptide availability, can somewhat improve the association between divergence and VLs relative to the method based solely on genetic distances (*5*). Although we find highly significant effects of functional divergence for HLA-B and HLA-A, particularly the combination of HLA-A and HLA-B, in analyses of three outcomes to HIV infection (disease progression in seroconverters, elite versus noncontrol of HIV, and mean VL over time), we found no evidence that genetic distance predicts outcome in any of these cohorts. This was true for analyses performed by dichotomizing the top one-fourth of the genetic divergence spectrum versus the remainder of the study participants, as was used in cancer immunotherapy studies (*8*) or using divergence as a continuous variable, as done in HIV studies (*5*). It is possible that our cohorts were not of sufficient size to detect the relatively weaker effect of genetic divergence compared with the more direct and sensitive approach used herein. The strength of the effect of functional divergence as measured using our metric was among the greatest observed in predicting HIV outcome, underscoring the advancement gained in measuring divergence using metrics employing endogenous peptides that have been processed and presented on the cell surface through physiological means.

Previous data have shown that HIV escape mutations occur most frequently at anchor residues (*27*) and these mutations were computationally predicted to severely affect binding of the peptide to the corresponding HLA allotype. A key feature of the metric described herein includes the grouping of peptides by motif similarity where motifs are primarily defined by the anchor residues and noise contributed by minor peptide variation is eliminated by applying low entropy cutoffs. Given that immune escape generally occurs at the same positions defining the submotif categories, it is not surprising that this metric of functional divergence is strongly predictive of HIV outcomes. In our cohort, simple quantifiers such as aggregate peptide repertoire sizes do not show a clear association with HIV progression in heterozygotes, most likely as a consequence of the large proportion of peptides in the repertoire that differ slightly in sequence but are functionally and immunologically

equivalent. Removal of this "noise" is necessary to enable focus on peptide groupings that contribute distinctly to immunological response.

Our findings in an extended cohort of HIV seroconverters confirm that genetic heterozygosity at both *HLA-A* and *HLA-B* associate with protection against HIV progression, but there is no clear effect of *HLA-C* zygosity. Likewise, greater functional divergence at both HLA-A and HLA-B independently or combined significantly associates with slower disease progression and there is no apparent effect of functional divergence at HLA-C. Comparison of individuals with the greatest functional divergence at both HLA-A and HLA-B to those with the lowest divergence levels shows the greatest differentiation in protection, indicating the importance of the two loci acting in unison. The superior ability of HLA-A and HLA-B to present a broad repertoire of peptides as compared with HLA-C was previously reported (*28*, *29*). Furthermore, HLA-A and HLA-B are thought to present a complementary set of peptides in which HLA-A is more fastidious at the F pocket and HLA-B is more fastidious at the B pocket (*29*), which may explain the strength of the effects reached when considering functional divergence at HLA-A and HLA-B combined. Disease-specific refinement of the submotif pool used to estimate functional divergence shows promise as an approach to improve accuracy to some extent but attaining the complete, unbiased disease-specific epitope pool will be challenging. Use of new mass spectrometry technologies to characterize HLA allotype-specific HIV epitopes (*30*) should provide a broader and less biased definition of true epitopes and lead to improved accuracy in tests of functional divergence. However, submotifs determined from the self peptide pool successfully detect the impact of functional divergence and can be applied across disease types.

Our metric is likely to be applicable to other viral diseases, particularly those that undergo immune escape through mutation of epitopes. It is unclear, however, whether the same metric will be predictive of disease outcome when the epitopes are much more similar to self, as in the case of cancer neoantigens where generally only a single amino acid change from self is often observed (*31*, *32*), possibly requiring a more granular approach to defining peptide overlap. Furthermore, the class of the neoantigens (e.g., genomic variants, transcriptomic variants, proteomic variants, or viral-derived variants) can vary depending on tumor type (*33*), which may also result in the necessity to adjust the properties of the metric. The work described herein represents a foundation for determining functional divergence that can be readily modified to suit the particular outcome in question. The metric may be applicable in predicting clinical outcomes such as efficacy of vaccination or immunotherapy as well as the natural history of disease pathogenesis.

REFERENCES AND NOTES

1. M. Carrington *et al*., *Science* **283**, 1748–1752 (1999).
2. A. L. Hughes, T. Ota, M. Nei, *Mol. Biol. Evol.* **7**, 515–524 (1990).
3. E. K. Wakeland *et al*., *Immunol. Res.* **9**, 115–122 (1990).
4. F. Pierini, T. L. Lenz, *Mol. Biol. Evol.* **35**, 2145–2158 (2018).
5. J. Arora *et al*., *Mol. Biol. Evol.* **37**, 639–650 (2020).
6. A. M. Daull *et al*., *Front. Immunol.* **13**, 841470–841470 (2022).
7. C. Féray *et al*., *Ann. Intern. Med.* **174**, 1385–1394 (2021).
8. D. Chowell *et al*., *Nat. Med.* **25**, 1715–1720 (2019).
9. J. Sidney, B. Peters, N. Frahm, C. Brander, A. Sette, *BMC Immunol.* **9**, 1 (2008).
10. S. Sarkizova *et al*., *Nat. Biotechnol.* **38**, 199–209 (2020).
11. M. P. Martin, M. Carrington, *Immunol. Rev.* **254**, 245–264 (2013).
12. M. Cullen, S. P. Perfetto, W. Klitz, G. Nelson, M. Carrington, *Am. J. Hum. Genet.* **71**, 759–776 (2002).
13. *MMWR Recomm. Rep.* **41**, 1–19 (1992).
14. CDC, *MMWR Suppl.* **36**, 1S–15S (1987).
15. C. L. Roark *et al*., *Hum. Immunol.* **83**, 730–735 (2022).
16. N. Shah *et al*., *Leukemia* **25**, 1036–1039 (2011).
17. J. Tang *et al*., *AIDS Res. Hum. Retroviruses* **15**, 317–324 (1999).
18. S. S. Wang *et al*., *Cancer Res.* **78**, 4086–4096 (2018).
19. P. C. Doherty, R. M. Zinkernagel, *Nature* **256**, 50–52 (1975).
20. R. Greenwood, Y. Shimizu, G. S. Sekhon, R. DeMars, *J. Immunol.* **153**, 5525–5536 (1994).
21. C. A. Peh *et al*., *Immunity* **8**, 531–542 (1998).
22. C. A. Peh, N. Laham, S. R. Burrows, Y. Zhu, J. McCluskey, *J. Immunol.* **164**, 292–299 (2000).
23. A. W. Purcell *et al*., *J. Immunol.* **166**, 1016–1027 (2001).
24. D. Zernich *et al*., *J. Exp. Med.* **200**, 13–24 (2004).
25. A. A. Bashirova *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28232–28238 (2020).
26. S. M. Rizvi *et al*., *J. Immunol.* **192**, 4967–4976 (2014).
27. J. M. Carlson *et al*., *J. Virol.* **86**, 13202–13216 (2012).
28. M. Rasmussen *et al*., *J. Immunol.* **193**, 4790–4802 (2014).
29. D. Di, J. M. Nunes, W. Jiang, A. Sanchez-Mazas, *Mol. Biol. Evol.* **38**, 1580–1594 (2021).
30. S. Sengupta *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2123406119 (2022).
31. T. N. Schumacher, R. D. Schreiber, *Science* **348**, 69–74 (2015).
32. L. P. Richman, R. H. Vonderheide, A. J. Rech, *Cell Syst.* **9**, 375–382.e4 (2019).
33. N. Xie *et al*., *Signal Transduct. Target. Ther.* **8**, 9 (2023).

ACKNOWLEDGMENTS

We would like to thank G. Nelson for helpful discussions. **Funding:** This project has been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under contract no. 75N91019D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the NIH, Frederick National Laboratory, Center for Cancer Research. See the extended acknowledgments in Supplementary Materials for full details. **Author contributions:** Conceptualization and study design: M.C., M.V., and C.O. Bioinformatic and statistical analyses: M.V., C.O. HLA genotyping: Y.Y. Supervision: M.C. Cohort sources: S.W., S.B., G.D.K., J.J.G., N.L.M., D.W.H., S.G.D., B.W., X.Y. Writing – original draft: C.O., M.V., A.A.B, M.C. Writing – review and editing: all authors **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** All results described in the paper are presented in the main manuscript or the supplementary materials. The functional divergence between every allotypic pair at each locus is provided in data S1 and is available at https://github.com/matviard/FD/. All clinical data underlying the study are available from the cohorts upon application to the cohort curation websites indicated by links in the supplementary materials *Study participants* section. The peptide dataset is publicly available (*10*). The HLA allelic sequences are available from the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk0777
Materials and Methods
Extended Acknowledgements
Figs. S1 to S4
Tables S1 to S4
References (*34*, *35*)
Data S1

Submitted 1 August 2023; accepted 7 December 2023
10.1126/science.adk0777