Immunoinformatic Analysis Reveals Antigenic Heterogeneity of Epstein-Barr Virus Is Immune-Driven

Ana Cirac [1,2†], Remy Poirey [3†], Michael Dieckmeyer [4], Klaus Witter [5], Henri-Jacques Delecluse [3], Uta Behrends [1,2,6‡] and Josef Mautner [1,2,6*‡]

[1] Children's Hospital, School of Medicine, Technische Universität München, Munich, Germany, [2] German Centre for Infection Research (DZIF), partner site Munich, Munich, Germany, [3] German Cancer Research Center (DKFZ) Unit F100 and Institut National de la Santé et de la Recherche Médicale Unit U1074, Heidelberg, Germany, [4] Department of Diagnostic and Interventional Neuroradiology, Technische Universität München, Munich, Germany, [5] Laboratory of Immunogenetics, Ludwig-Maximilians-Universität, München, Germany, [6] Institute of Virology, Helmholtz Zentrum München, Munich, Germany

Whole genome sequencing of Epstein-Barr virus (EBV) isolates from around the world has uncovered pervasive strain heterogeneity, but the forces driving strain diversification and the impact on immune recognition remained largely unknown. Using a data mining approach, we analyzed more than 300 T-cell epitopes in 168 published EBV strains. Polymorphisms were detected in approximately 65% of all CD8+ and 80% of all CD4+ T-cell epitopes and these numbers further increased when epitope flanking regions were included. Polymorphisms in CD8+ T-cell epitopes often involved MHC anchor residues and resulted in changes of the amino acid subgroup, suggesting that only a limited number of conserved T-cell epitopes may represent generic target antigens against different viral strains. Although considered the prototypic EBV strain, the rather low degree of overlap with most other viral strains implied that B95.8 may not represent the ideal reference strain for T-cell epitopes. Instead, a combinatorial library of consensus epitopes may provide better targets for diagnostic and therapeutic purposes when the infecting strain is unknown. Polymorphisms were significantly enriched in epitope versus non-epitope protein sequences, implicating immune selection in driving strain diversification. Remarkably, CD4+ T-cell epitopes in EBNA2, EBNA-LP, and the EBNA3 family appeared to be under negative selection pressure, hinting towards a beneficial role of immune responses against these latency type III antigens in virus biology. These findings validate this immunoinformatics approach for providing novel insight into immune targets and the intricate relationship of host defense and virus evolution that may also pertain to other pathogens.

Keywords: Epstein-Barr virus, virus evolution, immunoinformatics, T-cell epitope, strain variants

INTRODUCTION

The Epstein-Barr virus (EBV) establishes lifelong persistent infections in more than 90% of the human population by latently infecting B cells. Primary infection with EBV usually occurs early in life by parent-to-child oral transmission in an almost always asymptomatic fashion. Delayed primary infection in adolescence or adulthood may cause the syndrome of infectious mononucleosis (IM). Besides IM, EBV has been associated with a growing number of non-malignant and malignant

diseases, such as chronic active EBV infection, hemophagocytic lymphohistiocytosis, post-transplant lymphoproliferative disease, NK/T-cell, Burkitt, and Hodgkin lymphoma, as well as nasopharyngeal and gastric carcinoma (1, 2). In addition, EBV infection has been linked to the etiology of several autoimmune diseases, e.g. multiple sclerosis (1, 3).

Based on the EBV nuclear antigen (EBNA) 2 and EBNA-3 gene sequences, EBV strains have been grouped into type 1 and type 2 (4–6). More recently, sequence analyses of a rapidly growing number of viral isolates from different healthy virus carriers and tumor specimens have identified numerous viral variants and complex variations in EBV strains within the 1/type 2 classification system (5, 7–10). Moreover, a comparative analysis of whole genome sequences identified a number of polymorphisms that were shared by tumor-derived strains but absent in viral isolates from healthy study participants (7). These findings suggest that some sequence variations may impact on viral phenotype and pathogenicity, a notion recently substantiated by *in vitro* studies (11–14).

In most virus carriers, EBV is contained as asymptomatic infection mainly by T cells recognizing various viral latent and lytic cycle antigens (15, 16). Consequently, patients with T-cell dysfunction are at increased risk of developing life-threatening EBV-associated lymphoproliferative disease and reconstitution of EBV-specific immunity in these patients by the adoptive transfer of virus-specific T-cells has been shown to be effective in preventing and treating such EBV-associated pathologies (17–21). However, response rates in apparently immunocompetent patients with EBV-associated malignancies are still unsatisfactory (19).

Owing to high viral titers that can be readily obtained from the supernatant of the producer cell line, the B95.8 EBV strain has become the prototypic type 1 strain and is commonly used for generating EBV-specific T-cell lines for clinical application (22). Whether immune responses elicited against B95.8 also protect against a rapidly growing number of heterogeneous field strains is still unknown. Amino acid sequence exchanges have been identified in several T-cell epitopes. Also, differences in T-cell recognition of target cells infected with different viral strains (23), or loaded with peptide epitope variants, have been described. In fact, the first EBV-derived CD8+ T-cell epitope that was molecularly defined was shown to differ between B95.8 and the prototypic type 2 strain AG876 in two amino acids (24). Among the CD8+ T-cell epitopes identified in subsequent studies, some proved to be conserved in different viral strains (25), while others were found to vary between virus types (26), or even within the same virus type (27–32). Depending on the epitope and the polymorphism, T-cell recognition was unaffected, diminished, or abolished (27–30, 32, 33).

While these investigations had focused on a limited number of CD8+ T-cell epitopes and diverse viral isolates, a recent study compared the majority of all published CD8+ and CD4+ EBV T-cell epitopes in the three viral strains B95.8, AG876, and M81 (34). Polymorphisms were detected in about half of all T-cell epitopes and the individual T-cell responses ignored some epitope variants, which may impair immunity to strain variants (34).

These findings raised several questions including (i) the overall CD4+ and CD8+ T-cell epitope variation rate in a broader spectrum of viral isolates, (ii) whether conserved epitopes exist that represent generic targets for immunotherapeutic intervention against different viral strains across the world, and (iii) whether random genetic drift or immune selection is driving strain diversification.

Here, we compared more than 300 T-cell epitopes in 168 published EBV strains isolated from a variety of geographical regions and histological sample types. This analysis revealed amino acid exchanges in a large proportion of all CD4+ and CD8+ T-cell epitopes, identified clinically relevant epitopes with no/low inter-strain variability, and provided evidence for immune pressure shaping the evolution of the virus. These results validate our immunoinformatics approach that can readily be adapted to other complex viruses such as CMV and SARS-CoV-2.

MATERIAL AND METHODS

Epitope Analysis

All viral strain sequences were downloaded from NCBI. Epitope sequences including their flanking regions were aligned using in-house routines built with MATLAB 9.4 (2018a) (The MathWorks Inc., Natick, MA, USA). If epitopes differed from the B95.8 reference strain, including one or multiple amino acid exchanges, deletions, insertions or strong dissimilarities, they were classified as containing "polymorphisms" either within the epitope, the flanking region, or in both. Results were obtained in absolute or relative numbers, the latter being calculated by dividing individual strain numbers sharing the same epitope variant by the number of total strains that provided sequence data for the respective epitope. Data was further analyzed in GraphPad Prism 7 and MATLAB. Statistical analysis (alpha=0.05) was performed with MATLAB using Chi^2 goodness of fit test for uniform distribution as well as two-sided binomial test with Bonferroni correction. Allele frequencies were obtained from http://www.allelefrequencies.net.

Consensus Library

The viral strains sequences alignments using ClustalW algorithm and the resulting consensus library were generated with the MacVector software.

Assessing Polymorphism Probabilities in Epitope vs. Non-Epitope Regions

Calculation of polymorphism probabilities and adjusting weights were all performed with MATLAB 9.4 (R2018a) as follows:
Definitions:

P_i: i-th EBV protein

S_i^p: protein size, as total number of protein P_i

S_i^e: total number of amino acids of the epitope regions in protein P_i

$S_i^{ex} = S_i^p - S_i^e$

μ_i^e: number of polymorphisms within the epitope regions (S_i^e) of protein P_i

μ_i^{ex}: number of polymorphisms within the regions excluding epitopes (S_i^{ex}) of protein P_i

P_i^e: probability of polymorphisms occurring within epitope regions of protein P_i

P_i^{ex}: probability of polymorphisms occurring within the regions excluding epitopes (S_i^{ex}) of protein P_i

P^e: overall probability of polymorphisms occurring within epitope regions of all proteins

P^{ex}: overall probability of polymorphisms occurring within the regions excluding epitopes (S_i^{ex}) of all proteins

WP_i^e: weighted probability of polymorphisms occurring within epitope regions of protein P_i

WP_i^{ex}: weighted probability of polymorphisms occurring within the regions excluding epitopes (S_i^{ex}) of protein P_i

WP^e: overall weighted probability of polymorphisms occurring within epitope regions of all proteins

WP^{ex}: overall weighted probability of polymorphisms occurring within the regions excluding epitopes (S_i^{ex}) of all proteins

Calculating Probabilities of Polymorphisms Within Different Protein Regions

The probability of polymorphisms occurring in the epitope regions of one protein P_i is the number of polymorphisms within the epitope region divided by the total number of amino acids of the epitope regions. The overall probability of polymorphisms occurring within the epitope regions is the sum of all polymorphisms found in all epitope regions divided by the total number of amino acids of all the epitope regions. The same argument applies to the non-epitope regions.

$$P_i^e = \frac{\mu_i^e}{s_i^e}; \quad P^e = \frac{\Sigma_i\, \mu_i^e}{\Sigma_i\, s_i^e}$$

$$P_i^{ex} = \frac{\mu_i^{ex}}{s_i^{ex}}; \quad P^{ex} = \frac{\Sigma_i\, \mu_i^{ex}}{\Sigma_i\, s_i^{ex}}$$

Adjusting the Weights of the Proteins´ Individual Probabilities for the Overall Analysis According to Epitope and Protein Size

The weighted probability of polymorphisms occurring within epitope regions of protein P_i is the probability of polymorphisms occurring within epitope regions of protein P_i weighted with the ratio of the number of amino acids of the epitope region in protein P_i and the total number of amino acids of the epitope regions of all proteins. This coincides with number of polymorphisms within the epitope regions of protein P_i divided by the total number of amino acids of the epitope regions of all proteins. The same argument applies to the non-epitope regions.

$$WP_i^e = P_i^e\, \frac{s_i^e}{\Sigma_n\, s_n^e} = \frac{\mu_i^e}{s_i^e}\, \frac{s_i^e}{\Sigma_n\, s_n^e} = \frac{\mu_i^e}{\Sigma_n\, s_n^e}$$

$$WP_i^{ex} = P_i^{ex}\, \frac{s_i^{ex}}{\Sigma_n\, s_n^{ex}} = \frac{\mu_i^{ex}}{s_i^{ex}}\, \frac{s_i^{ex}}{\Sigma_n\, s_n^{ex}} = \frac{\mu_i^{ex}}{\Sigma_n\, s_n^{ex}}$$

With these definitions, the overall probability of polymorphisms occurring within epitope regions of all proteins coincides with the sum of the weighted probabilities with respect to all proteins. The same argument applies to the non-epitope regions. This means that the weighted probabilities decompose the total probability of polymorphisms within an epitope or non-epitope region according to the protein sizes.

$$\Sigma_i\, WP_i^e = \frac{\Sigma_i\, \mu_i^e}{\Sigma_i\, s_i^e} = P^e; \quad \Sigma_i\, WP_i^{ex} = \frac{\Sigma_i\, \mu_i^{ex}}{\Sigma_i\, s_i^{ex}} = P^{ex}$$

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov Test was performed including each data point of the different EBV strains for all 32 EBV proteins using weight-adjusted probabilities.

For each protein variant, WP_i^e and WP_i^{ex} were calculated and the two sets of data were analyzed for statistical significance using the Kolmogorov-Smirnov test using MATLAB 9.4 (R2018a).

RESULTS

Geographical and Tissue Origin of EBV Strains

In this analysis of T-cell epitope polymorphisms, genomic sequences of 168 distinct EBV strains from various geographical regions and tissues were included (**Table S1** and **Figure S1A**). Most of the strains had been isolated from donors in Asia (n=58) and Africa (n=42), fewer had been obtained from individuals in North America (n = 28), Australia (n=21), Europe (n=9), and South America (n=9). In accordance with their worldwide prevalence (35), the vast majority of all strains from the different geographical regions were classified as type 1, except for African virus isolates of which 25% were type 2. Furthermore, genomic sequences from almost one third of all Asian strains were recombined or incomplete and classified neither as type 1, nor as type 2. One type 2 strain sample was of unknown origin (**Figure S1A**).

Most strains had been isolated from malignant tissues of epithelial (e.g. nasopharyngeal, gastric and lung carcinomas) or lymphoid origin (e.g. Burkitt lymphoma, PTLD, Hodgkin lymphoma). About one third of the viral strains had been isolated from non-malignant tissues from donors from different parts of the world, including PBMC of patients with infectious mononucleosis (IM) and healthy virus carriers, as well as spontaneous LCL (sLCL) and saliva (**Figure S1B**).

Frequency of Polymorphisms in CD4+ and CD8+ T-Cell Epitopes

This collection of viral strains from diseased and healthy tissues from around the world was used to assess polymorphisms in EBV-specific T-cell epitopes. To this end, we compared amino

acid sequences of 185 CD8+ T-cell epitopes and 120 CD4+ T-cell epitopes that had been retrieved from literature and also included some unpublished CD4+ T-cell epitopes identified by our group (34, 36). Since amino acid sequences flanking an epitope can impact its antigenicity, flanking regions (FR) were defined as five N- and five C-terminal amino acids immediately adjacent to the epitope and were also included in this analysis. Most of the CD4+ (72%) and almost half (46%) of the CD8+ T-cell epitopes were derived from latent cycle antigens. The distribution of T-cell epitopes across EBV antigens is displayed in **Table S2**.

A list of all CD4+ and CD8+ T-cell epitopes and their flanking regions as well as their variants in the analyzed EBV strains is shown in **Table S3**. Because some viral genome sequences are still incomplete, not all epitopes could be analyzed in all viral strains, especially when located close to or within DNA repeats. However, most epitopes were analyzed in more than 150 viral strains.

Of the 185 CD8+ T-cell epitopes, 65 were conserved in all viral strains. When taking the flanking regions also into account, this number dropped to 42. Accordingly, less than 25% of all CD8+ epitopes plus flanking regions were found to be conserved in all viral strains examined. Of the 120 CD4+ T-cell epitopes, 24 were found to be conserved and this number dropped to 15, or 12.5%, when flanking regions were included (**Figure 1A**).

Polymorphic CD4+ T-cell epitopes were found to have a mean number of 7 and a maximum number of 36 variants, while CD8+ epitopes varied less with a mean number of 4 and a maximum number of 25 variants (**Figure 1B**). Detailed polymorphism analysis for every CD4+ and CD8+ T-cell epitope is provided in **Table S3**.

Epitope Variability Within EBV Strains

These results suggested that only a limited number of conserved T-cell epitopes represent generic immune targets against all viral variants. However, several epitopes were conserved or differed in only a small number of viral isolates and thus might still be effective against a broad set of viral strains. Therefore, median sequence variations in CD4+ and CD8+ T-cell epitopes were assessed by calculating the percentage of strains in which each epitope was found to be conserved with respect to the B95.8 sequence (**Figure 2A**). On average, any given CD4+ T-cell epitope was estimated to be conserved in almost 80% of viral strains. If flanking regions were considered, CD4+ T-cell epitopes were conserved in almost 70% of all viral strains. In the case of CD8+ T-cell epitopes, the degree of conservation was even higher. On average, any given CD8+ T-cell epitope was found to be conserved in almost 99% of all viral strains, and this number dropped to 90% when flanking regions were included into the analysis. While these numbers implied that T-cell responses against these epitopes may be effective against infection with almost all viral strains, the high level of scattering suggested that strains with vastly disparate sets of T-cell epitopes may exist. To identify such variants that are possibly refractory to B95.8 antigen sequence-based immunotherapy, viral strains were ranked according to the number of epitope
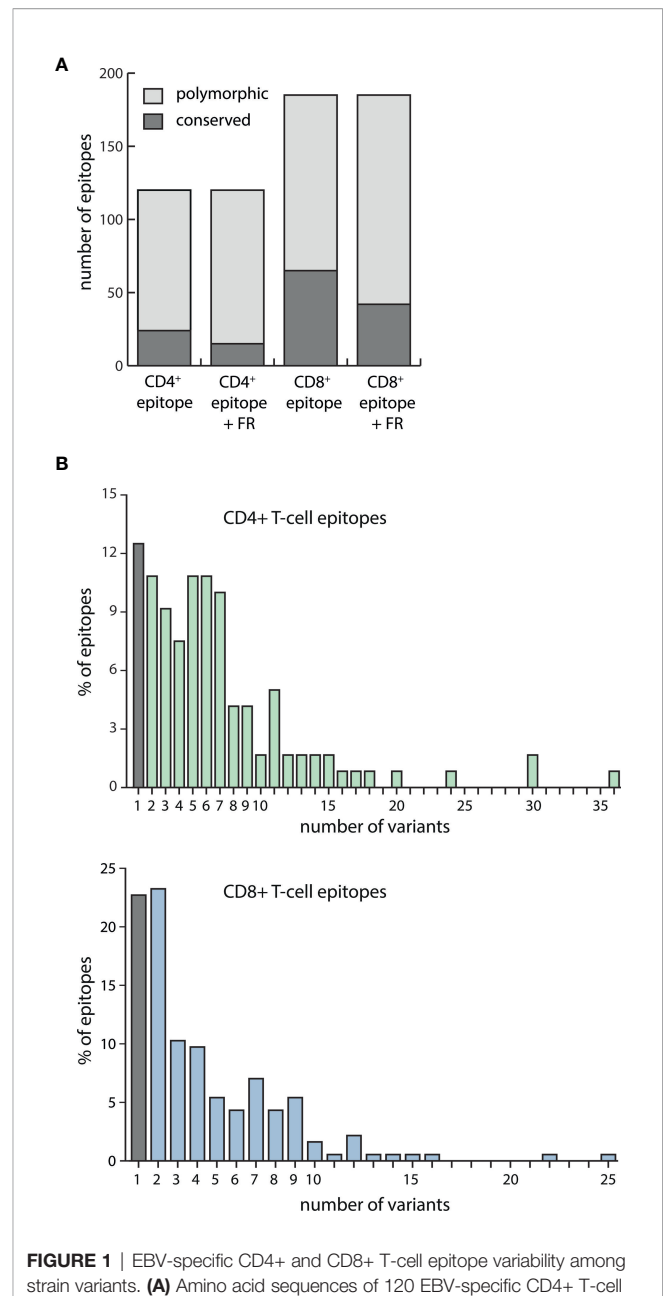

FIGURE 1 | EBV-specific CD4+ and CD8+ T-cell epitope variability among strain variants. **(A)** Amino acid sequences of 120 EBV-specific CD4+ T-cell epitopes and 185 CD8+ T-cell epitopes, either alone or including their flanking regions (FR), were assessed for polymorphisms in 168 EBV strains. Exchanges, insertions, and deletions of single or several amino acids were regarded as polymorphisms. Depicted is the number of epitopes +/- their flanking regions that were found to be conserved or polymorphic among the analyzed viral strains. **(B)** For each of the 120 CD4+ (upper graph) and 185 CD8+ (lower graph) T-cell epitopes including their flanking regions, the number of sequence variants among 168 EBV strains was determined.

variants they contained. This analysis revealed that in any given strain, on average 40% of all CD4+ and 20% of all CD8+ T-cell epitopes including their flanking regions differed from the B95.8 sequence (**Figure 2B**). Strains that differed greatly from B95.8 consensus were mostly of type 2. With more than 70% divergence in CD4+ and more than 50% in CD8+ T-cell
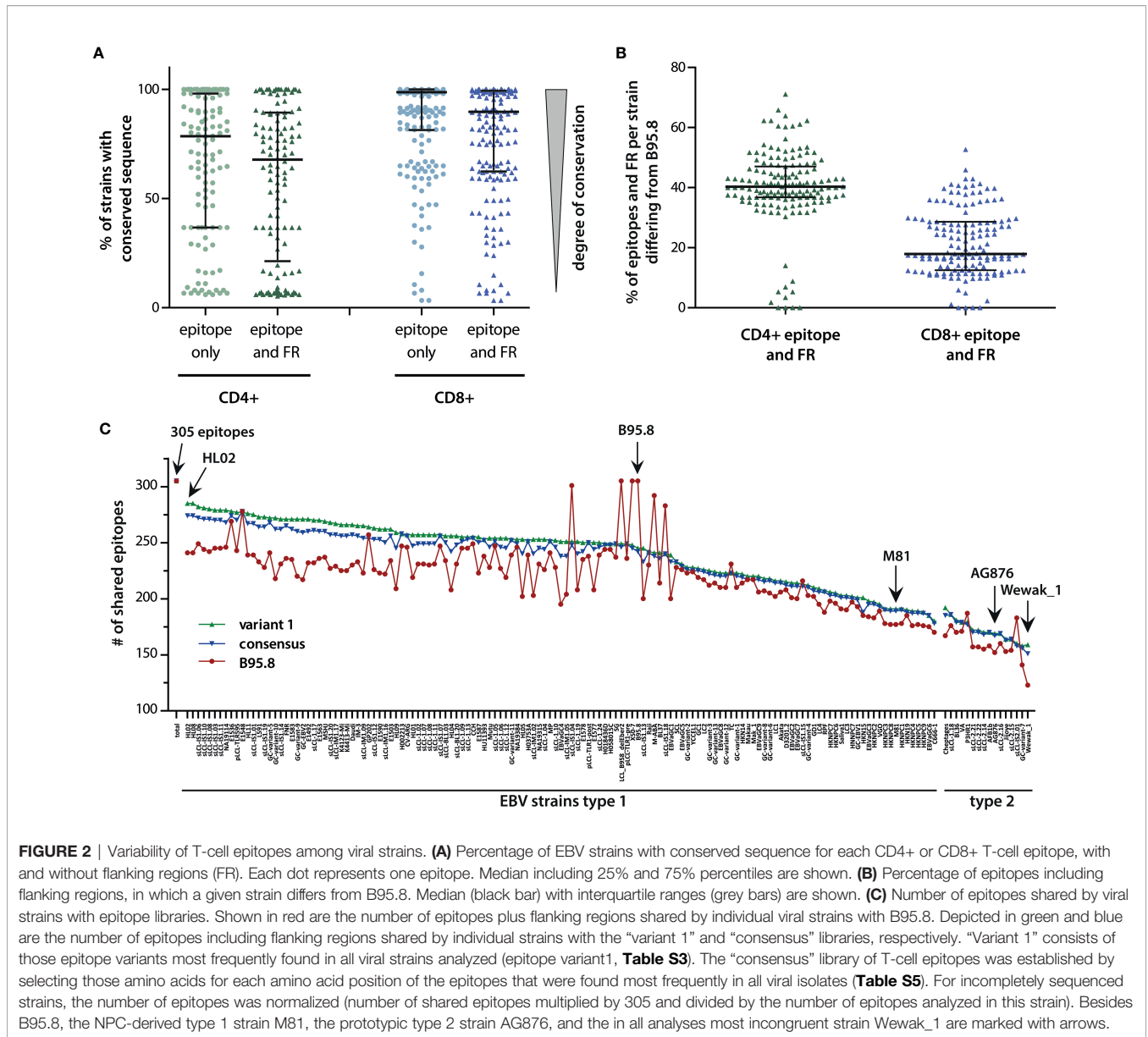
FIGURE 2 | Variability of T-cell epitopes among viral strains. **(A)** Percentage of EBV strains with conserved sequence for each CD4+ or CD8+ T-cell epitope, with and without flanking regions (FR). Each dot represents one epitope. Median including 25% and 75% percentiles are shown. **(B)** Percentage of epitopes including flanking regions, in which a given strain differs from B95.8. Median (black bar) with interquartile ranges (grey bars) are shown. **(C)** Number of epitopes shared by viral strains with epitope libraries. Shown in red are the number of epitopes plus flanking regions shared by individual viral strains with B95.8. Depicted in green and blue are the number of epitopes including flanking regions shared by individual strains with the "variant 1" and "consensus" libraries, respectively. "Variant 1" consists of those epitope variants most frequently found in all viral strains analyzed (epitope variant1, **Table S3**). The "consensus" library of T-cell epitopes was established by selecting those amino acids for each amino acid position of the epitopes that were found most frequently in all viral isolates (**Table S5**). For incompletely sequenced strains, the number of epitopes was normalized (number of shared epitopes multiplied by 305 and divided by the number of epitopes analyzed in this strain). Besides B95.8, the NPC-derived type 1 strain M81, the prototypic type 2 strain AG876, and the in all analyses most incongruent strain Wewak_1 are marked with arrows.

epitopes, the most divers strain was found to be the type 2 strain Wewak. As there were only few viral isolates that shared all or almost all epitopes with B95.8, this analysis indicated that B95.8 may not represent the ideal reference strain for T-cell epitopes. To investigate this in more detail, the number of epitopes plus flanking regions shared by viral strains with B95.8 was determined (**Figure 2C**). Five strains shared almost all 305 epitopes with B95.8 but this number declined to 163 within the group of EBV type 1, and ranged between 191 and 126 in type 2 strains with Wewak_1 showing the least overlap. To explore the possibility of higher and more balanced coverage, two epitope libraries were designed. The "variant 1" library consisted of those epitope and flanking region variants that were most frequently found in all strains analyzed (epitope variant 1; **Table S3**). The "consensus" library was generated by selecting those amino acids

for each position of a given epitope plus flanking regions that were most frequently found at this position in viral isolates (**Table S5**). Next, the number of T-cell epitopes shared by a given viral strain with these designer libraries was determined (**Figure 2C**). Compared to B95.8, an overall higher degree of congruity and a narrower range was noted for both libraries (average number of epitopes shared by viral strains with B95.8: 217, with consensus library: 230, with variant 1 library: 235; range B95.8: 123-305, consensus library: 151-278, variant 1 library: 158-285). Due to the high proportion of type 1 strains in this collection, congruity was highest for type 1 strains, whereas NPC-derived type 1 and type 2 strains clustered at the lower end. Thus, the separation of strains observed in principal component analyses (5) was also reflected at the T-cell epitope level (**Figure 2C**).

Immune-Driven Epitope Diversification

To investigate whether these sequence variations were caused by random genetic drift or by an immune-driven selection process, the location of amino acid exchanges in octamer and nonamer CD8+ T-cell epitopes and their flanking regions was investigated. In the case of the 7 octamer epitopes and their variants (n=24), most polymorphisms were located within the epitopes and barely within the flanking regions. Furthermore, most amino acid exchanges were noted at position 2 of the epitope, which is required for anchoring the peptide to the MHC binding pocket (**Figure 3A**). To assess whether these polymorphisms changed the chemical characteristics of the peptides and thereby impacted on peptide/MHC interactions, amino acids were divided into charged, polar, and hydrophobic subgroups. All polymorphisms at position 1 and 2 of the epitope resulted in a change of the subgroup (**Figure 3B**). While these results were indicative of an immune-driven diversification process, the low number of 8-mer epitopes (n=24) precluded statistical validation.

Therefore, the same analysis was performed with the 78 polymorphic nonamer epitopes and their variants (n=416) listed in **Table S3**. In this case, a statistically significant increase in amino acid exchanges at position nine was noted (**Figures 3C, D**). Most peptides that bind to MHC I have a hydrophobic (or sometimes basic) anchor residue at the carboxy terminus. Although not reaching statistical significance, frequent amino acid exchanges were also noted at position two, which is another important anchor residue for several HLA class I alleles. While amino acid substitutions at position 1 (**Figure 3E**), which is not critical for binding to most HLA class I molecules, were mostly of the same subgroup, amino acids at positions 2 (**Figure 3F**) and 9 (**Figure 3G**) were often replaced by members of different subgroups. Overall, approximately 50% of the amino acids at P9 were exchanged for amino acids of other subgroups. Approximately one third of the hydrophobic amino acids present in the B95.8 epitopes were exchanged for polar or charged amino acids and the majority of charged amino acids at P9 were exchanged for amino acids of the polar subgroup. Such amino acids are not favored by most MHC class I alleles at this position. Almost half of the polymorphisms at position two resulted in a change in the amino acid subgroup. In addition, amino acids at positions 3 and 7, considered secondary anchor residues for several HLA alleles, were exchanged for members of other subgroups in the majority of cases. Likewise, polymorphisms in the first amino acid downstream of the epitope (position 6 FR), which impacts on proteasomal cleavage, often altered the amino acid subgroup (**Figure 3D**). Collectively, these findings indicated that the polymorphisms in epitopes are likely to have consequential effects on the immune recognition.

Polymorphisms Are Enriched in Epitope *Versus* Non-Epitope Sequences

To investigate the role of immune pressure as a driving force behind viral strain diversification, we assessed the locations of T-cell epitopes and polymorphisms in proteins. As exemplified for LMP2A in **Figure 4**, polymorphic regions and epitopes often colocalized, especially when polymorphisms occurred in a large number of strains. A systematic analysis of polymorphism rates in epitope and non-epitope sequences in all 32 viral proteins, in which

T-cell epitopes had been identified, demonstrated that polymorphisms were significantly enriched in both CD4+ and CD8+ T-cell epitopes (p-value <0.005) (**Figure 5A**). The probability that any given amino acid was affected by a polymorphism and belonged to a CD4+ T-cell epitope was found to be almost threefold higher than in regions of the protein that are not known to be targeted by T cells. For CD8+ T-cell epitopes, this ratio was found to be 1,5-fold increased. When the epitopes were subdivided in latency type III antigens (EBNA2, EBNA-LP, and EBNA3 family) and all antigens except latency type III antigens, polymorphisms in CD4+ T-cell epitopes in the latter group were further enriched, but reduced in epitopes in latency type III antigens (**Figure 5B**). Because similar numbers of lytic cycle (n=34) and latency type III-associated (n=35) CD4+ T-cell epitopes have been analyzed (**Table S2**), this discrepancy is unlikely to be caused by differences in sample sizes. Furthermore, a similar trend, albeit less pronounced, became apparent for CD8+ T-cell epitopes. These results identify immune pressure as driver in EBV strain diversification, except for CD4+ T-cell epitopes in latency type III antigens, which appear to be under negative selection.

DISCUSSION

To study the clinical implications of EBV strain heterogeneity, we compared EBV T-cell epitopes in more than 150 viral strains derived from normal and diseased tissues and from different geographical regions. Approximately 30% of all CD8+ and 80% of all CD4+ T-cell epitopes were found to be polymorphic and these numbers further increased when the epitope flanking regions were included. However, one has to keep in mind that many of the published CD4+ T-cell epitopes were defined using overlapping peptide libraries of 15 or 20 amino acids in length (**Table S3**). The core sequence of CD4+ T-cell epitopes, however, usually encompasses only 9 amino acids. Consequently, some of the identified polymorphisms may reside outside the core epitope or its flanking regions and have led to an over-estimation in the number of polymorphisms in CD4+ T-cell epitopes.

In addition, several of the analyzed viral strains were derived from tumor cell lines in which mutations in T-cell epitopes may have accumulated due to genetic instability and/or immune pressure and thereby contributed to these high polymorphism rates. This, however, appears unlikely because rare (tumor-specific) epitopes were not found to be enriched in strains isolated from tumor cells (**Table S3**).

On average, viral strains differed from B95.8 in 40% of all CD4+ and 20% of all CD8+ T-cell epitopes including flanking regions and variability extended over a broad range, challenging the role of B95.8 as reference strain for T-cell epitopes. Polymorphisms in T-cell epitopes have previously been implicated as potential reason for low clinical efficacy of B95.8 LCL-stimulated T-cell preparations in some PTLD patients and their failure to recognize spontaneous LCL from these patients (37–45). Although LCL usually present multiple viral antigens on MHC class I and II, including some that were probably shared between B95.8 and the endogenous virus of the patients, the CTL response is often dominated by few epitopes so that even polyclonal EBV-specific T-cell preparations may be
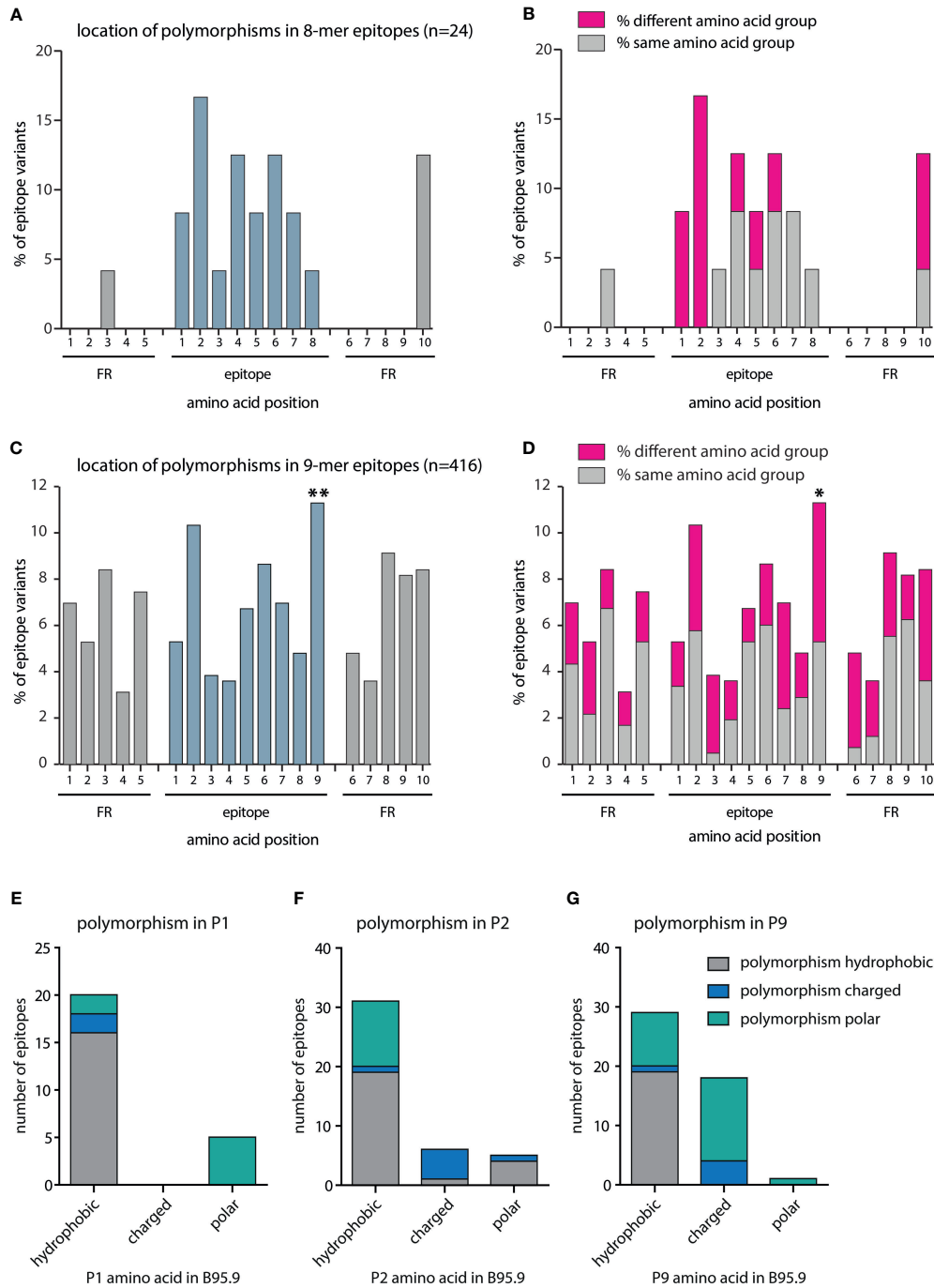
FIGURE 3 | Quantitative and qualitative analysis of amino acid exchanges in CD8+ T-cell epitopes and epitope flanking regions. All polymorphisms in octamer and nonamer CD8+ T-cell epitopes and their flanking regions were included in this analysis. The number of polymorphisms was determined separately for each amino acid position in all variants. Depicted are the frequencies of polymorphisms for each amino acid position in **(A)** octamers plus flanking regions (n=24), and **(C)** nonamers plus flanking regions (n=416). Chi^2-test goodness of fit test for uniform distribution was performed using MATLAB: **(A)** Chi^2-value 25,45 (p = 0.0851), **(C)** Chi^2-value 71,15 (p < 0.00001). For each position, a two-sided binomial test was performed using Bonferroni adjustment. P-values were indicated as follows *$p \leq 0.002$; **$p \leq 0.0002$. **(B, D)** Amino acids were divided into the subgroups charged (D, E, K, R), polar (C, H, N, Q, S, T, W, Y), and hydrophobic (A, F, G, I, L, M, P, V). For each position, the percentage of polymorphisms was determined that resulted in a change of the amino acid subgroup. In nonamer epitopes, the number of hydrophobic, charged, and polar amino acids present at positions one **(E)**, two **(F)**, and nine **(G)** was determined and the number of amino acid exchanges resulting in the same or a different subgroup defined.
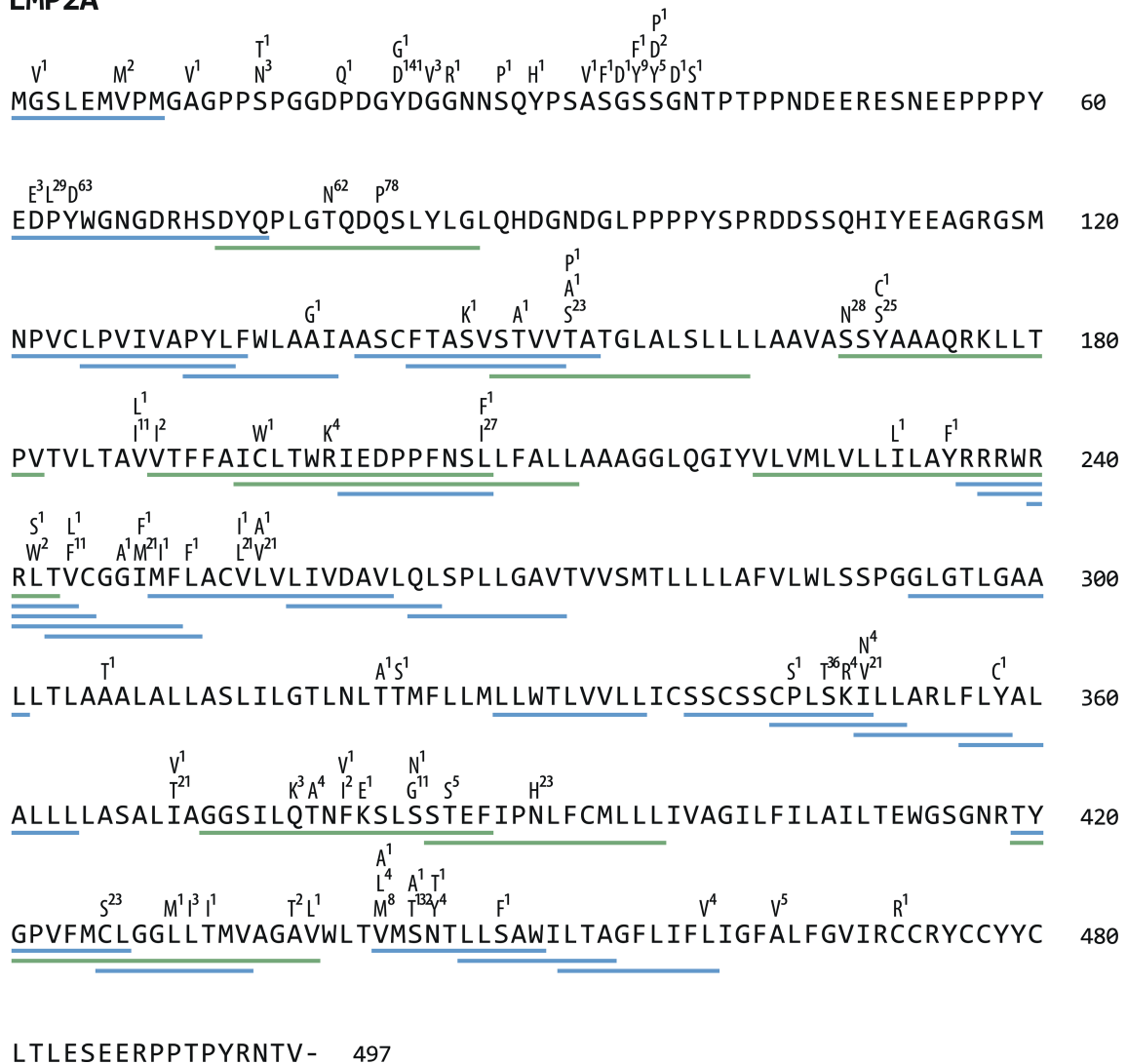
FIGURE 4 | Location of polymorphisms and T-cell epitopes in LMP2A. LMP2A protein sequences of 155 viral strains were aligned and polymorphisms identified using Clustal Omega software (https://www.ebi.ac.uk/Tools/msa/clustalo/). All identified amino acid exchanges and the number of strains in which they occurred are displayed above the LMP2A reference sequence from B95.8. Underneath the sequence, green (CD4+) and blue (CD8+) lines indicate the location of T-cell epitopes (further details can be found in **Table S3**).

oligoclonal in specificity (46–48). Consequently, epitope libraries containing the most frequent epitope variants or designed by selecting those amino acids most frequently found at a given position of an epitope in all EBV strains, may constitute better targets for diagnostic and therapeutic purposes when the infecting strain is unknown. Alternatively, BAC technology may be used to generate viral strains with higher and more balanced epitope coverage such as HL02 (**Figure 2C**) and used for T-cell stimulation instead of B95.8 (49).

Despite the overall high variability, several T-cell epitopes were found to be conserved in all strains analyzed, including some immunodominant EBV antigens (**Table S4**), such as the GLC-epitope derived from BMLF1, or the YVL-epitope derived from

BRLF1, that may represent generic targets for clinical application. The reason for this conservation is not known, but as described for other viruses, these epitopes may be located in regions critical for protein function in which escape comes at the cost of viral fitness (50). By further analogy to other pathogens, the development of such escape mutants would also be unlikely to occur if T-cell responses against these epitopes had beneficial effects on virus pathobiology or negligible effects on viral transmission (51, 52).

Previous analyses demonstrated that amino acid exchanges in epitopes and their flanking regions can retain, reduce, or abolish T-cell recognition [(53) and references therein]. Polymorphisms within or in close proximity to epitopes may affect T-cell recognition in various ways, e.g. by altering antigen processing,
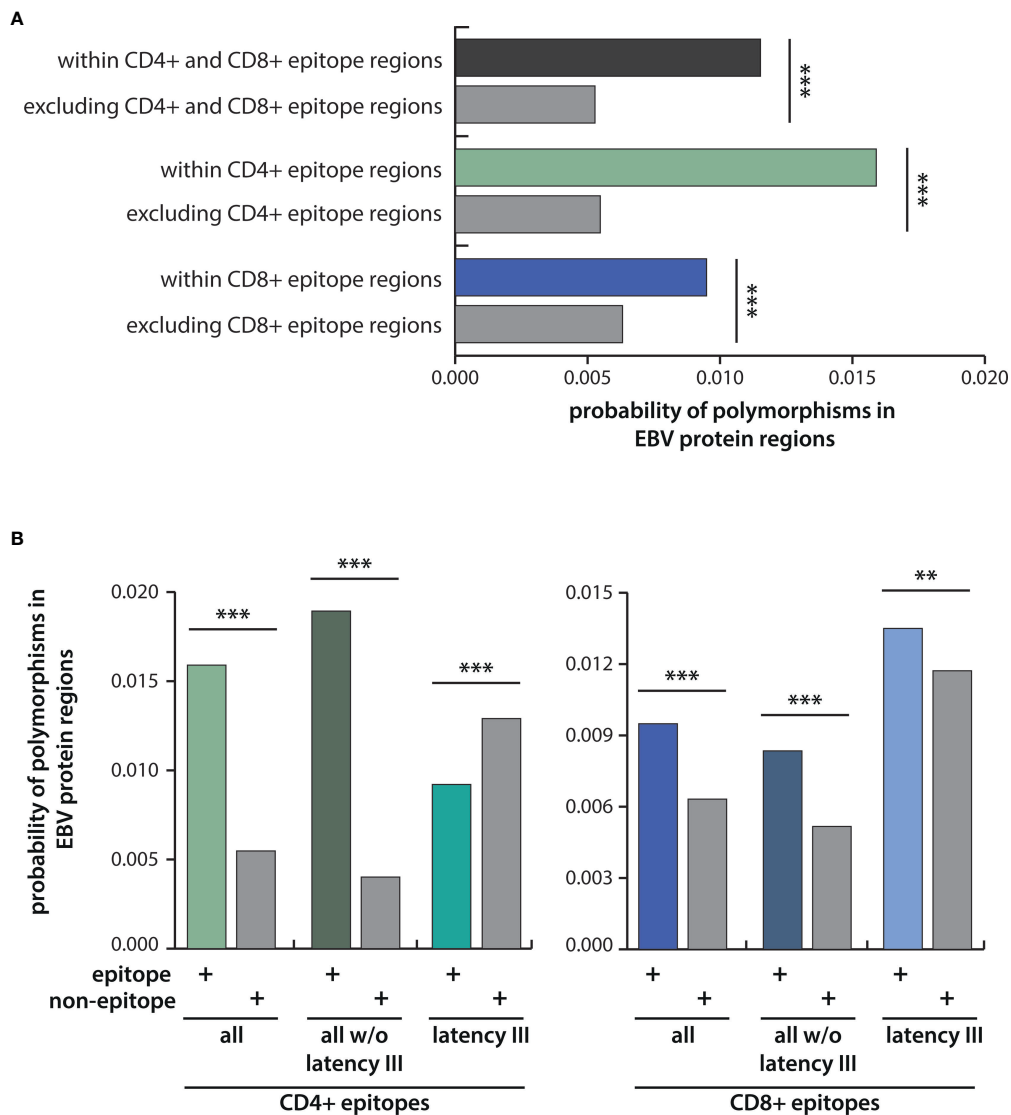
FIGURE 5 | Probabilities of polymorphisms occurring within epitope regions in comparison to the rest of the protein. **(A)** Depicted are the probabilities of amino acid exchanges occurring within T-cell epitopes in comparison to the rest of the protein for all CD4+ and CD8+ T-cell epitopes taken together, and for CD4+ and CD8+ T-cell epitopes separately. **(B)** Probabilities for polymorphisms in CD4+ (left) and CD8+ (right) T-cell epitopes were further analyzed for latency III antigens (EBNA2, EBNA-LP, EBNA3A, B, C) and for all antigens except latency III antigens. Statistical significance was tested with the Kolmogorov-Smirnov test, which was performed including each data point of the different EBV strains for all 32 EBV proteins using weight-adjusted probabilities (for details see Methods section). **p < 0.05; ***p < 0.005.

presentation, and binding to MHC and the T-cell receptor (54–58). Therefore, the impact of single amino acid exchanges on the T-cell response is generally difficult to predict. In the case of CD8+ T-cell epitopes, stable binding to MHC class I molecules depends on distinct anchor residues, usually the amino acids at position 2 and at the C-terminus (59). These residues were often found to be polymorphic and, in many cases, exchanged for amino acids from different biochemical subgroups. Similarly, polymorphisms in the first residue in the C-terminal flanking region often resulted in a switch in the amino acid subgroup. The C-terminus of most MHC I ligands is generated by proteasomal cleavage and although amino acid preferences downstream of the

scissile peptide bond seem to vary between model proteins, polar amino acids, particularly serine, were recently found to be enriched at this position (60, 61). Although still requiring experimental verification, these finding are suggestive of immune escape due to impaired epitope generation and MHC binding. However, due to differences in HLA genotype prevalence around the world, we cannot exclude that some of the variants may represent immunogenic epitopes in association with certain HLA molecules. Furthermore, no such analysis was performed for CD4+ T-cell epitopes because only few core sequences had been mapped. In addition, CD4+ T-cell epitopes usually contain four anchor residues and the contribution of single anchor amino acids

to MHC class II binding is still poorly defined. Nevertheless, the notion of antigenic drift was further substantiated when polymorphism rates in epitopes and non-epitope sequences were compared. Amino acid exchanges were significantly more frequent in CD4+ and CD8+ T-cell epitope sequences as compared to the rest of the proteins. The probability of a given polymorphism to be part a CD4+ T-cell epitope versus non-epitope region was almost threefold higher. For CD8+ T-cell epitopes, this probability was 1.5-fold increased. These results infer an important role of T-cell responses in driving virus diversification.

The question whether adaptive immunity is shaping the evolution of the virus is longstanding and has engendered controversial debates in the past following the discovery that EBV type 1 strains from highly HLA-A*11 positive Asian populations consistently display variations within two HLA-A*11-restricted, immunodominant CD8+ T-cell epitopes derived from EBNA3B, which render these epitope variants non-immunogenic *in vivo* (29, 62, 63). Such epitope loss would confer a selective advantage to strain variants in this particular host community and was interpreted by some as evidence of immune selection. Others, however, inferred random genetic drift because polymorphisms in these as well as additional epitopes from the EBNA3 family of proteins coherently mapped onto the phylogenetic tree of Chinese EBV type 1 strains and thereby implied a common viral origin of these strains. Moreover, such epitope-loss mutations were not replicated at other epitope loci (11, 33, 64, 65), which is consistent with recent findings showing that most positively selected codons in EBNA3 genes map to regions outside established CD8+ T-cell epitopes (5).

In contrast to these studies, which were often performed on a limited number of epitopes and viral isolates, the present analysis of 305 epitopes in more than 150 strains provides clear evidence that the T-cell response is driving virus diversification. However, immune pressure seems to shape the virus in opposite ways. While polymorphisms were generally found to be enriched in CD4+ and CD8+ T-cell epitopes, CD4+ T-cell epitopes in latency type III antigens appear to be under negative selection pressure, resulting in sequence diversification and conservation, respectively. While diversification is pointing towards immune evasion of the virus, the conservation of epitopes within the most variable viral proteins that are considered immunodominant targets of the virus-specific T-cell responses, appears more difficult to reconcile. Besides functional constraints, low variability in latency type III epitopes may have evolved in benefit of the virus. Since expression of these antigens is normally restricted to newly infected B cells, their conservation may ensure that any aberrantly expanding cell population expressing these antigens will be efficiently eliminated by specific T cells and not kill the host, as proposed earlier (66). In this case, however, a stronger negative selection pressure on epitopes recognized by cytotoxic CD8+ T cells would have been expected. In an alternative scenario, immune responses against these antigens may facilitate access of the virus to the memory compartment, the actual site of viral persistence. According to the germinal center model, EBV infects naïve B cells *in vivo* and expression of the growth program enables EBV-infected blasts to

migrate into follicles of secondary lymphoid organs where they switch to latency type I and leave the follicle as resting memory B cells (66). This transition in latency type is thought to depend on exogenous signals provided by activated CD4+ T cells (35, 66, 67), such as IL21 and soluble CD40L (67–69). The conservation of CD4+ T-cell epitopes in type III latency antigens may increase the likelihood of B cells infected with different viral strains for encountering and receiving T-cell help and thereby foster coinfections. Of note, increasing evidence suggests that coinfections are the rule, rather than the exception (70, 71).

If correct, this coinfection model would posit that persistent infection is a much more dynamic process than hitherto thought. For verifying this model, several predictions can be made and tested experimentally. For instance, do polymorphisms in latency type III epitopes have a milder immunological phenotype, i.e. lower impact on T-cell recognition, than epitopes under positive selection, do co-resident viral strains express mostly shared sets of latency type III epitopes, and do CD4+ T cells specific for these antigens form an integral part of the follicular CD4+ T-cell compartment? Ultimately, these studies may contribute to a better understanding of the biology of this virus and its intricate relationship with the human host, and inform on the targets best suited for clinical intervention.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

UB and JM conceived and designed the study. AC and RP collected and analyzed sequences. AC and MD performed the statistical analyses. KW provided information on HLA genotypes. H-JD contributed to the acquisition, analysis, and interpretation of data. AC and JM wrote the paper and all authors made substantial contributions to data analysis and interpretation, manuscript editing, review and approval.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.796379/full#supplementary-material

REFERENCES

1. Longnecker RM, Kieff E, Cohen JI. Epstein-Barr Virus. In: *Fields Virology*, *Sixth Edition*, vol. 2. Philadelphia: Wolters Kluwer Health Adis (ESP (2013). p. 1898–959.

2. Farrell PJ. Epstein-Barr Virus and Cancer. *Annu Rev Pathol* (2019) 14:29–53. doi: 10.1146/annurev-pathmechdis-012418-013023

3. Ascherio A, Munger KL. EBV and Autoimmunity. *Curr Top Microbiol Immunol* (2015) 390(Pt 1):365–85. doi: 10.1007/978-3-319-22822-8_15

4. Rickinson AB, Kieff E. Epstein-Barr Virus. In: DM Knipe, PM Howley, editors. *Fields Virology, 5th ed*. Philadelphia: Lippincott Williams & Wilkins (2007). p. 2575–627.

5. Palser AL, Grayson NE, White RE, Corton C, Correia S, Ba Abdullah MM, et al. Genome Diversity of Epstein-Barr Virus From Multiple Tumor Types and Normal Infection. *J Virol* (2015) 89(10):5222–37. doi: 10.1128/JVI.03614-14

6. Dambaugh T, Hennessy K, Chamnankit L, Kieff E. U2 Region of Epstein-Barr Virus DNA may Encode Epstein-Barr Nuclear Antigen 2. *Proc Natl Acad Sci USA* (1984) 81(23):7632–6. doi: 10.1073/pnas.81.23.7632

7. Zhou L, Chen JN, Qiu XM, Pan YH, Zhang ZG, Shao CK. Comparative Analysis of 22 Epstein-Barr Virus Genomes From Diseased and Healthy Individuals. *J Gen Virol* (2017) 98(1):96–107. doi: 10.1099/jgv.0.000699

8. Kwok H, Wu CW, Palser AL, Kellam P, Sham PC, Kwong DL, et al. Genomic Diversity of Epstein-Barr Virus Genomes Isolated From Primary Nasopharyngeal Carcinoma Biopsy Samples. *J Virol* (2014) 88(18):10662–72. doi: 10.1128/JVI.01665-14

9. Yao Y, Xu M, Liang L, Zhang H, Xu R, Feng Q, et al. Genome-Wide Analysis of Epstein-Barr Virus Identifies Variants and Genes Associated With Gastric Carcinoma and Population Structure. *Tumour Biol J Int Soc Oncodevelopmental Biol Med* (2017) 39(10):1010428317714195. doi: 10.1177/1010428317714195

10. Zanella L, Riquelme I, Buchegger K, Abanto M, Ili C, Brebi P. A Reliable Epstein-Barr Virus Classification Based on Phylogenomic and Population Analyses. *Sci Rep* (2019) 9(1):9829. doi: 10.1038/s41598-019-45986-3

11. Feederle R, Klinke O, Kutikhin A, Poirey R, Tsai MH, Delecluse HJ. Epstein-Barr Virus: From the Detection of Sequence Polymorphisms to the Recognition of Viral Types. *Curr topics Microbiol Immunol* (2015) 390(Pt 1):119–48. doi: 10.1007/978-3-319-22822-8_7

12. Tsai MH, Raykova A, Klinke O, Bernhardt K, Gartner K, Leung CS, et al. Spontaneous Lytic Replication and Epitheliotropism Define an Epstein-Barr Virus Strain Found in Carcinomas. *Cell Rep* (2013) 5(2):458–70. doi: 10.1016/j.celrep.2013.09.012

13. Coleman CB, Wohlford EM, Smith NA, King CA, Ritchie JA, Baresel PC, et al. Epstein-Barr Virus Type 2 Latently Infects T Cells, Inducing an Atypical Activation Characterized by Expression of Lymphotactic Cytokines. *J Virol* (2015) 89(4):2301–12. doi: 10.1128/JVI.03001-14

14. Tzellos S, Correia PB, Karstegl CE, Cancian L, Cano-Flanagan J, McClellan MJ, et al. A Single Amino Acid in EBNA-2 Determines Superior B Lymphoblastoid Cell Line Growth Maintenance by Epstein-Barr Virus Type 1 EBNA-2. *J Virol* (2014) 88(16):8743–53. doi: 10.1128/JVI.01000-14

15. Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop AD. The Immunology of Epstein-Barr Virus-Induced Disease. *Annu Rev Immunol* (2015) 33:787–821. doi: 10.1146/annurev-immunol-032414-112326

16. Long HM, Taylor GS, Rickinson AB. Immune Defence Against EBV and EBV-Associated Disease. *Curr Opin Immunol* (2011) 23(2):258–64. doi: 10.1016/j.coi.2010.12.014

17. Heslop HE, Rooney CM. Adoptive Cellular Immunotherapy for EBV Lymphoproliferative Disease. *Immunol Rev* (1997) 157:217–22. doi: 10.1111/j.1600-065X.1997.tb00984.x

18. Gottschalk S, Rooney CM. Adoptive T-Cell Immunotherapy. *Curr topics Microbiol Immunol* (2015) 391:427–54. doi: 10.1007/978-3-319-22834-1_15

19. Merlo A, Turrini R, Dolcetti R, Zanovello P, Rosato A. Immunotherapy for EBV-Associated Malignancies. *Int J Hematol* (2011) 93(3):281–93. doi: 10.1007/s12185-011-0782-2

20. Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained Complete Responses in Patients With Lymphoma Receiving Autologous Cytotoxic T Lymphocytes Targeting Epstein-Barr Virus Latent Membrane Proteins. *J Clin Oncol Off J Am Soc Clin Oncol* (2014) 32(8):798–808. doi: 10.1200/JCO.2013.51.5304

21. McLaughlin LP, Bollard CM, Keller MD. Adoptive T Cell Therapy for Epstein-Barr Virus Complications in Patients With Primary Immunodeficiency Disorders. *Front Immunol* (2018) 9:556. doi: 10.3389/fimmu.2018.00556

22. Miller G, Shope T, Lisco H, Stitt D, Lipman M. Epstein-Barr Virus: Transformation, Cytopathic Changes, and Viral Antigens in Squirrel Monkey and Marmoset Leukocytes. *Proc Natl Acad Sci USA* (1972) 69 (2):383–7. doi: 10.1073/pnas.69.2.383

23. Moss DJ, Misko IS, Burrows SR, Burman K, McCarthy R, Sculley TB. Cytotoxic T-Cell Clones Discriminate Between A- and B-Type Epstein-Barr Virus Transformants. *Nature* (1988) 331(6158):719–21. doi: 10.1038/331719a0

24. Burrows SR, Sculley TB, Misko IS, Schmidt C, Moss DJ. An Epstein-Barr Virus-Specific Cytotoxic T Cell Epitope in EBV Nuclear Antigen 3 (EBNA 3). *J Exp Med* (1990) 171(1):345–9. doi: 10.1084/jem.171.1.345

25. Burrows SR, Misko IS, Sculley TB, Schmidt C, Moss DJ. An Epstein-Barr Virus-Specific Cytotoxic T-Cell Epitope Present on A- and B-Type Transformants. *J Virol* (1990) 64(8):3974–6. doi: 10.1128/jvi.64.8.3974-3976.1990

26. Schmidt C, Burrows SR, Sculley TB, Moss DJ, Misko IS. Nonresponsiveness to an Immunodominant Epstein-Barr Virus-Encoded Cytotoxic T-Lymphocyte Epitope in Nuclear Antigen 3A: Implications for Vaccine Strategies. *Proc Natl Acad Sci USA* (1991) 88(21):9478–82. doi: 10.1073/pnas.88.21.9478

27. Apolloni A, Moss D, Stumm R, Burrows S, Suhrbier A, Misko I, et al. Sequence Variation of Cytotoxic T Cell Epitopes in Different Isolates of Epstein-Barr Virus. *Eur J Immunol* (1992) 22(1):183–9. doi: 10.1002/eji.1830220127

28. Lee SP, Thomas WA, Murray RJ, Khanim F, Kaur S, Young LS, et al. HLA A2.1-Restricted Cytotoxic T Cells Recognizing a Range of Epstein-Barr Virus Isolates Through a Defined Epitope in Latent Membrane Protein LMP2. *J Virol* (1993) 67(12):7428–35. doi: 10.1128/JVI.67.12.7428-7435.1993

29. Midgley RS, Bell AI, Yao QY, Croom-Carter D, Hislop AD, Whitney BM, et al. HLA-A11-Restricted Epitope Polymorphism Among Epstein-Barr Virus Strains in the Highly HLA-A11-Positive Chinese Population: Incidence and Immunogenicity of Variant Epitope Sequences. *J Virol* (2003) 77(21):11507–16. doi: 10.1128/JVI.77.21.11507-11516.2003

30. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, Sullivan LC, et al. Allelic Polymorphism in the T Cell Receptor and Its Impact on Immune Responses. *J Exp Med* (2010) 207(7):1555–67. doi: 10.1084/jem.20100603

31. Bell MJ, Brennan R, Miles JJ, Moss DJ, Burrows JM, Burrows SR. Widespread Sequence Variation in Epstein-Barr Virus Nuclear Antigen 1 Influences the Antiviral T Cell Response. *J Infect Dis* (2008) 197(11):1594–7. doi: 10.1086/587848

32. Brooks JM, Croom-Carter DS, Leese AM, Tierney RJ, Habeshaw G, Rickinson AB. Cytotoxic T-Lymphocyte Responses to a Polymorphic Epstein-Barr Virus Epitope Identify Healthy Carriers With Coresident Viral Strains. *J Virol* (2000) 74(4):1801–9. doi: 10.1128/JVI.74.4.1801-1809.2000

33. Burrows JM, Burrows SR, Poulsen LM, Sculley TB, Moss DJ, Khanna R. Unusually High Frequency of Epstein-Barr Virus Genetic Variants in Papua New Guinea That Can Escape Cytotoxic T-Cell Recognition: Implications for Virus Evolution. *J Virol* (1996) 70(4):2490–6. doi: 10.1128/jvi.70.4.2490-2496.1996

34. Cirac A, Stutzle S, Dieckmeyer M, Adhikary D, Moosmann A, Korber N, et al. Epstein-Barr Virus Strain Heterogeneity Impairs Human T-Cell Immunity. *Cancer Immunol Immunother CII* (2018) 67(4):663–74. doi: 10.1007/s00262-018-2118-z

35. Longnecker RM, Kieff E, Cohen JI. Epstein-Barr Virus. In: DM Knipe, PM Howley, editors. *Fields Virology, 6th ed*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins (2013). p. 1898–959.

36. Hislop AD, Taylor GS, Sauce D, Rickinson AB. Cellular Responses to Viral Infection in Humans: Lessons From Epstein-Barr Virus. *Annu Rev Immunol* (2007) 25:587–617. doi: 10.1146/annurev.immunol.25.022106.141553

37. Taylor GS, Steven NM. Therapeutic Vaccination Strategies to Treat Nasopharyngeal Carcinoma. *Chin Clin Oncol* (2016) 5(2):23. doi: 10.21037/cco.2016.03.20

38. Smith C, Khanna R. The Development of Prophylactic and Therapeutic EBV Vaccines. *Curr topics Microbiol Immunol* (2015) 391:455–73. doi: 10.1007/978-3-319-22834-1_16

39. Cohen JI. Epstein-Barr Virus Vaccines. *Clin Trans Immunol* (2015) 4(1):e32. doi: 10.1038/cti.2014.27

40. Bollard CM, Heslop HE. T Cells for Viral Infections After Allogeneic Hematopoietic Stem Cell Transplant. *Blood* (2016) 127(26):3331–40. doi: 10.1182/blood-2016-01-628982

41. Moosmann A, Bigalke I, Tischer J, Schirrmann L, Kasten J, Tippmer S, et al. Effective and Long-Term Control of EBV PTLD After Transfer of Peptide-Selected T Cells. *Blood* (2010) 115(14):2960–70. doi: 10.1182/blood-2009-08-236356

42. Icheva V, Kayser S, Wolff D, Tuve S, Kyzirakos C, Bethge W, et al. Adoptive Transfer of Epstein-Barr Virus (EBV) Nuclear Antigen 1-Specific T Cells as Treatment for EBV Reactivation and Lymphoproliferative Disorders After Allogeneic Stem-Cell Transplantation. *J Clin Oncol Off J Am Soc Clin Oncol* (2013) 31(1):39–48. doi: 10.1200/JCO.2011.39.8495

43. Haque T, Wilkie GM, Jones MM, Higgins CD, Urquhart G, Wingate P, et al. Allogeneic Cytotoxic T-Cell Therapy for EBV-Positive Posttransplantation Lymphoproliferative Disease: Results of a Phase 2 Multicenter Clinical Trial. *Blood* (2007) 110(4):1123–31. doi: 10.1182/blood-2006-12-063008

44. O'Reilly RJ, Prockop S, Hasan AN, Koehne G, Doubrovina E. Virus-Specific T-Cell Banks for 'Off the Shelf' Adoptive Therapy of Refractory Infections. *Bone Marrow Transplant* (2016) 51(9):1163–72. doi: 10.1038/bmt.2016.17

45. Bollard CM. Improving T-Cell Therapy for Epstein-Barr Virus Lymphoproliferative Disorders. *J Clin Oncol Off J Am Soc Clin Oncol* (2013) 31(1):5–7. doi: 10.1200/JCO.2012.43.5784

46. Steven NM, Leese AM, Annels NE, Lee SP, Rickinson AB. Epitope Focusing in the Primary Cytotoxic T Cell Response to Epstein-Barr Virus and Its Relationship to T Cell Memory. *J Exp Med* (1996) 184(5):1801–13. doi: 10.1084/jem.184.5.1801

47. Gottschalk S, Ng CY, Perez M, Smith CA, Sample C, Brenner MK, et al. An Epstein-Barr Virus Deletion Mutant Associated With Fatal Lymphoproliferative Disease Unresponsive to Therapy With Virus-Specific CTLs. *Blood* (2001) 97(4):835–43. doi: 10.1182/blood.V97.4.835

48. Houssaint E, Saulquin X, Scotet E, Bonneville M. Immunodominant CD8 T Cell Response to Epstein-Barr Virus. *BioMed Pharmacother* (2001) 55(7):373–80. doi: 10.1016/S0753-3322(01)00082-8

49. Delecluse HJ, Feederle R, Behrends U, Mautner J. Contribution of Viral Recombinants to the Study of the Immune Response Against the Epstein-Barr Virus. *Semin Cancer Biol* (2008) 18(6):409–15. doi: 10.1016/j.semcancer.2008.09.001

50. Kloverpris HN, Leslie A, Goulder P. Role of HLA Adaptation in HIV Evolution. *Front Immunol* (2015) 6:665. doi: 10.3389/fimmu.2015.00665

51. Li ZT, Zarnitsyna VI, Lowen AC, Weissman D, Koelle K, Kohlmeier JE, et al. Why Are CD8 T Cell Epitopes of Human Influenza A Virus Conserved? *J Virol* (2019) 93(6):1–13. doi: 10.1128/JVI.01534-18

52. Coscolla M, Copin R, Sutherland J, Gehre F, de Jong B, Owolabi O, et al. M. Tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens. *Cell Host Microbe* (2015) 18(5):538–48. doi: 10.1016/j.chom.2015.10.008

53. Cirac Ana BU, Mautner J. Clinical Implications of Epstein-Barr Virus Strain Diversity. *J Immunol Sci* (2018) 2(3):51–5. doi: 10.29245/2578-3009/2018/3.1145

54. Ossendorp F, Eggers M, Neisig A, Ruppert T, Groettrup M, Sijts A, et al. A Single Residue Exchange Within a Viral CTL Epitope Alters Proteasome-Mediated Degradation Resulting in Lack of Antigen Presentation. *Immunity* (1996) 5(2):115–24. doi: 10.1016/S1074-7613(00)80488-4

55. Hastings KT. GILT: Shaping the MHC Class II-Restricted Peptidome and CD4(+) T Cell-Mediated Immunity. *Front Immunol* (2013) 4:429. doi: 10.3389/fimmu.2013.00429

56. Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules. *Annu Rev Immunol* (2015) 33:169–200. doi: 10.1146/annurev-immunol-032414-112334

57. Neefjes J, Ovaa H. A Peptide's Perspective on Antigen Presentation to the Immune System. *Nat Chem Biol* (2013) 9(12):769–75. doi: 10.1038/nchembio.1391

58. Cole DK, Gallagher K, Lemercier B, Holland CJ, Junaid S, Hindley JP, et al. Modification of the Carboxy-Terminal Flanking Region of a Universal Influenza Epitope Alters CD4(+) T-Cell Repertoire Selection. *Nat Commun* (2012) 3:665. doi: 10.1038/ncomms1665

59. Gfeller D, Bassani-Sternberg M. Predicting Antigen Presentation-What Could We Learn From a Million Peptides? *Front Immunol* (2018) 9:1716. doi: 10.3389/fimmu.2018.01716

60. Groettrup M, Kirk CJ, Basler M. Proteasomes in Immune Cells: More Than Peptide Producers? *Nat Rev Immunol* (2010) 10(1):73–8. doi: 10.1038/nri2687

61. Martin LK, Hollaus A, Stahuber A, Hubener C, Fraccaroli A, Tischer J, et al. Cross-Sectional Analysis of CD8 T Cell Immunity to Human Herpesvirus 6B. *PloS Pathog* (2018) 14(4):e1006991. doi: 10.1371/journal.ppat.1006991

62. de Campos-Lima PO, Gavioli R, Zhang QJ, Wallace LE, Dolcetti R, Rowe M, et al. HLA-A11 Epitope Loss Isolates of Epstein-Barr Virus From a Highly A11+ Population. *Science* (1993) 260(5104):98–100. doi: 10.1126/science.7682013

63. Midgley RS, Bell AI, McGeoch DJ, Rickinson AB. Latent Gene Sequencing Reveals Familial Relationships Among Chinese Epstein-Barr Virus Strains and Evidence for Positive Selection of A11 Epitope Changes. *J Virol* (2003) 77(21):11517–30. doi: 10.1128/JVI.77.21.11517-11530.2003

64. Khanna R, Slade RW, Poulsen L, Moss DJ, Burrows SR, Nicholls J, et al. Evolutionary Dynamics of Genetic Variation in Epstein-Barr Virus Isolates of Diverse Geographical Origins: Evidence for Immune Pressure-Independent Genetic Drift. *J Virol* (1997) 71(11):8340–6. doi: 10.1128/jvi.71.11.8340-8346.1997

65. Lee SP, Morgan S, Skinner J, Thomas WA, Jones SR, Sutton J, et al. Epstein-Barr Virus Isolates With the Major HLA B35.01-Restricted Cytotoxic T Lymphocyte Epitope Are Prevalent in a Highly B35.01-Positive African Population. *Eur J Immunol* (1995) 25(1):102–10. doi: 10.1002/eji.1830250119

66. Thorley-Lawson DA. EBV Persistence–Introducing the Virus. *Curr topics Microbiol Immunol* (2015) 390(Pt 1):151–209. doi: 10.1007/978-3-319-22822-8_8

67. Nagy N, Adori M, Rasul A, Heuts F, Salamon D, Ujvari D, et al. Soluble Factors Produced by Activated CD4+ T Cells Modulate EBV Latency. *Proc Natl Acad Sci USA* (2012) 109(5):1512–7. doi: 10.1073/pnas.1120587109

68. Klein E, Nagy N, Rasul E. Modification of Cell Differentiation, One of the Mechanisms in the Surveillance of Malignancy. *Cancer Immunol Res* (2015) 3(2):97–102. doi: 10.1158/2326-6066.CIR-14-0238

69. Heuts F, Rottenberg ME, Salamon D, Rasul E, Adori M, Klein G, et al. T Cells Modulate Epstein-Barr Virus Latency Phenotypes During Infection of Humanized Mice. *J Virol* (2014) 88(6):3235–45. doi: 10.1128/JVI.02885-13

70. Srivastava G, Wong KY, Chiang AK, Lam KY, Tao Q. Coinfection of Multiple Strains of Epstein-Barr Virus in Immunocompetent Normal Individuals: Reassessment of the Viral Carrier State. *Blood* (2000) 95(7):2443–5. doi: 10.1182/blood.V95.7.2443.007k18_2443_2445

71. Weiss ER, Lamers SL, Henderson JL, Melnikov A, Somasundaran M, Garber M, et al. Early Epstein-Barr Virus Genomic Diversity and Convergence Toward the B95.8 Genome in Primary Infection. *J Virol* (2018) 92(2):1–19. doi: 10.1128/JVI.01466-17

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.