Received: 19 October 2022 | Revised: 30 October 2023 | Accepted: 4 November 2023

DOI: 10.1111/tan.15298

ORIGINAL ARTICLE

HLA Immune Response Genetics WILEY

Impact of predicted HLA class I immunopeptidome on viral reservoir in a cohort of people living with HIV in Italy

Pier Edoardo Rovatti[1,2] | **Camilla Muccini**[2,3] | **Marco Punta**[1,4] | **Laura Galli**[3] | **Ilaria Mainardi**[2] | **Giacomo Ponta**[2] | **Luca Aldo Edoardo Vago**[1,2,5] | **Antonella Castagna**[2,3]

[1]Unit of Immunogenetics, Leukemia Genomics and Immunobiology, IRCCS San Raffaele Scientific Institute, Milan, Italy

[2]Vita-Salute San Raffaele University, Milan, Italy

[3]Infectious Diseases Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

[4]Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Italy

[5]Hematology and Bone Marrow Transplantation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

Correspondence
Pier Edoardo Rovatti, Unit of Immunogenetics, Leukemia Genomics and Immunobiology, IRCCS San Raffaele Scientific Institute, Milan, Italy.
Email: rovatti.pieredoardo@hsr.it

Camilla Muccini, Vita-Salute San Raffaele University, Milan, Italy.
Email: muccini.camilla@hsr.it

Funding information
Italian Ministry of Health, Grant/Award Numbers: GR-2018-12367860, RF-2011-02348034, RF-2011-02351998; Associazione Italiana per la Ricerca sul Cancro, Grant/Award Numbers: IG #22197, #14162

The class I HLA genotype has been widely recognized as a factor influencing HIV disease progression in treatment-naïve subjects. However, little is known regarding its role in HIV disease course and how it influences the size of the viral reservoir once anti-retroviral therapy (ART) is started. Here, leveraging on cutting-edge bioinformatic tools, we explored the relationship between HLA class I and the HIV reservoir in a cohort of 90 people living with HIV (PLWH) undergoing ART and who achieved viral suppression. Analysis of HLA allele distribution among patients with high and low HIV reservoir allowed us to document a predominant role of *HLA-B* and *-C* genes in regulating the size of HIV reservoir. We then focused on the analysis of HIV antigen (Ag) repertoire, by investigating immunogenetic parameters such as the degree of homozygosity, HLA evolutionary distance and Ag load. In particular, we used two different bioinformatic algorithms, NetMHCpan and MixMHCpred, to predict HLA presentation of immunogenic HIV-derived peptides and identified *HLA-B*57:01* and *HLA-B*58:01* among the highest ranking HLAs in terms of total load, suggesting that their previously reported protective role against HIV disease progression might be linked to a more effective viral recognition and presentation to Cytotoxic T lymphocytes (CTLs). Further, we speculated that some peptide-HLA complexes, including those produced by the interaction between *HLA-B*27* and the HIV Gag protein, might be particularly relevant for the efficient regulation of HIV replication and containment of the HIV reservoir. Last, we provide evidence of a possible synergistic effect between the CCR5 Δ32 mutation and Ag load in controlling HIV reservoir.

KEYWORDS

*B*27*, *B*57:01*, *B*58:01*, CCR5, computational biology, HIV, HLA antigens, immunogenetics, in silico prediction, T-cell epitopes

Pier Edoardo Rovatti and Camilla Muccini have contributed equally as first authors.

Luca Aldo Edoardo Vago and Antonella Castagna have contributed equally as senior authors.

1 | INTRODUCTION

Since the start of the pandemic in 1981, Acquired Immunodeficiency Syndrome (AIDS) caused by the Human Immunodeficiency Virus (HIV) has claimed approximately 40.1 million lives worldwide. Thanks to the development of anti-retroviral therapies (ARTs), the mortality and morbidity rates associated to HIV-related illnesses has dramatically slowed down. However, as of 2021, there are still nearly 38.4 million people living with HIV (PLWH),[1] and the road to HIV eradication continues to look steep. Identifying methods to eradicate in ART-treated subjects the persistent viral reservoir, which contains integrated, replication-competent provirus within host cellular DNA, is currently one of the major hurdles in the battle against HIV infection. When ART is interrupted, viruses from this reservoir cause rebound viraemia and a status of immune dysregulation, which results in subsequent clinical progression.[2]

While the development of a sterilizing, drug-based cure has been traditionally identified and targeted as the main avenue toward eradication of the HIV reservoir in PLWH, in recent years we have witnessed increasing interest in finding a functional cure, aimed at controlling HIV replication without ART.[3] Models of this functional cure are those individuals defined as "elite controllers," who maintain virological suppression without any treatment despite the persistence of replication-competent viruses, remaining clinically and/or immunologically stable for years without ART.[4] A number of previous studies have linked HIV control to the presence of specific HLA alleles[5–7] in patients, suggesting that some HLA alleles may favor immune control of viral replication.

HLA molecules are key players in the immune response to foreign pathogens and they are responsible, in particular, for the presentation of viral antigens (Ags) to T cells. HIV viral control has been associated with HLA class I molecules and, specifically, with a handful of *HLA-B* alleles.[8] Class I molecules are expressed on the cell surface of most nucleated cells, where they present intracellularly produced Ags to CD8+ cytotoxic T lymphocytes (CTLs), the major effector cells of the adaptive immune system.[9] *HLA-B*57:01*,[10,11,5] *HLA-B*27*[12–14] and HLA-*B*58:01*[15,16] have been shown to be associated to a slower progression to AIDS and low levels of pro-viral HIV DNA; additionally, they are strongly enriched in cohorts of HIV controllers compared to non-controllers, whereas several other alleles and allele groups, including *HLA-B*35-Px*,[16] HLA-A*29,[17] *HLA-B*22*,[17] and *HLA-B*58:02*,[18] have been linked to reduced time to progression to AIDS. Several authors reported also a protective effect of high *HLA-C* expression levels on HIV-1 progression,[19–21] hinting at a possible influence of this gene on HIV disease course.

A number of functional mechanisms underlying this relationship between CTL responses, HLA class I polymorphism, and HIV progression have been unveiled, including T cells polyfunctional capabilities, increased clonal turnover, and superior functional avidity in responses to HLA-B-restricted Gag protein-derived epitopes.[22,23]

Further, a number of studies has focused on Ag load as a correlate to HIV-1 clinical outcome,[24–27] with the idea that the capability of an individual to present a larger array of peptides of viral origin might result in a broader and stronger immune response.

Building upon these insights, it is important to further investigate the role of HLA class I molecules in influencing the HIV reservoir and to explore the potential of the presented peptidome as a therapeutic target for HIV treatment.

Leveraging state-of-the-art bioinformatic algorithms, this study aims to evaluate the relationship between HLA class I molecules and the HIV reservoir in PLWH by focusing on patients' HLA genotype diversity and on the HLA-presented viral peptide repertoire.

2 | METHODS

This is a post-hoc analysis of the Intensively Monitored Antiretroviral Pause in Chronic HIV-Infected Subjects with Long-Lasting Suppressed Viremia (APACHE) study,[28] conducted on adult PLWH followed up at IRCCS San Raffaele Scientific Institute, Milan, Italy. The APACHE study is registered with ClinicalTrials.gov (NCT03198325). Inclusion criteria for this particular subgroup analysis were achievement of long-term virological suppression, as defined by HIV-1 RNA <50 copies/mL for at least 10 years, the availability of HIV-1 DNA quantification in PBMCs at sampling time and *HLA-A, -B* and *-C* low-resolution typing. The San Raffaele Ethics Committee approved the study protocol (on 17 May 2016; approval reference number: 31/2016) and the patients signed written informed consent.

We selected a total of 114 adults with chronic HIV-1 infection, HIV-1 RNA <50 copies/mL for ≥10 years, absence of plasma residual viremia for ≥5 years without any viral blips and CD4+ >500 cells/μL, who had been screened for HIV-1 DNA and typed at one-field resolution for the *HLA-A*, *-B* and *-C* loci.

Total HIV-1 DNA was amplified as formerly described[29] and quantified in peripheral blood mononuclear cells (PBMCs) by Real Time PCR (ABI Prism 7900). To distinguish between patients with low and high HIV reservoir, we used the threshold of 100 copies/10^6 PBMCs HIV-DNA.

Low-resolution typing of *HLA-A, -B* and *-C* was performed using genomic DNA and sequence-specific oligonucleotide (HISTO SPOT SSO).

Based on availability, HLA and CCR5 genotyping were performed on $N = 60$ samples by the DKMS Life Science Lab applying a high-resolution amplicon-based approach leveraging Illumina instruments.[30] To help overcome incomplete HLA information on remaining patients, two-field level genotypes were imputed using HaploStats (http://www.haplostats.org), a web application provided by the National Marrow Donor Program (NMDP) Bioinformatics group.

Haplostats uses an expectation–maximization algorithm to leverage a large pool of haplotype samples (for the population of European descent, which represents the majority of patients in our dataset, 395,676 samples with full *HLA-A*, *-B* and *-C* typing[31]). The method takes as input the population of origin (e.g., "African American" or "Filipino") and the one-field genotype from more than one locus of the patient/individual. In our case, one-field genotype input included *HLA-A*, *-B* and *-C* alleles. We used as reference the "NMDP full 2011" dataset.

Median likelihood for 2-field imputation was 74.4% (range 10.8–99.7) (Supplementary Figure 2A). To assess HaploStats accuracy, we compared predicted HLA genotypes and high-resolution HLA typing results in patients with both sets of data available ($N = 56$). $N = 30$ patients (54%) had 6/6 matching alleles, $N = 13$ (23%) had 5/6 matching alleles, $N = 11$ (20%) had 4/6, $N = 1$ (2%) patient had 3/6, $N = 1$ (2%) had 2/6 and no patients had less than 2/6 matching alleles between HaploStats predictions and high-resolution typing results (Supplementary Figure 2B). We used a linear regression model to assess the relationship between the number of matches and the HaploStats likelihood variable, and the "likelihood" coefficient (estimate = 0.0196, $p < 0.001$) indicated a statistically significant positive relationship. Thus, to ensure the integrity of the imputed HLA alleles and to minimize the inclusion of less likely predictions, a stringent threshold was set, and only genotypes that possessed a predicted likelihood equal to or exceeding 90% were retained.

Thus, for all analyses carried out on 2-field HLA genotypes, a total of $N = 73$ patients from our study cohort were used, for $N = 60$ of which we had high-resolution HLA typing data.

Odds ratio (OR) analysis was performed by fitting a generalized linear model to our data (using the function *or_plot()* from R package finalfit[32]), when considering as independent variables all HLAs present in at least 10 individuals of our cohort.

Evolutionary divergence of *HLA-A, -B*, and *-C* diplotypes was estimated using the Grantham Distance[33,34] (GD). We define the HLA class I GD for an individual as the average of their *HLA-A, HLA-B* and *HLA-C* GDs.

For Ag load calculation, we downloaded the HIV-1 M group subtype B reference proteome (NCBI accession number NC_001802.1). In-house generated scripts were used to create from all annotated HIV proteins a list comprising all possible peptides of length 8 to 11, which is the typical length of peptides binding into HLA class I molecules' binding groove. The R package antigengarnish[35] was then used to filter out peptides characterized by a low dissimilarity to the human genome or low similarity to epitopes known for their immunogenic properties (the latter, also known as foreignness), as we hypothesized that these characteristics would increase the likelihood of a given peptide to be under tolerance and therefore to be less immunogenic.[36] Based on the distribution of both parameters across all HIV-derived peptides, we selected peptides with dissimilarity >0.45 or foreignness >0.3 (Supplementary Figure 1).

Peptide HLA-presentation predictions for each HLA two-field allele were performed using NetMHCpan version 4.1[37] and MixMHCpred version 2.1.[38] Strong Binders (SB) were defined by a % Elution Rank ≤0.5 for both predictors. Antigen load (Ag load) was calculated for each patient as the total number of SBs, obtained by adding up the number of SBs for each of the patient's HLA class I alleles.

Statistical analyses were carried out in R, version 4.2.1, except for values reported in Table 2 (difference of the means, Hedge's g effect sizes and associated 95% confidence intervals), which were determined using the calculator available at https://www.cem.org/effect-size-calculator, after verifying that Ag loads for the two HIV-DNA groups calculated with NetMHCpan and MixMHCpred followed normal distributions (Shapiro–Wilk test) and had equal variance (*F*-test). The finalfit[32] R package was used for data table generation (Table 1) and Odds Ratio (OR) analysis; corrplot[39] for visualization of correlation matrices and ggplot[40] for data visualization.

3 | RESULTS

Out of the total 114 PLWH evaluated, $N = 36$ (32%) featured HIV-1 DNA <100 copies/10^6 peripheral blood mononuclear cells (PBMCs), while $N = 78$ (68%) had HIV-1 DNA ≥100 copies/10^6 PBMCs. Overall, $N = 80$ (70%) were male, $N = 34$ (30%) were female. $N = 110$ (96%) were European Caucasian, $N = 3$ (3%) were South or Central American Hispanic, and $N = 1$ (1%) was Caribbean Hispanic, according to HaploStats nomenclature.[31] Median age at HIV diagnosis was 32 (IQR 25–39)

TABLE 1 Patient characteristics. Demographic and clinical characteristics of patients included in the study, separated by HIV-DNA group.

Variable	*N*	Overall, *N* = 114[a]	HIV-DNA group		*p*-value[b]
			<100 copies/10^6 PBMCs, *N* = 36[a]	>100 copies/10^6 PBMCs, *N* = 78[a]	
Gender	114				0.037
F		34 (30%)	6 (17%)	28 (36%)	
M		80 (70%)	30 (83%)	50 (64%)	
Origin	114				0.48
CARHIS		1 (0.9%)	0 (0%)	1 (1.3%)	
EURCAU		110 (96%)	34 (94%)	76 (97%)	
SCAHIS		3 (2.6%)	2 (5.6%)	1 (1.3%)	
Age at diagnosis	112	32 (25, 39)	34 (29, 40)	31 (25, 38)	0.27
Exposure to HIV (years)	112	20.2 (16.7, 23.4)	19.4 (16.6, 22.4)	20.9 (16.8, 24.9)	0.27
ART duration (years)	114	18.3 (16.0, 19.8)	18.2 (15.2, 19.1)	18.4 (16.4, 20.1)	0.37
Time from HIV diagnosis to ART initiation (months)	112	8 (2, 45)	6 (2, 29)	12 (2, 55)	0.48
Time with HIV-1 RNA < 50 copies/mL (years)	114	11.94 (10.63, 14.23)	13.72 (10.98, 15.69)	11.47 (10.60, 13.09)	0.036
CD4 count (cells/μL)	113	790 (644, 938)	744 (596, 924)	800 (661, 946)	0.26

Note: Counts are shown as *N* (%) for categorical variables and Median (IQR) for continuous variables.
Abbreviations: CARHIS, Caribbean Hispanic; EURCAU, European Caucasian; SCAHIS, Hispanic–South or Central American.
[a]Median (IQR) or Frequency (%).
[b]Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test.

years, median exposure to HIV infection was 20 (17–23) years and to ART 18 (16–20) years. Median time from HIV diagnosis to ART initiation was 8 (2–45) months. Median time spent with HIV-1 RNA <50 copies/mL was 12 (11–14) years; median CD4 count at HIV-1 DNA determination was 790 (644–938) cells/μL. These statistics are summarized in Table 1. Generally, HIV-DNA group correlated poorly with the demographic and clinical variables we collected. Partial exceptions were gender (p-value = 0.04) and time spent with HIV-1 RNA <50 copies/mL (p-value = 0.04).

We first investigated whether the immune response to HIV infection could be influenced by a specific HLA locus or allele. No significant global imbalance of *HLA-A, B* or *-C* alleles (Chi-squared test, p-value = 0.37) could be detected between HIV-1 DNA groups in our cohort (Figure 1A). However, after assessing frequencies of HLA Class I alleles for each HIV-1 DNA group, Odds Ratio (OR) analysis of most frequent ($N \geq 10$) individual alleles in our cohort identified *HLA-A*01* (uncorrected p-value = 0.016), *HLA-A*32* (0.018), *HLA-B*07* (0.005), *HLA-B*18* (0.02), and *HLA-B*44* (0.002) as associated to disease progression, mostly in agreement with existing literature,[6,19] and *HLA-C*02* (0.004), *HLA-C*04* (0.036), *HLA-C*05* (0.002), *HLA-C*07* (0.007) and *HLA-C*12* (0.012) as protective toward a lower viral burden, in accordance with known evidence of the importance of *HLA-C* in HIV control[19] (Figure 1B).

When focusing on specific HLA Class I alleles reported to be protective in HIV disease progression, in our cohort *HLA-B*27* was found in 5/31 (16.1%) patients with HIV-1 DNA <100 copies/10^6 PBMCs and in 3/42 (7.1%) with HIV-1 DNA ≥100 copies/10^6 PBMCs (p-value = 0.2721, Fisher's exact test), *HLA B*57:01* in 0 (0%) and in 3/42 (7.1%, p-value = 0.2568), while *HLA B*58:01* in 2/31 (6.5%) and in 2/42 (4.8%, p-value>0.99, Figure 1C).

An efficient adaptive response to infection by intracellular pathogens, such as HIV, largely depends on the ability of an individual's immune system to recognize pathogen-derived peptides and trigger a specific response. Thus, diversity and breadth of the repertoire of HLA-presented peptides is key to an efficient viral clearance. Individuals that are heterozygous for a given HLA locus are more likely to present a broader array of peptides than their homozygous counterparts.[41] Similarly, a higher sequence divergence and a consequent increased structural heterogeneity of an individual's HLA allele repertoire can, in principle, further boost the ability to present a wider range of antigens to immune effector cells.[42]

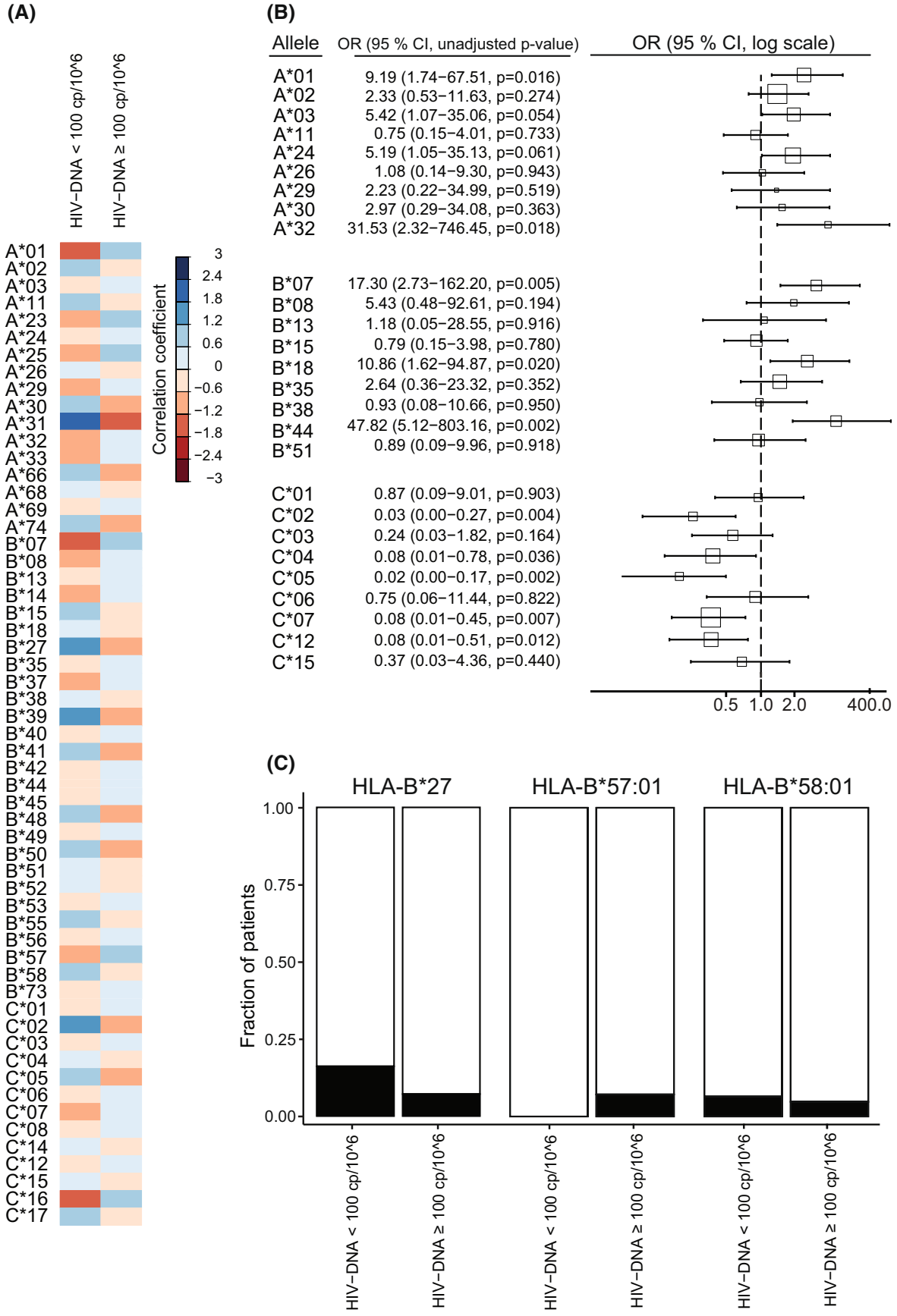

FIGURE 1 Legend on next page.

Indeed, *HLA-B* and *HLA-C* heterozygosity, as well as, *HLA-B* allele divergence have been recently demonstrated to be negatively associated with pre-treatment HIV-1 viral load, thanks to a broader array of HLA-bound peptides.[26]

To investigate these aspects in our cohort, we compared the fraction of patients that were homozygous for the HLA loci and calculated the distribution of mean HLA class I Grantham Distance (GD) in our two HIV-DNA groups.

To carry out both these analyses at a two-field resolution, we combined available high resolution HLA-typing information and two-field HLA allele imputation using the webtool HaploStats,[31] for a total of $N = 73$ patients in our cohort (See Methods section).

In our cohort, $N = 9$ (12%), $N = 0$ (0%) and $N = 2$ (3%) patients were homozygous for one, two and three loci, respectively while $N = 62$ (85%) had no homozygous locus at a one-field level (Figure 2A). No statistical difference was present among the two HIV-DNA groups (Fisher's test, p-value>0.99).

The median GD was 6.742 (IQR 5.82–7.844, range 0–9.354) for HIV-DNA <100 copies/10^6 PBMCs group, and 7.137 (IQR 5.962–7.801, range 0–10.341) for HIV-1 DNA ≥100 copies/10^6 PBMCs group, with no statistically significant difference between the two (p-value = 0.8104 by Mann–Whitney test, Figure 2B).

Finally, we asked whether the effective control achieved over HIV viral replication could be related to the number of highly immunogenic peptides presented to host T cells.

From a reference HIV genome, we derived a pool of all possible peptides of length ranging from 8 to 11 amino acids ($N = 12{,}268$ unique peptides). We then selected peptides with high dissimilarity to the human self-peptidome or high similarity to known immunogenic epitopes (foreignness), in order to derive a list of peptides with a higher probability of eliciting an immunological response.[43] This resulted in $N = 4356$ total peptides. We then used two state-of-the-art bioinformatic tools, NetMHCpan4.1[37] and MixMHCpred2.1,[38] to predict in silico binding likelihoods between these peptides and HLA alleles present in our cohort, along with the most frequent HLA alleles in the Italian population[44] (frequency ≥ 1%). In total, we obtained $N = 2225$ strong binders according to NetMHCpan, and $N = 3077$ for MixMHCpred (out of $N = 357{,}192$ possible peptide-HLA pairs). Interestingly, we documented a higher amount of predicted strong binding peptides for protective alleles *HLA-B*57:01* and *HLA-B*58:01* compared to other HLA class I alleles with both predictors (Figure 3A,B), although this was not the case for *HLA-B*27:05* and *HLA-B*27:02*, found to be protective in other studies.[6] The association between *HLA-B*27* and the immune control of HIV-1 has been linked, in particular, to the targeting of *Gag* specific epitopes. To verify this relationship in our study, we calculated Ag load of *HLA-B* alleles only considering *Gag*-derived peptides (Figure 3C). Interestingly, neither *HLA-B*27:05* or *HLA-B*27:02* were characterized by a higher Ag load. Nonetheless, supporting previous evidence that the association between *HLA-B*27* and the immune control of HIV-1 is linked to the targeting of *Gag* specific epitopes by CTLs,[22] *Gag* peptide KRWIILGLNK (KK10)[45] was predicted to be a strong binder for *HLA-B*27:05* both by NetMHCpan (%EL rank = 0.149) and MixMHCpred (%EL rank = 0.2), while HLA-B*27:02 showed strong binding properties to KK10 according to MixMHCpred (%EL rank = 0.4) and weak binding properties according to NetMHCpan (%EL rank = 0.824).

Finally, Ag load values in patients stratified by HIV-DNA group were compared. Neither NetMHCpan4.1 (Kruskal–Wallis test, p-value = 0.89) (Figure 3D), nor MixMHCpred2.1 (p-value = 0.86) (Figure 3E) predicted Ag load as significantly different between the two HIV-DNA groups. This was further confirmed by the absence of a significant linear correlation between HIV-DNA levels and Ag load ($R = 0.053$ for NetMHCpan and $R = 0.051$ for MixMHCpred) (Figure 3F,G).

A key factor in determining HIV infection susceptibility and disease progression is CCR5, a chemokine receptor that acts as a coreceptor for HIV entry into immune cells.[46] Individuals homozygous for a 32-bp deletion (Δ32/Δ32) in *CCR5* are resistant to HIV,[47] and individuals heterozygous for *CCR5* Δ32 have a slower progression to AIDS.[48] To assess whether differences observed in HIV reservoir depth in our cohort could be influenced by *CCR5* mutational status, HIV-1 tropism and *CCR5* genotype were assessed. In our cohort, $N = 77/114$ (68%) individuals had a CCR5-tropic infection, and $N = 8/57$ (14%)

FIGURE 1 (A) Correlogram showing the association between HLAs and HIV-DNA groups in our dataset. Color intensity indicates the absolute value of the correlation coefficients, with shades of blue expressing positive values and shades of red negative ones (anti-correlation). (B) Forest plot displaying the Odds ratio (OR) between HIV-DNA≥100 copies/10^6 PBMCs and HIV-DNA < 100 copies/10^6 PBMCs for individual HLA alleles. Only HLAs with N ≥ 10 were considered in this analysis. (C) Histograms showing the fraction of patients for each HIV-DNA group with at least one *B*27*, *B*57:01* or *B*58:01* allele, respectively.

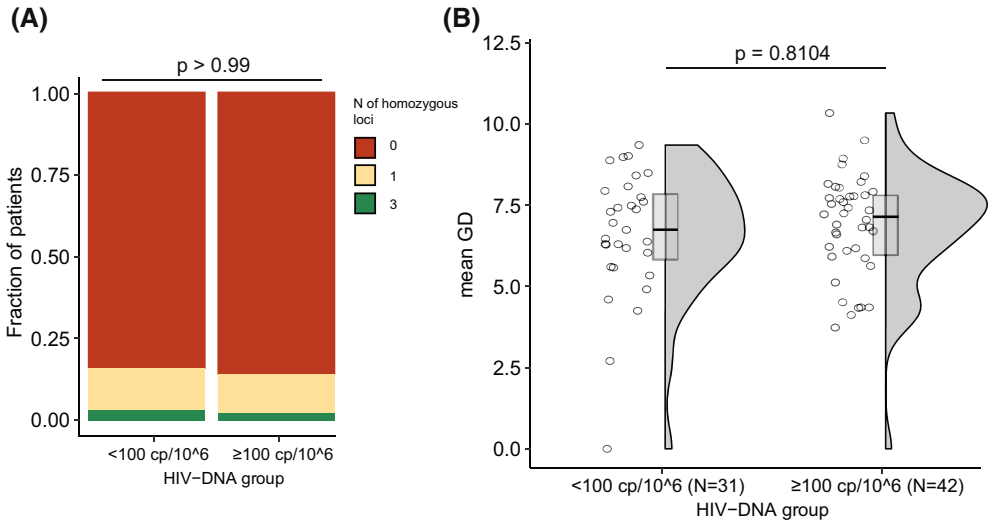

FIGURE 2 (A) Bar plot showing the proportion of patients with 1, 2, 3 or no homozygous loci per HIV-DNA group. (B) Raincloud[54] plot displaying GD values for patients, separated by HIV-DNA group. Dots indicate individual raw values; boxplots show median, interquartile range (IQR) and highest and lowest values (whiskers) excluding outliers; half-violins provide an explicit representation of the distributions. Lower "mean GD" values point to a less diverse set of HLA alleles for a patient compared to patients with higher "mean GD" values.

were heterozygous for *CCR5* Δ32, while no patient was homozygous for the mutation.

Δ32 heterozygosity was found in 4/27 (15%) patients with HIV-1 DNA <100 copies/10^6PBMCs and in 3/23 (12%) with HIV-1 DNA ≥100 copies/10^6PBMCs (p-value > 0.99, Chi-square test, Figure 4A). Next, we compared Ag load obtained from NetMHCpan and MixMHCpred of both *CCR5* wild-type (WT) and Δ32 heterozygous individual among HIV DNA groups.

When using NetMHCpan, no difference in the predicted Ag load could be identified among HIV-DNA groups, neither in *CCR5* Δ32 nor WT patients (Figure 4B). While similar results were obtained comparing predicted Ag load with MixMHCpred on *CCR5* WT patients (Figure 4C, left panel), heterozygous *CCR5* Δ32 patients in the HIV-1 DNA <100 copies/10^6 PBMCs group exhibited a slight significantly higher (Kruskal–Wallis test, p-value = 0.05) Ag load compared to those in the HIV-1 DNA ≥100 copies/10^6PBMCs group.

4 | DISCUSSION

By focusing on the quantitative and qualitative properties of the interaction between HLAs and the HIV-derived peptide repertoire and utilizing a bioinformatic pipeline relying on state-of-the-art packages and algorithms, we studied the potential link between HLA class I molecules and the HIV reservoir in PLWH. We started from a cohort of $N = 114$ patients with a long history of infection and ART exposure and that have been controlling the infection (HIV-RNA <50 copies/mL) for ≥10 years. We used HIV-DNA as a marker for quantification of their HIV reservoir. Total HIV-DNA level in PBMCs is an important predictive factor for the risk of progression to AIDS and death[49,50] in PLWH, and it has been traditionally used as a proxy for the total HIV reservoir. We divided our cohort into two groups, using the threshold of 100 copies/10^6 PBMCs to identify patients with low and high reservoir, and explored differences in HLA and Ag repertoire between these two subgroups. Ag repertoire was assessed using proxies such as degree of homozygosity and Grantham Distance and more directly by predicting Ag load. We were able to confirm some previously published findings, such as the relevance of *HLA-C* genes and of the *HLA-B*07* allele in HIV viral control. We also provided a novel putative explanation for the importance of certain *HLA-B* alleles, in particular *B*57:01* and *B*58:01*, in HIV disease progression. Indeed, HLA alleles that are capable of binding and effectively presenting a wider range of peptides of viral origin may confer a better immunological control and slow down the course of HIV infection. Moreover, we hypothesized that specific peptide–MHC complexes, such as those obtained from the interaction between Gag-derived peptides and *HLA-B*27*, might be key to successful control of HIV replication and containment of the HIV reservoir.

Lastly, we observed that patients in our cohort who were heterozygous for the *CCR5* Δ32 mutation and belonged to the group with a smaller HIV reservoir had higher Ag loads. Conversely, in patients with the *CCR5* wild-type genotype, there was no significant difference in

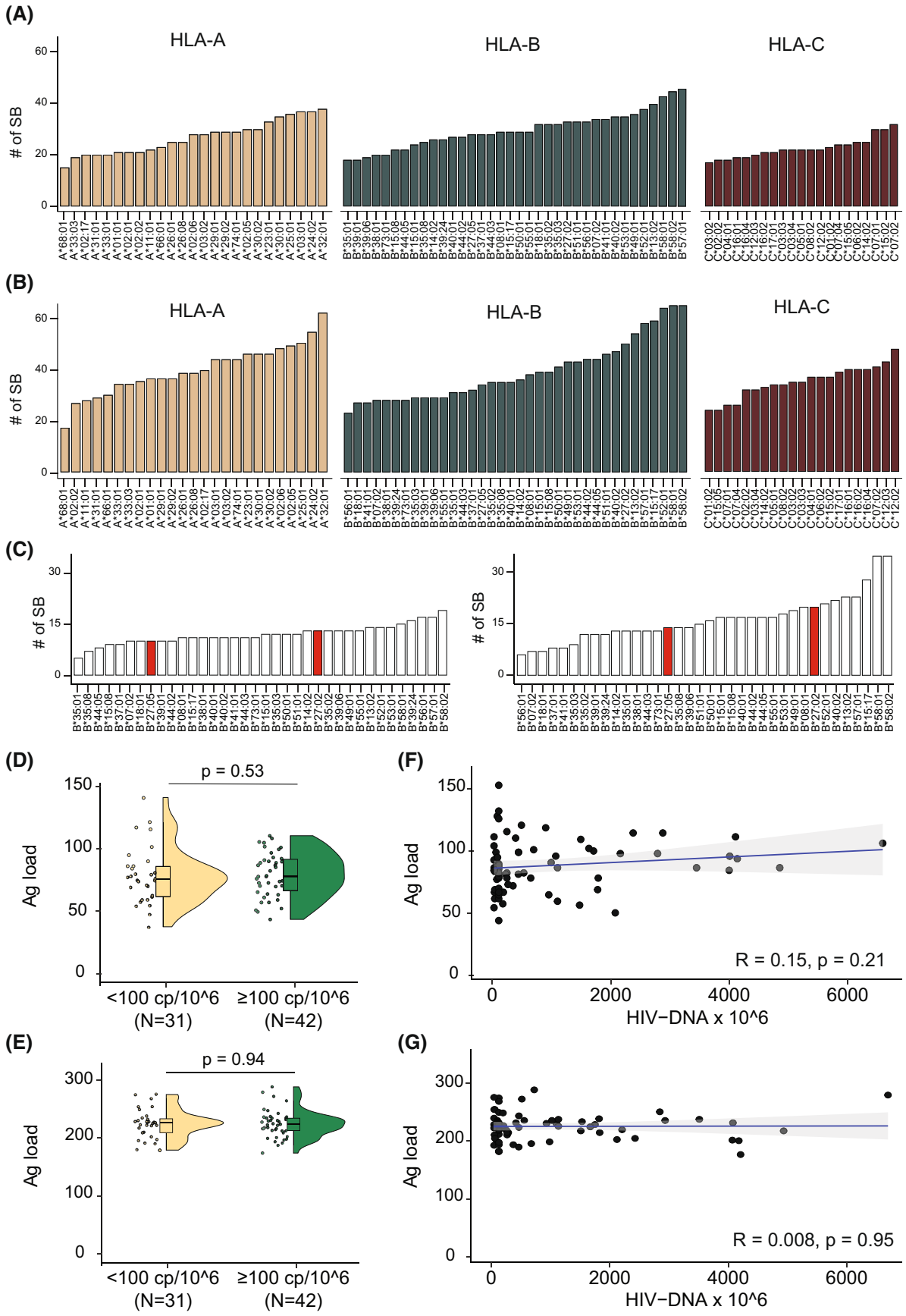

FIGURE 3 Legend on next page.

Ag load. Although this could be observed only using MixMHCpred to predict Ag load, and though the number of heterozygous *CCR5* Δ32 patients in our cohort is very limited (N = 6), we can speculate that in a setting where HIV entry in host's cell is more difficult, a stronger HLA-mediated viral recognition may help control the disease.

Our study presents a number of limitations. First, HLA typing was achieved experimentally at a low-resolution level for about a half of our cohort, and for these patients two-field resolution was imputed using a computational method, HaploStats. To increase typing reliability, we decided to remove from our dataset patients whose 2-field imputation likelihood (from HaploStats) was less than 90%. Second, despite recent, significant advances, accurate Ag load calculation remains a difficult task.[51] Here, we tried to obviate these difficulties in two ways: we used two separate algorithms to calculate load (NetMHCpan and MixMHXpred) and, additionally, applied a pre-filtering step for peptide immunogenicity,[35] which should narrow down the number of peptides to those that are most likely to induce an immune response. Still, as new data about presentability

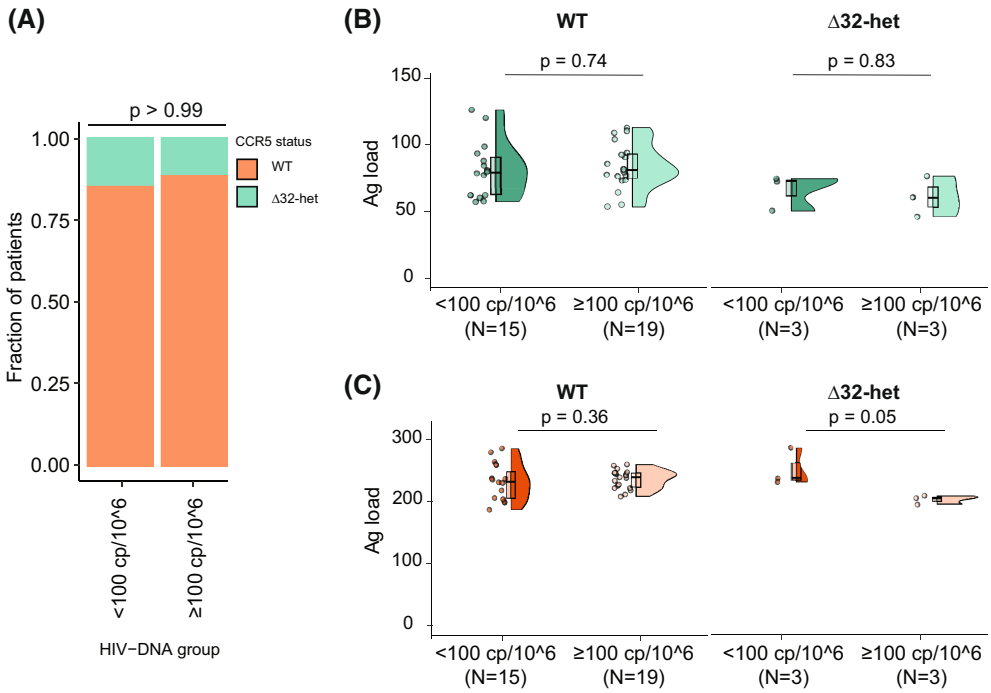

FIGURE 4 (A) Bar plot showing the proportion of patients with CCR5 WT and heterozygous CCR5 Δ per HIV-DNA group. (B, C) Raincloud plots depicting predicted Ag load by NetMHCpan (B) and MixMHCpred (C) for each patient in the two HIV-DNA groups, split by CCR5 status. On the left, Ag load for WT CCR5 ($N = 34$), on the right heterozygous CCR5 Δ32 ($N = 6$).

TABLE 2 Unstandardized effect size (mean difference) and standardized, unbiased effect size (Hedge's g) with 95% Confidence Intervals (CI) (see Methods) for the comparison between the two different HIV-DNA groups for Ag load calculated with NetMHCpan or MicMHCpred, respectively.

Study	Mean Diff	Lower 95% CI	Upper 95% CI	Hedge's g	Lower 95% CI	Upper 95% CI
Ag load (NetMHCpan)	1.06	−8.55	10.67	0.05	−0.41	0.52
Ag load (MixMHCpred)	−0.3	−11.54	10.94	−0.01	−0.48	0.45

FIGURE 3 (A, B) Histograms displaying the total number of strong binding peptides (see Methods) for HLA alleles with a frequency ≥1% in the Italian population, or present in our cohort, split by HLA class I gene and calculated by NetMHCpan (A) and MixMHCpred (B). HLA alleles for each gene are ordered according to increasing number of strong binders. (C) Number of strong binders predicted for *HLA-B* alleles by NetMHCpan (left) and MixMHCpred (right) using peptides derived solely from the HIV Gag protein. Highlighted in red are *HLA-B*27:05* and *HLA-B*27:02* (see main text). (D, E) Raincloud plots showing Ag load predicted by NetMHCpan (D) and MixMHCpred (E) for each patient in our two HIV-DNA groups. (F, G) Dot plot of Ag load (*y*-axis) predicted by NetMHCpan (F) and MixMHCpred (G) and HIV-DNA levels (*x*-axis) in our cohort. Blue line of best fit is obtained by linear regression method and confidence intervals are shown in gray. Correlation coefficients (R) and *p*-values are shown at the bottom right of each panel.

and immunogenicity of peptides are being produced and new methods are being developed, precision in immunopeptidome prediction is bound to improve. The major limitation of our work, however, is the small sample size. Since this was a post-hoc study, no preliminary power calculations were performed. Post-hoc power calculations are known to be directly related to p-values[52] and, as such, to be of little additional value with respect to the latter. Thus, to provide additional information useful for interpretation of the Ag load comparisons, we have calculated 95% confidence intervals for differences between means and for Hedge's g effect sizes (standardized differences between means)[53] (Table 2, Supplementary Figure 3). From these calculations we can see that our Ag load comparisons between different HIV-1 DNA groups feature rather large confidence intervals, implying that our results should be further validated on a larger cohort of patients.

In conclusion, HLA class I genotype has been consistently shown to be one of the strongest genetic factors influencing HIV disease progression. Its roles range from the interaction with cells of the innate immune system, that is, by modulation of natural killer (NK) cell activity, to interaction with CD4+ and CD8+ T cells. Here, we provide an immune-informatic approach to shed new light on the mechanisms underlying HIV immune control, which could advance the quest for an effective functional cure of HIV or for the design of novel vaccines.

AUTHOR CONTRIBUTIONS

Pier Edoardo Rovatti: Conceptualization; data analysis and interpretation; writing–original draft; writing–review and editing. **Camilla Muccini**: Conceptualization; data analysis and interpretation; writing–original draft. **Marco Punta**: Conceptualization; supervision; writing–review and editing. **Laura Galli, Ilaria Mainardi, and Giacomo Ponta**: Writing–review and editing. **Luca Aldo Edoardo Vago and Antonella Castagna**: Conceptualization; supervision; funding and resources acquisition; writing–original draft; writing–review and editing.

ACKNOWLEDGMENTS

The authors thank all members of the Vago laboratory and the Unit of Infectious Diseases at the IRCCS San Raffaele Scientific Institute, Milano, for critical discussion on the experiments and the manuscript. This work was supported by the Italian Ministry of Health (RF-2011-02351998 to F. Ciceri and L. Vago, RF-2011-02348034 to L. Vago, GR-2018-12367860 to L. Vago, and TRANSCAN HLALOSS to L. Vago), by the Associazione Italiana per la Ricerca sul Cancro (Start-Up Grant #14162 and IG #22197 to F. Ciceri). Open access funding provided by BIBLIOSAN.

CONFLICT OF INTEREST STATEMENT

L. Vago reports grants from the Italian Ministry of Health and Associazione Italiana per la Ricerca sul Cancro during the conduct of the study, as well as personal fees from GEN-DX and grants and personal fees from Moderna Therapeutics outside the submitted work. A. Castagna has received consultancy payments and speaking fee from Bristol-Myers Squibb, Gilead, ViiV Health Care, Merck Sharp & Dohme, Janssen-Cilag. No disclosures were reported by the other authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Pier Edoardo Rovatti https://orcid.org/0000-0001-5399-5222

REFERENCES

1. UNAIDS. 2022 epidemiological estimates. https://www.unaids.org
2. Churchill MJ, Deeks SG, Margolis DM, Siliciano RF, Swanstrom R. HIV reservoirs: what, where and how to target them. *Nat Rev Microbiol*. 2016;14(1):55-60. doi:10.1038/nrmicro.2015.5
3. Woldemeskel BA, Kwaa AK, Blankson JN. Viral reservoirs in elite controllers of HIV-1 infection: implications for HIV cure strategies. *EBioMedicine*. 2020;62:103118. doi:10.1016/j.ebiom.2020.103118
4. Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity*. 2007;27(3):406-416. doi:10.1016/j.immuni.2007.08.010
5. Migueles SA, Sabbaghian MS, Shupert WL, et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A*. 2000;97(6):2709-2714. doi:10.1073/pnas.050567397
6. The International HIV Controllers Study. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. 2010;330(6010):1551-1557. doi:10.1126/science.1195271
7. Migueles SA, Connors M. Success and failure of the cellular immune response against HIV-1. *Nat Immunol*. 2015;16(6):563-570. doi:10.1038/ni.3161
8. Kiepiela P, Leslie AJ, Honeyborne I, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature*. 2004;432(7018):769-775. doi:10.1038/nature03113
9. Kawashima Y, Pfafferott K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature*. 2009;458(7238):641-645. doi:10.1038/nature07746
10. Bailey JR, Williams TM, Siliciano RF, Blankson JN. Maintenance of viral suppression in HIV-1–infected HLA-B*57+ elite suppressors despite CTL escape mutations. *J Exp Med*. 2006;203(5):1357-1369. doi:10.1084/jem.20052319

11. Lambotte O, Boufassa F, Madec Y, et al. HIV controllers: A homogeneous group of HIV-1: Infected patients with spontaneous control of viral replication. *Clin Infect Dis*. 2005;41(7):1053-1056. doi:10.1086/433188
12. McNeil AJ, Yap PL, Gore SM, et al. Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease. *QJM*. 1996;89(3):177-186. doi:10.1093/qjmed/89.3.177
13. Costello C, Tang J, Rivers C, et al. HLA-B*5703 independently associated with slower HIV-1 disease progression in Rwandan women. *Aids*. 1999;13:1990-1991. doi:10.1097/00002030-199910010-00031
14. Kløverpris HN, Leslie A, Goulder P. Role of HLA adaptation in HIV evolution. *Front Immunol*. 2016;6:6. doi:10.3389/fimmu.2015.00665
15. Emu B, Sinclair E, Hatano H, et al. HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. *J Virol*. 2008;82(11):5398-5407. doi:10.1128/JVI.02176-07
16. Jin X, Gao X, Ramanathan M, et al. Human immunodeficiency virus type 1 (HIV-1)-specific CD8 $^{+}$ −T-cell responses for groups of HIV-1-infected individuals with different HLA-B*35 genotypes. *J Virol*. 2002;76(24):12603-12610. doi:10.1128/JVI.76.24.12603-12610.2002
17. Hendel H, Caillat-Zucman S, Lebuanec H, et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol*. 1999;162(11):6942-6946.
18. Ngumbela KC, Day CL, Mncube Z, et al. Targeting of a CD8 T cell env epitope presented by HLA-B*5802 is associated with markers of HIV disease progression and lack of selection pressure. *AIDS Res Hum Retroviruses*. 2008;24(1):72-82. doi:10.1089/aid.2007.0124
19. Apps R, Qi Y, Carlson JM, et al. Influence of HLA-C expression level on HIV control. *Science*. 2013;340(6128):87-91. doi:10.1126/science.1232685
20. Kulpa DA, Collins KL. The emerging role of HLA-C in HIV-1 infection: HLA-C in HIV-1 infection. *Immunology*. 2011;134(2):116-122. doi:10.1111/j.1365-2567.2011.03474.x
21. Kulkarni S, Savan R, Qi Y, et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature*. 2011;472(7344):495-498. doi:10.1038/nature09914
22. Almeida JR, Price DA, Papagno L, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *J Exp Med*. 2007;204(10):2473-2485. doi:10.1084/jem.20070784
23. Berger CT, Frahm N, Price DA, et al. High-functional-avidity cytotoxic T lymphocyte responses to HLA-B-restricted gag-derived epitopes associated with relative HIV control. *J Virol*. 2011;85(18):9334-9345. doi:10.1128/JVI.00460-11
24. Bashirova AA, Viard M, Naranbhai V, et al. HLA tapasin independence: broader peptide repertoire and HIV control. *Proc Natl Acad Sci U S A*. 2020;117(45):28232-28238. doi:10.1073/pnas.2013554117
25. Khairkhah N, Namvar A, Kardani K, Bolhassani A. Prediction of cross-clade HIV-1 T-cell epitopes using immunoinformatics analysis. *Proteins*. 2018;86(12):1284-1293. doi:10.1002/prot.25609
26. Arora J, Pierini F, McLaren PJ, Carrington M, Fellay J, Lenz TL. HLA heterozygote advantage against HIV-1 is driven by quantitative and qualitative differences in HLA allele-specific peptide presentation. *Mol Biol Evol*. 2020;37(3):639-650. doi:10.1093/molbev/msz249
27. Barquera R, Collen E, Di D, et al. Binding affinities of 438 HLA proteins to complete proteomes of seven pandemic viruses and distributions of strongest and weakest HLA peptide binders in populations worldwide. *HLA*. 2020;96(3):277-298. doi:10.1111/tan.13956
28. Castagna A, Muccini C, Galli L, et al. Analytical treatment interruption in chronic HIV-1 infection: time and magnitude of viral rebound in adults with 10 years of undetectable viral load and low HIV-DNA (APACHE study). *J Antimicrobial Chemother*. 2019;74(7):2039-2046. doi:10.1093/jac/dkz138
29. De Rossi A, Zanchetta M, Vitone F, et al. Quantitative HIV-1 proviral DNA detection: a multicentre analysis. *New Microbiol*. 2010;33(4):293-302.
30. Lange V, Böhme I, Hofmann J, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. 2014;15(1):63. doi:10.1186/1471-2164-15-63
31. https://bioinformatics.bethematchclinical.org/workarea/downloadasset.aspx?id=6339
32. Harrison E, Drake T, Ots R. finalfit: Quickly create elegant regression results tables and plots when modelling. https://github.com/ewenharrison/finalfit
33. Bentham R. HLAdivR (Version v1.0). 2020. P. R Package. 2020 https://rdrr.io/github/rbentham/HLAdivR/
34. Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974;185(4154):862-864. doi:10.1126/science.185.4154.862
35. Richman LP, Vonderheide RH, Rech AJ. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst*. 2019;9(4):375-382.e4. doi:10.1016/j.cels.2019.08.009
36. Wells DK, Van Buuren MM, Dang KK, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell*. 2020;183(3):818-834.e13. doi:10.1016/j.cell.2020.09.015
37. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48(W1):W449-W454. doi:10.1093/nar/gkaa379
38. Gfeller D, Guillaume P, Michaux J, et al. The length distribution and multiple specificity of naturally presented HLA-I ligands. *JI*. 2018;201(12):3705-3716. doi:10.4049/jimmunol.1800914
39. Wei T, Simko V. R package "corrplot": Visualization of a Correlation Matrix. https://github.com/taiyun/corrplot
40. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag; 2016. https://ggplot2.tidyverse.org
41. Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*. 1975;256(5512):50-52. doi:10.1038/256050a0
42. Chowell D, Krishna C, Pierini F, et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat Med*. 2019;25(11):1715-1720. doi:10.1038/s41591-019-0639-4
43. Schmidt J, Smith AR, Magnin M, et al. Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Rep Med*. 2021;2(2):100194. doi:10.1016/j.xcrm.2021.100194
44. Sacchi N, Castagnetta M, Miotti V, Garbarino L, Gallina A. High-resolution analysis of the HLA-A, -B, -C and -DRB1

alleles and national and regional haplotype frequencies based on 120 926 volunteers from the Italian Bone Marrow Donor Registry. *HLA*. 2019;94(3):285-295. doi:10.1111/tan.13613

45. Payne RP, Kløverpris H, Sacha JB, et al. Efficacious early antiviral activity of HIV gag- and pol-specific HLA-B*2705-restricted CD8 $^{+}$ T cells. *J Virol*. 2010;84(20):10543-10557. doi:10.1128/JVI.00793-10
46. Dragic T, Litwin V, Allaway GP, et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*. 1996;381(6584):667-673. doi:10.1038/381667a0
47. Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*. 1996;382(6593):722-725. doi:10.1038/382722a0
48. Cohen OJ, Vaccarezza M, Lam GK, et al. Heterozygosity for a defective gene for CC chemokine receptor 5 is not the sole determinant for the immunologic and virologic phenotype of HIV-infected long-term nonprogressors. *J Clin Invest*. 1997;100(6):1581-1589. doi:10.1172/JCI119682
49. Rouzioux C, Hubert J, Burgard M, et al. Early levels of HIV-1 DNA in peripheral blood mononuclear cells are predictive of disease progression independently of HIV-1 RNA levels and CD4 $^{+}$ T cell counts. *J Infect Dis*. 2005;192(1):46-55. doi:10.1086/430610
50. Goujard C, Bonarek M, Meyer L, et al. CD4 cell count and HIV DNA level are independent predictors of disease progression after primary HIV type 1 infection in untreated patients. *Clin Infect Dis*. 2006;42(5):709-715. doi:10.1086/500213
51. Sellars MC, Wu CJ, Fritsch EF. Cancer vaccines: building a bridge over troubled waters. *Cell*. 2022;185(15):2770-2788. doi:10.1016/j.cell.2022.06.035
52. Althouse AD. Post hoc power: not empowering, just misleading. *J Surg Res*. 2021;259:A3-A6. doi:10.1016/j.jss.2019.10.049
53. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol*. 2016;69(6):555-562. doi:10.4097/kjae.2016.69.6.555
54. Allen M, Poggiali D, Whitaker K, Marshall TR, van Langen J, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res*. 2021;4:63. doi:10.12688/wellcomeopenres.15191.2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rovatti PE, Muccini C, Punta M, et al. Impact of predicted HLA class I immunopeptidome on viral reservoir in a cohort of people living with HIV in Italy. *HLA*. 2024;103(1):e15298. doi:10.1111/tan.15298