Frontiers in Bioscience Scholar

Front. Biosci. (Schol Ed) **2026**; 18(1): 46442
https://doi.org/10.31083/FBS46442

Original Research

Molecular Characterization of the Epstein–Barr Virus in Malignant Tumors of Different Origins: An Emphasis on *EBER* Promoter Value for EBV Classification

Valerija Begić[1,2], Petra Šimičić[3], Kristian Bodulić[4], Slavko Gašparov[5,6], Suzana Hančić[5], Roberto Galov[1], Marija Rozman[7], Paula Gršković[1], Snježana Židovec-Lepej[7], Petra Korać[1,*]

[1]Division of Molecular Biology, Department of Biology, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia
[2]Primary School "Sesvetski Kraljevec", 10361 Sesvetski Kraljevec, Croatia
[3]Department of Clinical Chemistry, University Hospital Centre Sestre Milosrdnice, 10000 Zagreb, Croatia
[4]Department for Bioinformatics and Statistics, University Hospital for Infectious Diseases, 10000 Zagreb, Croatia
[5]Institute of Clinical Pathology and Cytology, Merkur University Hospital, 10000 Zagreb, Croatia
[6]Department of Pathology, Medical School Zagreb, University of Zagreb, 10000 Zagreb, Croatia
[7]Department of Immunological and Molecular Diagnostics, University Hospital for Infectious Diseases, 10000 Zagreb, Croatia
*Correspondence: petra.korac@biol.pmf.hr (Petra Korać)
Academic Editor: Gustavo Caetano-Anollés
Submitted: 9 September 2025 Revised: 1 December 2025 Accepted: 16 December 2025 Published: 23 March 2026

Abstract

Background: Epstein–Barr virus (EBV) is a well-known cause of infectious mononucleosisand it has also been linked to the development of various malignant tumors. EBV can be classified into two genotypes based on *EBNA-2* and *EBNA-3* polymorphisms, with further subclassification predominantly based on *BNLF-1* and *EBNA-1* variations. Currently, EBV subtyping has not been performed using *EBER* genes. Thus, this study aimed to determine EBV subtypes in various malignant tumors and to assess the potential of *EBER* polymorphisms for EBV classification. **Methods**: DNA was isolated from tissue samples of patients diagnosed with classical Hodgkin lymphoma (cHL), angioimmunoblastic T cell lymphoma (AITL), and nasopharyngeal carcinoma (NPC). Specific fragments of the *BNLF-1*, *EBNA-1*, *EBNA-3C*, and *EBER* genes were amplified, and population sequencing was conducted. **Results**: Based on *EBNA-3C* sequences, all tumor samples were genotype EBV-1, whereas the *BNLF-1*, *EBNA-1*, and *EBER* sequences indicated that the tumor tissues were infected with different EBV subtypes. The B95-8 subtype predominated in cHL, whereas North Carolina and P-ala were the most frequent subtypes in NPC. In addition, we propose a newly designed algorithm for *EBER*-based subclassification, which showed that the subtype Cro2 is more frequently present in cHL and AITL than in NPC. **Conclusion**: The results of this study show characteristic pattern of EBV diversity based on the *EBNA-3C*, *EBNA-1*, and *BNLF-1* gene sequences, as well as the *EBER* promoter sequence in AITL, NPC, and cHL cohorts specific for the Caucasian population of southeastern Europe; however, these results may not relate to the distribution of EBV variants in other geographic areas.

Keywords: Epstein–Barr virus; classical Hodgkin lymphoma; angioimmunoblastic T cell lymphoma; nasopharyngeal carcinoma; genotyping; phylogenetics; *EBER*

1. Introduction

Formally named human herpesvirus 4 (HHV-4), the Epstein–Barr virus (EBV) is one of eight known human herpesviruses (*Herpesviridae* family, genus *Lymphocryptovirus*). Notably, over 90% of the global population is estimated to be infected with EBV [1], which causes infectious mononucleosis [2] and has been associated with the development of various malignancies [3]. The virus exhibits tropism for oropharyngeal epithelial cells, B cells, T cells, and natural killer (NK) cells [4].

The EBV replication cycle consists of lytic and latent phases, each characterized by distinct gene expression patterns. Five protein products are essential for B-cell transformation during the latent phase. This includes proteins encoded by four Epstein–Barr nuclear antigen genes (*EBNA-1*, *EBNA-2*, *EBNA-3A*, and *EBNA-3C*) and the Latent Membrane Protein 1 (LMP-1) encoded by the BamHI-N-fragment latent infection membrane protein 1 gene (*BNLF-1*) [1].

The EBV genome consists of double-stranded DNA of approximately 172 kilobase pairs (kb) in length. It was first sequenced in 1984 [5], with additional sequences reported over the next four decades [6], further complicating EBV classification. The simplest classification of EBV is based on variations in the *EBNA-2* gene, dividing EBV variants into type I (or type A) and type II (or type B) [1]. To a lesser extent, the *EBNA-3* gene sequence can also be used for EBV classification [7]. The two EBV types are characterized by different host immunobiological responses, as well as by distinct geographical distribution pat-

terns, with type 1 (including wild-type strain B95-8, Akata, Mutu, etc.) showing global prevalence and type 2 (for example, strain Ag876) usually restricted to sub-Saharan Africa, Papua New Guinea, and Alaska [6,8,9].

Based on the LMP-1 carboxyl end sequence and the associated difference from the wild-type strain B95-8 [5], EBV can also be classified into seven main variants: Alaskan, China 1, China 2, China 3, Mediterranean+ (Med+), Med-, and North Carolina (NC) [10]. LMP-1 is an essential EBV oncogenic protein that constitutively stimulates several signal transduction pathways, including those mediated by NF-κB, PI3K/AKT, and JAK/STAT, leading to the immortalization of B-cells [11]. Associations between variants on the C-terminal end of LMP-1 and the frequency of various malignant diseases have been studied with inconsistent results [12]. A 30-base-pair (bp) deletion in the *BNLF-1* gene (gene coding for the protein LMP-1) reduces the immunogenetic potential of LMP-1. This deletion is associated with a prolonged LMP-1 protein half-life and is considered an indication of a more aggressive clinical phenotype. In addition, a single nucleotide substitution G169425T results in the loss of the XhoI restriction site at the cytoplasmic N-terminal tail of LMP-1. Both of these polymorphisms have been detected at increased frequency in nasopharyngeal carcinoma (NPC) compared to samples from healthy controls [12].

Based on the amino acid variations at position 487 in EBNA-1, EBV can be classified into two prototype sequences, P-ala (B95-8 prototype) and P-thr, and three variant sequences, V-val, V-leu, and V-pro. Furthermore, EBV can be categorized into subvariants based on 22 additional polymorphic loci and variations in five amino acids present at the C-terminal end of EBNA-1 [13,14]. The increased frequency of the V-val variant has been documented in NPC from endemic areas and in lymphoma patients [13,14]. EBNA-1 is one of the essential viral proteins that plays a key role in EBV episome maintenance and episome proper segregation during mitosis through the interaction with the replication origin site OriP.

EBV-encoded small RNAs (EBERs) encode small non-coding RNAs involved in transformation and oncogenesis. EBER sequences are usually not used for virus classification, but EBERs are the main markers for EBV detection in malignant cells. Hui *et al*. [15] have identified a four-base deletion downstream of the *EBER-2* gene (EBER-del, coordinates 7188–7191) that is present in 96.8% of NPC cases and in 40.1% of population carriers from Hong Kong. This suggests that high-risk EBV subtypes with polymorphisms in the *EBER* locus may play an important role in EBV-associated malignancies [15].

Thus, changes in EBV classifications are expected as EBV sequence databases expand and analytical tools develop [16]. However, Sanger sequencing of the *BNLF-1* and *EBNA* genes remains the main tool for EBV strain determination.

In a previous study, we analyzed the *EBNA-3C* and *BNLF-1* gene sequences in classical Hodgkin lymphoma (cHL) [17]. Our results showed exclusive presence of EBV genotype 1, with a narrow repertoire of LMP-1 variants, including the B95-8 wild-type and the Mediterranean subtype with a 30 bp deletion, in the Croatian population of patients diagnosed with cHL [17]. This study expanded on the previous research by including additional EBV-related gene sequences (*EBNA-1* and *EBER*) and samples from patients diagnosed with angioimmunoblastic T cell lymphoma (AITL) and NPC.

T cell lymphomas are a heterogeneous group of malignant neoplasms that are difficult to classify, diagnose, and treat due to their diversity and low incidence. Due to their heterogeneity, T-cell lymphomas are divided into multiple subtypes based on their morphologic and molecular characteristics. The largest subclass of T cell lymphomas mainly comprises nodal tumors derived from mature T lymphocytes that have exited the thymus and entered the peripheral bloodstream. Therefore, these cells are termed peripheral T cell lymphomas (PTCLs) [18]. AITL represents one of the most common subtypes of PTCLs, originating from follicular helper T lymphocytes (T_{FH}) [19] and accounting for 15–30% of non-cutaneous T cell lymphomas and 1–2% of all non-Hodgkin lymphomas (NHL) [18]. Unlike most T cell lymphoma subtypes, AITL has a higher incidence in Europe (28.7% of all T cell lymphomas) than in Asia (17.9%) or North America (16%) [20]. Furthermore, AITL is more common in the middle-aged and older populations, with a higher incidence in men [21]. The poor prognosis for this type of lymphoma is additionally worsened by the presence of EBV [22,23]. While tumor cells are generally EBV-negative [18], EBV has been detected in 60–90% of AITL cases in the surrounding large B lymphocytes [22,24–27]; however, the EBV molecular diversity in AITL is currently unknown.

NPC is another tumor associated with EBV and is an epithelial tumor belonging to the group of head and neck carcinomas (HNCs). NPC accounts for 0.7% of the global cancer burden and is more common in men [28]. Several factors are associated with the risk of NPC development, including diet, smoking, and drinking habits [29], as well as human leukocyte antigen (HLA) genetics [30]. Most studies focus on the role of EBV in NPC biology, given the universal presence of EBV in NPC [31–33], prioritizing endemic areas with high NPC frequency.

This study aimed to analyze the distribution of EBV genotypes and variants based on *EBNA* and *BNLF-1* sequences and to evaluate the potential significance of *EBER* in EBV classification in malignant tumors, including cHL, AITL, and NPC from a non-endemic region. To our knowledge, this is the first study to analyze EBV molecular diversity in AITL and the first proposal of an *EBER*-based algorithm for EBV classification.

2. Materials and Methods

2.1 Tumor Tissue Samples

This study used tumor tissue samples from 91 patients. All samples were formalin-fixed paraffin-embedded (FFPE) blocks taken from Merkur University Hospital's archive. EBV infection was previously confirmed in all samples by immunohistochemical staining for LMP-1 and/or *in situ* hybridization for EBERs (EBER-ISH). The study was approved by the Ethics committee of the Merkur University Hospital, Zagreb, Croatia, on 25th September 2019 (UR. BR. 03/1-8793). For this study, informed consent was obtained from the patients or their families/legal representatives.

The group of EBV-positive patients analyzed in this study included: 52 patients diagnosed with cHL, 37 males and 15 females (median age: 56 years, range: 10–87 years), 11 patients diagnosed with AITL, 6 males and 5 females (median age: 70 years, range: 49–82 years), and 28 patients with NPC, 20 males and 8 females (median age: 56 years, range: 26–89 years). All patients included in the study were Caucasians and Croatian citizens.

2.2 DNA Extraction and PCR

DNA was extracted from the FFPE samples using a commercially available Quick-DNA/RNA FFPE Miniprep kit (ZymoResearch, Tustin, CA, USA) according to the manufacturer's instructions. Specific EBV sequences were amplified by PCR using GoTaq DNA polymerase (M300A, Promega, Madison, WI, USA) (**Supplementary Table 1**). The cHL tumor tissue samples were used for the *EBNA-1* and *EBER* analyses, while the AITL and NPC tumor tissue samples were used for the *BNLF-1*, *EBNA-3C*, *EBNA-1,* and *EBER* analyses. Data for the *BNLF-1* and *EBNA-3C* sequences in the same cHL samples used in this study were taken from the previously published paper [17].

2.3 DNA Sequencing

PCR products were purified using the ChargeSwitch PCR Clean-Up kit (CS12000-10, Life Technologies, Carlsbad, CA, USA). The purified PCR products were sent to Macrogen (Seoul, Republic of Korea) for Sanger sequencing.

2.4 Sequence Analysis

Forward and reverse sequences were overlapped using Benchling (Benchling Inc., San Francisco, CA, USA). The program was set to default settings, and the obtained consensus sequences were used for further analysis. Alignment of the consensus sequences was conducted using ClustalX v.2.1. (https://clustalx.software.informer.com/2.1/) [34]. Further analysis and sequence editing were performed in AliView v.1.17.1 (https://aliview.software.informer.com/) [35]. Variable-length sequence ends were clipped, and base readings were corrected manually if necessary.

Data for the *BNLF-1* and *EBNA-3C* sequences in the cHL samples were obtained from a previously published paper, in which the sequences had been analyzed using the same samples as in this study [17]. Analysis of the *BNLF-1* and *EBNA-3C* sequences in the AITL and NPC samples was performed as previously described for the cHL samples [17]. In short, the EBV genotype was determined using the *EBNA-3C* gene amplicon. The 153 bp amplicon was classified as the *EBV-1* genotype, and the 246 bp amplicon was classified as *EBV-2* [36]. The *BNLF-1* gene was used to determine six EBV subtypes: B95-8, China 1, China 2, Mediterranean (two subvariants: Med+, with deletion, and Med-, without deletion), Alaskan, and NC. Subtypes were determined based on LMP-1 C-terminal region sequence variability using an algorithm based on characteristic changes of amino acids at precisely defined positions (**Supplementary Table 2**). Analyzed sequences were compared with the reference sequence B95-8 (accession number V01555.2), which was downloaded from the GenBank database [10].

Analysis of the C-terminal region of the EBNA-1 protein in all tumor samples (cHL, AITL, and NPC) was performed as follows: classification into prototypes and variants was based on the sequence diversity of the C-terminal region. Determination of prototypes (P-ala and P-thr) and variants (V-val, V-leu, and V-pro) was based on the differences in amino acids at position 487, as well as on amino acid substitutions after position 487. We also identified EBNA-1-associated subvariants. The variant V-ala, which does not exhibit a mutation at position 487, but did at six other positions compared to the prototype sequence, was added to the existing classification (**Supplementary Table 3**) [13,37–43]. The obtained *EBNA-1* sequences were compared with the reference sequence B95-8 (accession number V01555.2) downloaded from GenBank.

2.5 EBER Evaluation

We designed an algorithm for determining EBV variants based on *EBER* promoter variability. Subsequently, two sequences of the prototype B95-8 were identified in the National Center for Biotechnology Information (NCBI) genome database. Both sequences contained mutually identical *EBER* promoter sequences of exactly 150 bp in the same locus. The *EBER* sequence of the prototype B95-8 (accession V01555) was then used to determine the precise position of the *EBER* promoter in the EBV genome and as a reference sequence in subsequent analyses (Table 1). Each *EBER* promoter sequence was compared with the reference sequence using the program MEGA X v.10.2.6 (https://www.megasoftware.net/) [44] with default settings. The exact position of these changes was marked in the EBV genome. Using the obtained data, the analyzed sequences were grouped based on similarity, and a new algorithm for EBV classification based on the *EBER* promoter sequence was created. The exact positions used to determine the sub-

Table 1. Prototype B95-8 *EBER* promoter sequence.

Access number	Length (bp)	*EBER* promoter sequence
V01555-8	150	gatccaaacttttgttttaggatttatgcatccattatcccgcagttccacctaaacgggcttaacgttgcatccca gaagatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc

Nucleotides in red show positions used in proposed EBV classification based on the *EBER* promoter sequence. EBV, Epstein–Barr virus.

Table 2. Nucleotide substitutions in the *EBER* promoter region.

Position in the genome	EBV subtypes and subvariants						
	Cro1		Cro2				
	Cro1^{i}	Cro1^{ii}	Cro2^{i}	Cro2^{ii}	Cro2^{iii}	Cro2^{iv}	Cro2^{v}
6829				A→G			
6834					C→T		
6854		G→A					
6856	G→T	G→T				G→T	
6884			G→A	G→A	G→A	G→A	G→A
6886			T→G	T→G	T→G	T→G	T→G
6911			A→G	A→G	A→G	A→G	A→G
6915							G→A

B95-8, prototype EBV strain; Cro1^{i}, subtype with substitution G→T (position of the genome 6856); Cro1^{ii}, subtype with substitutions G→A (6854) and G→T (6856); Cro2^{i}, subtype with substitutions G→A (6884), T→G (6886) and A→G (6911); Cro2^{ii}, subtype with substitutions A→G (6829), G→A (6884), T→G (6886) and A→G (6911); Cro2^{iii}, subtype with substitutions C→T (6834), G→A (6884), T→G (6886) and A→G (6911); Cro2^{iv}, subtype with substitutions G→T (6856), G→A (6884), T→G (6886) and A→G (6911); Cro2^{v}, subtype with substitutions G→A (6884), T→G (6886), A→G (6911) and G→A (6915).

types and subvariants are marked in red in the *EBER* sequence in Table 1. The observed nucleotide substitutions are shown in Table 2. The newly described subtypes and variants were named Cro1 and Cro2 for easier identification.

In addition, GenBank was searched for cHL, AITL, and NPC samples that contain sequences for all genes investigated in this study. EBV sequences found in this way were analyzed using the EBV classification algorithm based on the EBER promoter. Moreover, the newly proposed *EBER* algorithm was used for comparative analyses of high-risk EBER sequences published by Hui *et al*. [15]

2.6 Statistical Analysis

In addition to the results from this study, we used data from our previous research [17] to evaluate EBV molecular characterization in malignant tumors. Numerical variables are reported using the mean and standard deviation (SD), while categorical variables are reported using counts and percentages. The normality of numerical variable distributions was assessed graphically and with the Shapiro–Wilk test. Normally distributed variables were compared using the analysis of variance (ANOVA). The association of categorical variables was assessed using the chi-square test or Fisher's exact test, as appropriate. Confidence intervals (CIs) for the odds ratio (OR) were calculated using the conditional maximum likelihood estimate method. The *p*-values were corrected for multiple testing using the Benjamini–Hochberg method. All tests were two-tailed, with a significance level of $p < 0.05$. Statistical analysis was performed using R (v. 4.4.1, https://cran.r-project.org/bin/windows/base/old/).

The same set of analyses was used for sequences determined using GenBank data to validate the EBER-based algorithm.

3. Results

3.1 BNLF-1 Gene

Edwards' algorithm was used to determine the EBV subtypes based on the *BNLF-1* gene in tumor samples. As described previously, in the 46 cHL samples, three subtypes were detected: B95-8, Mediterranean (subvariant with deletion, Med+, and subvariant without deletion, Med-), and North Carolina [17]. In the 10 AITL samples, two subtypes were detected: the Mediterranean subtype in five (subvariants Med+ and Med-) and the B95-8 subtype in three. In two samples, a co-infection with subtypes B95-8 and Med- was detected (**Supplementary Table 4**).

In the 26 NPC samples, the B95-8, Mediterranean (subvariants Med+ and Med-), and North Carolina subtypes were detected, while co-infection was found in two samples (subtypes Med+ and NC). The Mediterranean subtype was the most abundant variant, detected in 10 of 26 samples.

Next were the B95-8 and North Carolina subtypes, noted in 7 and 6 out of 26 samples, respectively. In one sample, it was not possible to determine the subtype according to the Edwards' algorithm (**Supplementary Table 4**).

All detected deletions in the AITL and NPC samples were 30 bp in length and corresponded to the Mediterranean subtype. These deletions were detected in 4 of 10 AITL samples and 7 of 26 NPC samples. The deleted nucleotides were at positions 168285–168256, corresponding to amino acids 343–352. The China 1, China 2, and Alaskan subtypes were not detected. The overall distribution of EBV subtypes based on the *BNLF-1* gene in different malignant tumors is shown in **Supplementary Table 5**; there were no statistically significant differences between age, sex, and EBV subtypes ($p > 0.05$).

The B95-8, Mediterranean, and North Carolina subtypes were detected in both AITL and NPC samples. The North Carolina subtype was significantly more represented in NPC compared to AITL samples (OR = 7.86, 95% CI 1.27–86.05; $p = 0.016$) and was significantly more represented in secondary compared to primary tumor types, *i.e.*, in tumors of epithelial origin than in tumors of lymphocytic origin (OR = 7.69, 95% CI 1.17–50.20; $p = 0.016$).

The final length of the *BNLF-1* gene sequences was 157 bp or 127 bp (for sequences with the 30 bp deletion), corresponding to nucleotide positions 168194 to 168350 of the reference genome B95-8. In the phylogenetic analysis, the most likely substitution model for the combined data set was the JC + G model. The resulting phylogenetic tree showed that the sequences did not cluster by diagnosis, tumor stage, or cell of origin; instead, the sequences clustered into distinct clades according to EBV subtype. The Mediterranean subtype clustered into more than one clade, while co-infections clustered into one of the subtypes represented. Branching support was less than 70% except in the two clustering cases (Fig. 1, Ref. [17]).

Edwards' algorithm was also used to determine EBV subtypes based on the *BNLF-1* gene for 53 sequences obtained from GenBank, which were used to validate the EBER algorithm. We were unable to identify cHL samples with sequences for all genes investigated in this study, therefore additional analyses of the investigated gene was performed on 36 EBV sequences from the NPC samples (NCBI accession numbers: HQ020558.1, LC137018.1, KJ411974.1, KF992565.1, KF992566.1, KF992567.1, KF992568.1, KF992569.1, KF992570.1, KF992571.1, AB850643.1, AB850644.1, AB850645.1, AB850646.1, AB850647.1, AB850648.1, AB850649.1, AB850650.1, AB850651.1, AB850652.1, AB850653.1, AB850654.1, AB850655.1, AB850656.1, AB850657.1, AB850658.1, AB850659.1, AB850660.1, LC150743.1, LC150742.1, LC150741.1, LC150338.1, LC150337.1, LC150327.1, LC149491.1, AY961628.3) and 17 sequences from the AITL samples (NCBI accession numbers: MH837528.1, MH837527.1, MH837526.1, MH837525.1, MH837524.1, MH837523.1, MH837522.1, MH837521.1, MH837520.1, MH837519.1, MH837518.1, MH837517.1, MH837516.1, MH837515.1, MH837514.1, MH837513.1, MH837512.1). In the NPC samples, the Med+ subtype was found in 23 cases, the China 1 subtype in 7 cases, the B95-8 strain in 3 cases, and the *BNLF1* subtype could not be determined in 3 cases. In the AITL samples, five cases showed the B95-8 strain, five cases were the Med- subtype, three cases were the Med+ subtype, one case was the China 1 subtype, and one case was the Alaskan subtype. In two cases, it was not possible to determine EBV subtypes based on the *BNLF-1* gene sequence.

In the tumor samples data from GenBank, the Med+ subtype was significantly more represented in the NPC samples compared with AITL (OR = 7.9, 95% CI 1.8–50.9; $p = 0.008$), while the Med- subtype was significantly more represented in the AITL samples compared with NPC (OR = 14.8, 95% CI 1.6–73.0; $p = 0.008$).

3.2 EBNA-3C and EBNA-1 Genes

Out of 91 tumor tissue samples, 84 *EBNA-3C* sequences obtained by Sanger sequencing were of adequate quality for further analysis. This included 46 sequences from the cHL samples (from the previously published study) [17], 11 sequences from the AITL samples, and 27 sequences from the NPC samples.

All samples were classified as EBV-1, with 79 *EBNA-1* gene sequences available for further analysis (42 from cHL, 11 from AITL, and 26 from NPC).

Of 36 additional EBV sequences in NPC found in GenBank for EBER algorithm validation, 30 were EBV-1 genotype, 1 was EBV-2 genotype, and for 5 cases, genotype could not be determined based on EBNA3C classification. All 17 NPC samples from GenBank were determined as genotype EBV-1.

To determine the EBV subtype (prototypes: P-ala and P-thr, or variants: V-val, V-leu, V-pro, and V-ala), the variability of the sequences obtained was analyzed using previously established algorithms [13,37–43]. Two subtypes, P-thr and P-ala prototypes, were detected in all three tumor types (**Supplementary Table 6**). There was no statistically significant association of age and sex with EBV subtypes ($p > 0.05$).

However, determining the subvariants (P-thr′ or P-thr″) was not possible for the P-thr prototype, considering that the defining nucleotide position had not been sequenced. In one cHL sample with the EBV prototype P-thr, an additional Ala→Thr substitution was found at position 525. For the P-ala prototype, P-ala′ was the most common subvariant in cHL and AITL samples, except for the P-ala′ subvariant in one cHL sample. Two subvariants of the P-ala prototype were found in the NPC samples: P-ala′ and P-ala-v1 subvariants. The P-ala′ subvariant was significantly more frequent in the NPC samples from women than from men (OR = 8.33, 95% CI 1.04–102.44; $p = 0.030$). The V-val, V-leu, V-pro, and V-ala variants were not found in the

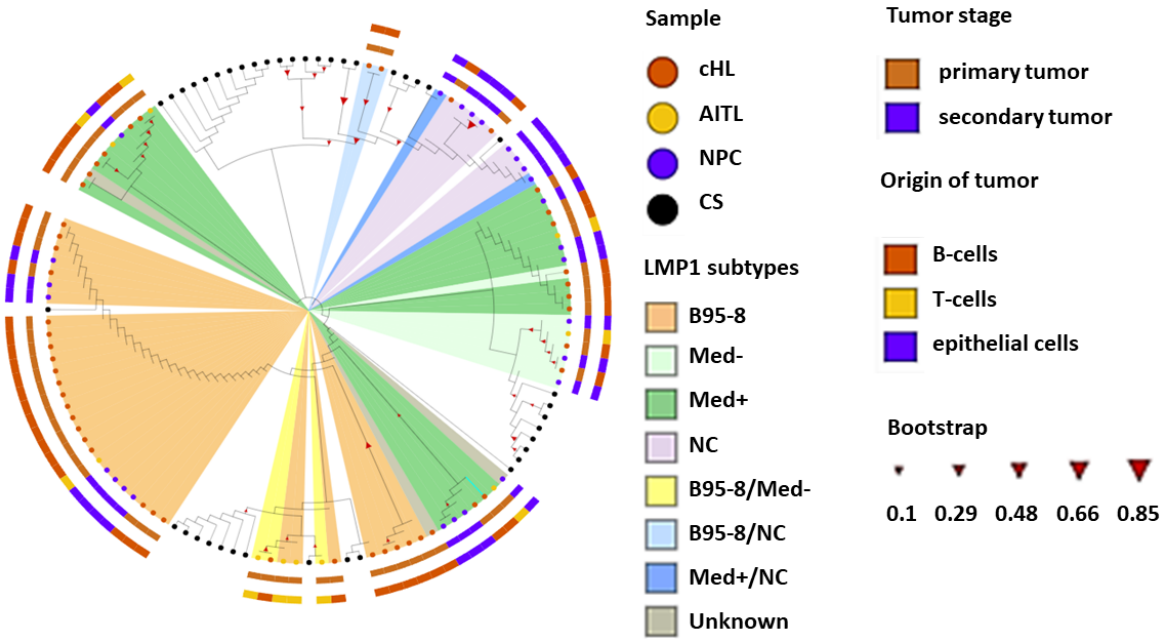

Fig. 1. A phylogenetic tree of the *BNLF-1* gene was constructed using the maximum likelihood method with the JC + G substitution model for 1000 pseudoreplications. Bootstrap values between 0.1 and 0.85 are shown at the nodes as red triangles, with size proportional to the strength of branch support. The three diagnoses, cHL, AITL, and NPC, are represented by red, yellow, and blue-colored circles, respectively, and black circles represent the control sequences (CS). The two-colored bar shows the developmental stage of the tumor, while the three-colored bar represents the tumor origin. The color range shows the clades of different subtypes and co-infections. No subtype was assigned to the control sequences. Data for *BNLF-1* sequences in cHL samples were used from the previously published paper [17]. cHL, classical Hodgkin's lymphoma; AITL, angioimmunoblastic T- cell lymphoma; NPC, nasopharyngeal carcinoma; CS, control sequences; NC, North Carolina subtype.

analyzed samples. The overall distribution of subtypes by tumor type is shown in **Supplementary Table 7**.

The P-ala subtype, subvariant P-ala′, was significantly more frequent than the P-thr subtype in NPC samples compared to the other analyzed tumor types (P-ala: OR = 9.89, 95% CI 2.96–37.05; $p < 0.001$, P-ala′: OR = 5.48, 95% CI 1.59–20.83; $p = 0.007$). The P-thr subtype was more frequent in primary tumors/tumors of lymphocytic origin than in secondary tumors/tumors of epithelial origin (OR = 9.89, 95% CI 2.96–37.05; $p < 0.001$). Additionally, the P-ala′ subvariant was significantly more frequently present in the secondary tumor type compared to P-thr and other P-ala subvariants (OR = 5.56, 95% CI 1.59–20.03; $p = 0.007$).

The final length of the *EBNA-1* sequences was 214 bp, corresponding to nucleotide positions 109377–109590 of the reference genome B95-8. In phylogenetic analyses, the JC + G model was the most probable substitution model for the combined data set. The phylogenetic tree shows that the sequences were not grouped by diagnosis, tumor stage, or tumor origin. The grouping of sequences into two clades is visible and depends on EBV subtypes. The highest branching support was 65% (Fig. 2).

Of 36 NPC samples found in GenBank for EBER algorithm validation, EBV subtype based on EBNA1 sequence was variant V-val in 31 cases, prototype P-ala in 3 cases, and determination could not be performed in 2 cases. Out of 17 AITL cases, 15 were P-thr prototypes, 1 was P-ala prototype, and 1 was V-val variant.

In the tumor samples from GenBank, the variant V-val was significantly more represented in NPC compared to AITL samples (OR = 85.4, 95% CI 9.9–415.2; $p < 0.001$), while P-thr prototype was significantly more represented in AITL compared to NPC samples (OR = 157.6, 95% CI 16.5–785.5; $p < 0.001$).

3.3 The Association Between BNLF-1 and EBNA-1 Subtypes

The NC subtype was significantly more frequently observed with the P-ala subtype compared to the P-thr subtype (OR = 7.14, 95% CI 1.07–100.51; $p = 0.024$). The combination of the B95-8 and P-thr subtypes was significantly more frequent in cHL samples compared to the other analyzed tumor types (OR = 5.47, 95% CI 1.50–25.54; $p = 0.009$), while the combination of the NC and P-ala subtypes was more frequently found in the NPC samples (OR = 13.59, 95% CI 1.25–701.54; $p = 0.019$) compared to cHL and AITL. The combination of Med+ and P-thr subtypes was significantly less frequent in NPC samples than in the other tumor types (OR = 0.15, 95% CI 0.02–0.77; $p = 0.016$). The association between the *BNLF-1* and *EBNA-1* subtypes and their distribution across tumor types is shown in Table 3.

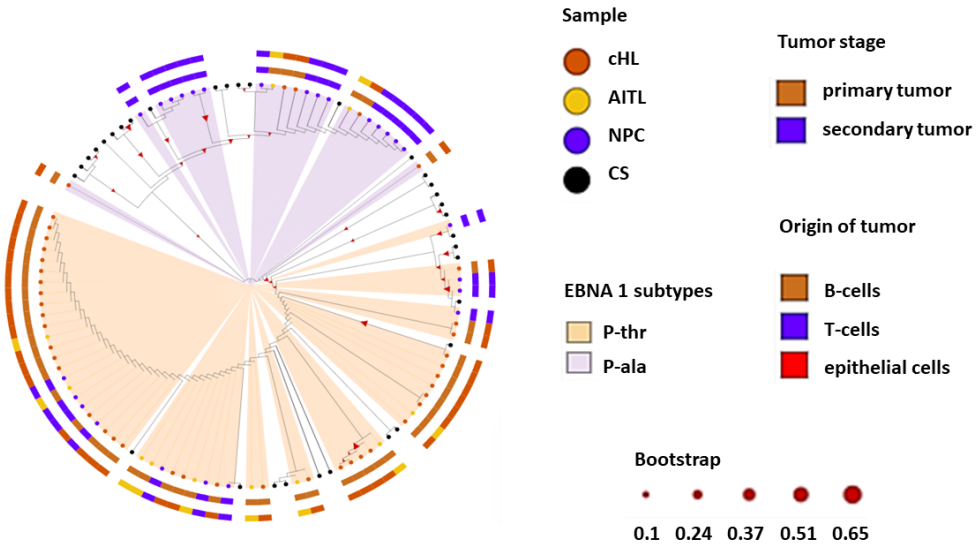

Fig. 2. Phylogenetic tree of the *EBNA-1* gene constructed using the maximum likelihood method with the JC + G substitution model for 1000 pseudoreplications. Bootstrap values between 0.1 and 0.65 are shown at the nodes as red triangles, with size proportional to the strength of branch support. The three diagnoses, cHL, AITL, and NPC, are represented by red, yellow, and blue-colored circles, respectively, and black circles represent the CS. The two-colored bar shows the developmental stage of the tumor, while the three-colored bar represents the tumor origin. Clades of two subtypes are shown with two colors. No subtype was assigned to the control sequences.

In the tumor samples from GenBank, the combination of the Med+/V-val subtypes was significantly more frequent in the NPC samples compared to the AITL samples (OR = 21.2, 95% CI 2.7–97.7; $p = 0.002$), while the combinations of the B95-8/P-thr and Med-/P-thr subtypes were more frequently found in AITL samples (OR = 16.2, 95% CI 1.7–80.0; $p = 0.004$) compared to NPC.

3.4 EBER Gene

Out of 91 tumor tissue samples, 81 *EBER* sequences were of adequate quality for further analyses (47 from cHL, 11 from AITL, and 23 from NPC samples). Recurrent mutations were observed at specific positions within the analyzed *EBER* promoter sequence. Thus, a new algorithm for EBV classification was designed based on these changes (Table 1), following a similar principle used in EBV classification algorithms based on the *BNLF-1* and *EBNA-1* genes.

The B95-8 subtype and the newly defined Cro2 subtype were detected in all tumor types, while the Cro1 subtype was detected exclusively in the NPC samples. Four subvariants of the Cro2 subtype were detected in the cHL and AITL samples, in addition to the B95-8 subtype (**Supplementary Table 8**). All three subtypes were detected in the NPC samples, which included the Cro1 subtype with both subvariants, and the Cro2 subtype with three subvariants (**Supplementary Table 8**); no statistically significant association was observed for age and sex with EBV subtypes ($p > 0.05$). The overall distribution of subtypes by tumor type is shown in **Supplementary Table 9**.

The Cro2 subtype was significantly more frequent in cHL and AITL compared to NPC (OR = 6.88, 95% CI 2.12–25.45; $p < 0.001$), while the B95-8 subtype was more frequent in NPC (OR = 4.51, 95% CI 1.31–16.28; $p = 0.015$). Accordingly, the Cro2 subtype was more frequent in the primary tumor type, *i.e.*, in a tumor of lymphocytic origin compared to the other *EBER* subtypes (OR = 6.88, 95% CI 2.12–25.45, $p < 0.001$), and the B95-8 subtype was more commonly found in the secondary tumor type, *i.e.*, in tumors originating from epithelial cells (OR = 4.51, 95% CI 1.31–16.28; $p = 0.015$).

Using GenBank data, the newly proposed algorithm identified four distinct EBER sequences in NPC. Out of 36 NPC samples 28 samples had the sequence of B95-8 strain with a substitution G>A on position 6866, three samples had the same sequence as strain B95-8, two samples were determined as Cro2 subtype (subvariant Cro2^{i}), one sample had duplication of sequence TCCATTAT (positions 6826–6833), while the EBER sequence could not be analyzed for one sample. The majority of AITL cases showed Cro2 subtype (12 in total, 11 subvariant Cro2^{i} and one subvariant Cro2^{iii}). In addition, two cases were determined as the B95-8 strain; in one case, substitutions G>T at position 6884 and C>T at position 6885 were observed in the B95-8 sequence, and in the other case, the sequence could not be analyzed.

The Cro2 subtype was significantly more frequent in the AITL samples compared to the NPC samples ($p < 0.001$), while the B95-8 sequence, with a substitution G>A on position 6866, was exclusively present in the NPC samples.

When the newly proposed algorithm was applied to the high-risk EBER deletion sequences published by Hui *et al.* [15], 54 sequences showed no newly proposed Cro subtypes; however, one sequence from the B95-8 strain

Table 3. The association of *BNLF-1* and *EBNA-1* subtypes and distribution of subtype combinations in the analyzed tumors.

Polymorphisms	Tumor samples		
	cHL	AITL	NPC
B95-8/P-thr	18 (46.5%)	2 (20%)	3 (14%)
B95-8/P-ala	2 (5%)	1 (10%)	3 (14%)
Med+/P-thr	11 (28.5%)	3 (30%)	1 (4.5%)
Med+/P-ala	1 (2.5%)	1 (10%)	4 (18%)
Med-/P-thr	3 (7.5%)	1 (10%)	2 (9%)
Med-/P-ala	-	-	1 (4.5%)
NC/P-thr	-	-	2 (9%)
NC/P-ala	1 (2.5%)	-	4 (18%)
B95-8/Med-/P-thr	1 (2.5%)	2 (20%)	-
B95-8/Med-/P-ala	-	-	-
B95-8/NC/P-thr	2 (5%)	-	-
B95-8/NC/P-ala	-	-	-
Med+/NC/P-thr	-	-	-
Med+/NC/P-ala	-	-	2 (9%)

B95-8, prototype EBV strain; Med+, Mediterranean subtype with deletion; Med-, Mediterranean subtype without deletion; NC, North Carolina subtype; P-thr, prototype of the EBV containing the amino acid Thr at position 487 of the EBNA-1 protein; P-ala, prototype containing the amino acid Ala at position 487.

showed a G>A substitution at position 6866. Sequence MH590551.1 (isolate HKNPC40) was determined as subtype Cro2^{i}, sequence MH590535.1 (isolate HKNPC24) could not be determined with the newly proposed algorithm; however, in addition to substitutions G>A at position 6866, a substitution G>T was also observed at position 6915, while the sequence MH590512.1 (isolate HKNPC1) was determined as the Cro1 subtype with an additional substitution G>A at position 6866.

3.5 The Association Between EBER, BNLF-1, and EBNA-1 Subtypes

The EBER Cro2^{i} subtype was more often present with the B95-8 or Med- subtypes compared to the Med+ or NC (OR = 6.21, 95% CI 1.91–22.18; $p < 0.001$), while the EBER B95-8 subtype was more frequently found with the NC subtype compared to the other *BNLF-1* subtypes (OR = 20.02, 95% CI 1.97–1024.12; $p = 0.007$). The Cro2^{iii} subtype was more often present with the B95-8 or Med+ LMP-1 subtypes compared to the Med- or NC subtypes (12.57, 95% CI 1.16–685.32; $p = 0.015$), while the Cro2^{iv} subtype was more commonly found with the Med+ subtype compared to other *BNLF-1* subtypes (OR = 19.82, 95% CI 1.09–1394.23; $p = 0.024$). When the individual tumor type was analyzed, the association between the Cro2^{i} subtype and the B95-8 or Med subtypes was observed in NPC (OR = 10.63, 95% CI 1.01–610.74; $p = 0.022$).

By studying the association with the *EBNA-1* subtypes, the Cro2^{i} *EBER* subtype was observed more often with the P-thr subtype compared to the P-ala subtype (OR = 50.32, 95% CI 6.72–2261.71; $p < 0.001$), while the B95-8 *EBER* subtype was more commonly found with the P-ala subtype (OR = 50.11, 95% CI 10.17–1000.26; $p < 0.001$).

When individual tumor types were analyzed, a more frequent association between the Cro2^{i} subtype and the P-thr subtype was found in cHL (OR = 16.14, 95% CI 1.35–893.58; $p = 0.015$) and NPC (OR = 12.92, 95% CI 1.04–762.14; $p = 0.024$) compared with the P-ala subtype. A more frequent association between the B95-8 *EBER* subtype and the P-ala subtype was also found in all analyzed tumor types, cHL (OR = 26.79, 95% CI 1.04–2017.52; $p = 0.024$), AITL (OR = 19.98, 95% CI 0.87–1691.98; $p = 0.022$), and NPC (OR = 17.52, 95% CI 1.69–945.93; $p = 0.007$) compared to the P-thr subtype.

By comparing the *BNLF-1*, *EBNA-1,* and *EBER* subtypes, the combination of the B95-8, P-thr, and Cro2^{i} subtypes was significantly more frequent in cHL compared to AITL and NPC (OR = 10.32, 95% CI 2.07–102.29; $p < 0.001$), and the combination of P-ala and B95-8 *EBER* subtypes was significantly more frequent in NPC compared to cHL and AITL (OR = 7.68, 95% CI 1.80–39.90; $p = 0.007$). The most common subtype combinations based on the *BNLF-1*, *EBNA-1*, and *EBER* gene polymorphisms are shown in Table 4.

When the GenBank EBV data were analyzed, a more frequent association between the Cro2^{i} subtype and the P-thr EBNA1 subtype and B-95-8 LMP-1 sequence was found in AITL (OR = 12.4, 95% CI 1.2–63.6; $p = 0.015$) compared to NPC. A more frequent association between the B95-8 EBER sequence and the Med+ LMP-1 subtype and V-val EBNA1 variant was found in NPC (OR = 13.3, 95% CI 1.7–60.6; $p = 0.029$) compared with AITL.

4. Discussion

The results of this study showed characteristic patterns of EBV genotype/variant distribution based on the sequences of the *EBNA-3C*, *EBNA-1,* and *BNLF-1* genes, as well as the *EBER* promoter, in three models of malignant diseases (AITL, NPC, and cHL) in a cohort of patients receiving care in a single clinical center from southeastern Europe. Based on *EBNA-3C* gene sequencing, all enrolled patients were infected with EBV genotype 1. The *BNLF-1* variant analysis demonstrated an exclusive presence of only three variants (Mediterranean subtype with or without 30 bp deletion and the wild-type strain B95-8) in AITL, with the addition of the North Carolina variant in NPC patients. Interestingly, the frequencies of EBV prototypes P-ala and P-thr, and the associated variants based on the *EBNA-1* gene sequence, were significantly associated with the type of malignant disease (increased in NPC) and the tumor type (increased in secondary tumor types). Our newly designed algorithm for the classification of EBV variants based on specific positions within the *EBER* promoter also identified a pattern of variant representation, with the Cro1 subtype be-

Table 4. Association of *BNLF-1*, *EBNA-1*, and *EBER* subtypes, and distribution of subtype combinations in the analyzed tumors.

Combination of *BNLF-1*, *EBNA-1*, and *EBER* subtypes	Tumor samples		
	cHL	AITL	NPC
B95-8, P-thr, Cro2^{i}	16 (43.2%)	1 (10.0%)	1 (5.0%)
Med+, P-thr, Cro2^{i}	7 (18.9%)	1 (10.0%)	0 (0.0%)
P-ala, EBER B95-8	2 (5.1%)	2 (18.2%)	9 (40.9%)
Med+, Cro2^{i}	8 (19.1%)	1 (10.0%)	0 (0.0%)

B95-8, prototype EBV strain; Med+, Mediterranean subtype with deletion; P-thr, prototype that contains the amino acid Thr at position 487 of the EBNA-1 protein; P-ala, prototype that includes the amino acid Ala at position 487 of the EBNA-1 protein; Cro2^{i}, subtype with substitutions G→A (position of the genome 6884), T→G (6886) and A→G (6911).

ing more common in the NPC samples. To our knowledge, this is the first study to analyze the molecular diversity of EBV in AITL and the first to classify EBV based on *EBER* promoter sequences.

The classification of EBV into genotypes 1 and 2 (or A and B) is usually the first step in studies analyzing the molecular diversity of this virus. However, it should also be emphasized that the two genotypes exhibit different biological features *in vitro* and *in vivo*. For example, *in vitro* studies have shown that EBV genotype 1, such as the wild-type B95-8, has a greater capacity to transform resting human B-cells into lymphoblastoid cell lines than genotype 2. The ability of EBV genotypes to determine cellular growth phenotype in this experimental model was associated with molecular features of EBNA-2 protein [9]. However, more recent studies in humanized mouse models have confirmed the ability of EBV type 2 to establish type III latency in T- and B-cells and to induce a malignant disease similar to diffuse large B-cell lymphoma *in vivo*, with features also comparable to those observed in human models [8]. In addition, EBV genotype 2 is characterized by distinct viral tropism for T cells [45].

A large number of studies analyzing the molecular epidemiology of EBV in various models of malignant and non-malignant diseases have shown a global predominance of EBV genotype 1. Furthermore, region-specific patterns with a high frequency of genotype 2 in selected patient populations (Burkitt's lymphoma in Africa and HL in China and Mexico) have been documented [32]. Our previous studies have shown the presence of genotype 1 in a cohort of 50 patients with histologically verified EBV-positive cHL from Croatia, as well as in a cohort of pediatric patients with EBV-associated infectious mononucleosis [17,46]. The region-specific predominance of EBV type 1 has also been confirmed in the present study, with this genotype exclusively present in the NPC and AITL cohorts. The NPC findings observed in our study are in concordance with the data collected by Banko *et al*. [47], who observed the presence of genotype 1 in 90.7% of samples from a cohort of 30 patients with infectious mononucleosis, six transplanted patients, 16 NPC patients, and one patient with cHL from Serbia.

Since LMP-1 is an essential oncogenic protein of EBV, the distribution of the viral *BNLF-1* variants has been extensively investigated in patients with malignant diseases of different cellular origin (cHL, NHL, NPC, T cell lymphomas, and Burkitt's lymphoma) as well as in patients with EBV-associated infectious mononucleosis and healthy subjects [8,47–53]. Recently, we have shown a very narrow repertoire of EBV *BNLF-1* variant distribution in patients with cHL from Croatia, with the predominance of the wild-type strain B95-8 and the Mediterranean subtype with a 30 bp deletion, while NC, the Mediterranean subtype without a 30 bp deletion, and co-infections were very rare [17]. Contrary to these findings, *BNLF-1* variant distribution in Croatian pediatric patients diagnosed with infectious mononucleosis was more diverse, including the China 1 variant, and revealed a very high percentage of *BNLF-1* variant recombinants (16.4% of patients) [46]. The *BNLF-1* variant analysis in this study showed a narrow repertoire of variants in AITL, including only the Mediterranean subtype (with or without the 30 bp deletion) and the wild-type strain B95-8. Since this is the first experimental evidence of *BNLF-1* variant distribution in AITL, it is not possible to determine whether this distribution reflects regional specificity or a more general skewing of gene diversity associated with this particular disease.

The possible clinical importance of *BNLF-1* gene variability has been extensively investigated in NPC. A recent systematic review and meta-analysis by Banko *et al*. [12] provided a proof of concept for the association between 30 bp deletion (OR = 3.53, 95% CI 1.48–8.43) and XhoI loss (OR = 14.17, 95% CI 4.99–40.20) and showed a skewing of the *BNLF-1* variant repertoire in this model by demonstrating a very rare occurrence of B95-8, China 1 and NC in NPC patients [12]. Contrary to these findings, our results suggest that in addition to the Mediterranean variant (mainly with a 30 bp deletion) and the wild-type B95-8, NC, and recombinants between the Mediterranean with deletion and NC are also present in Croatian NPC patients. These results once again emphasize the need for careful comparisons and interpretations of studies from different geographic regions that clearly exhibit specific variant distributions.

Due to the importance of the EBNA-1 protein in the process of infection, genome replication, transcription, maintenance, as well as in the transformation of target cells, several studies investigated the variability of the *EBNA-1* gene based on the amino acid at position 487 in malignant diseases associated with EBV.

The largest study of NPC patients to date was reported by Thuan *et al*. [13], which showed a clear predominance of the V-val subtype in about 80% of biopsy samples from Vietnamese patients. In addition, the authors identified a novel pattern of the V-val subtype with amino acid changes at positions 492, 528, 529, 553, 585, and 588 compared with the wild-type B95-8, suggesting a specific variation pattern of the *EBNA-1* gene in Vietnam [13]. These findings support the data from Zhang *et al*. [54], who demonstrated the exclusive presence of the V-val subtype in NPC patients obtained in endemic and non-endemic areas of China. Contrary to these findings, the *EBNA-1* gene variability analysis reported in our study focusing on NPC patients from a non-endemic European country revealed the complete absence of V-val subtype and the predominance of P-ala (subvariants P-ala′ and P-ala-v1) in comparison with the P-thr prototype. In addition, our findings showed an exceptional frequency for the P-thr subtype in cHL and AITL patients.

Recently, Vafapour *et al*. [55] compared the variability of the *EBNA-1* gene in 40 patients with chronic lymphocytic leukemia (CLL) and 21 healthy subjects, showing the highest frequency of the P-ala variant in both groups, but also suggesting that the P-thr and V-val variants might be associated with the pathogenesis of CLL.

Several recent studies on NPC from endemic populations have clearly demonstrated the association between variation in the EBV genome and functional consequences, as well as the pathogenesis of this malignant disease, including variants within the *BALF-2* [56] and *EBER-2* genes [15,57] (for review, see Young [58]). Since *EBERs* significantly contribute to the efficient *in vitro* transformation of B cells by enhancing the growth potential of transformed lymphocytes, polymorphisms in these genes are likely candidates for further analysis of their clinical significance in NPC and other diseases [59]. Here, we demonstrated the exclusive presence of the Cro1 subtype in NPC samples using a newly developed algorithm to characterize *EBER-1* subtypes in our cohort of patients, thereby providing an important candidate for further functional analyses that could lead to the possible identification of clinically relevant EBV variants in NPC from non-endemic areas. In addition, to validate our algorithm, we used the GenBank database to identify EBV sequences from tumors of the same type as those in our study. We found 36 available sequences from NPC patients (samples mostly from China) and 17 from AITL patients (samples mostly from France). After applying our algorithm to those sequences, we observed a very rare occurrence of the Cro2 subtype in NPC samples and a very frequent occurrence in AITL samples. Moreover, we applied our algorithm to known high-risk EBER deletion sequences [15] and found no correlation between high-risk deletion and EBER Cro subtypes.

Single-nucleotide polymorphisms (SNPs) in promoter regions, such as those described in this paper, can significantly influence gene regulation by altering transcription factor binding sites and, thereby, increasing or decreasing transcription factor affinity for those sites [60]. SNPs also influence promoter strength, with some positions in the promoter sequence exerting greater influence than others [61]. Recently, Kim *et al*. [62] showed that an SNP in the *BART* promoter region enhanced gene activity and may, therefore, contribute to the development of EBV-associated epithelial malignancies. We hypothesize that SNPs in the *EBER* promoter could have a similar effect. Although the exact functional effects of the SNPs described in this study remain unknown, these findings provide a foundation for future research into the mechanistic roles of SNPs. Additionally, further research could investigate the potential relevance of these effects for vaccine development and other therapeutic strategies.

It is also important to highlight the limitations of our study. Analyses of EBV genetic diversity in malignant diseases reported in the literature refer to various populations and geographic regions and, in some disease models, depend on region-specific epidemiology. Therefore, our results describing characteristic patterns of EBV diversity in the three selected malignant diseases are most relevant to Caucasian patients from the southeastern part of Europe and may not necessarily translate to findings in other geographic areas. In addition, viral diversity studies in any disease model can be influenced by the heterogeneity in the population regarding the their demographic and genetic backgrounds, as well as social and behavioral patterns.

Furthermore, translating data on EBV molecular diversity into the clinical context and considering the possible inclusion of EBV genotyping in the diagnostic work-up of patients will require prospective longitudinal studies that link EBV virological background to clinical outcomes. The evaluation of the individual EBV genotypes described in this study as possible predictors of favorable or less favorable outcomes across different models of malignant disease at the population level is a promising area of future research. Notably, collaborative studies across different geographic regions, for example, Europe, Asia, and Africa, might contribute to more accurate identification of pathogenic EBV variants in various disease models that may require adapted treatment protocols.

The majority of the literature on EBV diversity is based on Sanger sequencing, which, due to the inherent limitations associated with providing a consensus sequence, does not allow accurate determination of potential recombinants or co-infections with multiple viral variants. The use of high-throughput sequencing (HTS) by short-read, for example, next-generation sequencing on the Illumina platform, or long-read—such as Nanopore-based direct DNA sequencing—allows a more detailed analysis of the viral

population architecture and is likely to contribute to a more accurate analysis of EBV recombinants and co-infections in various models of malignant diseases. In our recent study of the *BNLF-1* genotype distribution in pediatric patients with EBV-associated infectious mononucleosis by Oxford Nanopore Technology, sequences were classified as recombinants when polymorphisms associated with more than one *BNLF-1* variant were found at distinct nucleotide positions [63]. In contrast, sequences were considered co-infections when polymorphisms from multiple variants occurred at the same position with approximately equal frequencies. The use of different HTS technologies to more in-depth analysis of EBV diversity in malignant diseases might also yield a more accurate map of viral co-infections or recombinants across various geographic regions.

5. Conclusion

In conclusion, this study has shown a characteristic pattern of EBV diversity based on sequences of the *EBNA-3C*, *EBNA-1*, and *BNLF-1* genes, as well as the *EBER* promoter, across AITL, NPC, and cHL cohorts from southeastern Europe. To our knowledge, this is the first study of EBV molecular diversity on AITL and the first to provide a novel EBV classification algorithm based on *EBER* promoter sequences. These data may provide an important contribution to the development of therapeutic or prophylactic vaccines, as well as novel treatment options for EBV-associated diseases.

Availability of Data and Materials

We have deposited the sequences in GeneBank (https://www.ncbi.nlm.nih.gov/nuccore/): PV020184-PV020264 (*EBER*), PV020265-PV020343 (*EBNA-1*), PV020344-PV020427 (*BNLF-1*), OM460003-OM460060 (*EBNA-3C*), OR340986-OR341017 (*EBNA-3C*). The other datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

Conceptualization, PK and SZ-L; methodology, PK, SG, SH, VB, KB, PG, RG, MR and PS; software, VB, RG, KB, MR and PS; validation, KB, MR and PS; formal analysis, VB, RG, SH, KB, PG and PK; investigation, VB, RG, PK, SH, PG, MR, PS and SG, resources, PK, SG and SZ-L; data curation, SG and PK; writing—original draft preparation, VB, PK and SZ-L; writing—review and editing, all authors; visualization, PK, SG and VB; supervision, SZ-L; project administration, SG and PK; funding acquisition, SZ-L. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The study was approved by the Ethics committee of the Merkur University Hospital, Zagreb, Croatia, on 25th September 2019 (UR. BR. 03/1-8793). For this study, informed consent was obtained from the patients or their families/legal representatives. The study was carried out in accordance with the guidelines of the Declaration of Helsinki.

Acknowledgment

Not applicable.

Funding

This work was supported by the Croatian Science Foundation Grant [grant number IP-2020-02-8961, "Molecular diversity of Epstein–Barr virus", to Prof. Snjezana Zidovec Lepej].

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/FBS46442.

References

[1] Kieff E, Rickinson AB. Epstein-Barr virus and its replication. In Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MM, Roizman B, *et al*. (eds.) Fields virology (pp. 2603–2654). 5th edn. Lippincott Williams & Wilkins: Philadelphia. 2007.

[2] Henle G, Henle W, Diehl V. Relation of Burkitt's tumor-associated herpes-ytpe virus to infectious mononucleosis. Proceedings of the National Academy of Sciences of the United States of America. 1968; 59: 94–101. https://doi.org/10.1073/pnas.59.1.94.

[3] Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. Frontiers in Oncology. 2019; 9: 713. https://doi.org/10.3389/fonc.2019.00713.

[4] Bu GL, Xie C, Kang YF, Zeng MS, Sun C. How EBV Infects: The Tropism and Underlying Molecular Mechanism for Viral Infection. Viruses. 2022; 14: 2372. https://doi.org/10.3390/v14112372.

[5] Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, *et al*. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature. 1984; 310: 207–211. https://doi.org/10.1038/310207a0.

[6] 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. Journal of Virology. 2015; 89: 5222–5237. https://doi.org/10.1128/JVI.03614-14.

[7] Rowe M, Young LS, Cadwallader K, Petti L, Kieff E, Rickinson AB. Distinction between Epstein-Barr virus type A (EBNA 2A) and type B (EBNA 2B) isolates extends to the EBNA 3 family of nuclear proteins. Journal of Virology. 1989; 63: 1031–1039. https://doi.org/10.1128/JVI.63.3.1031-1039.1989.

[8] Simbiri KO, Smith NA, Otieno R, Wohlford EEM, Daud II, Odada SP, *et al*. Epstein-Barr virus genetic variation in lymphoblastoid cell lines derived from Kenyan pediatric population. PloS One. 2015; 10: e0125420. https://doi.org/10.1371/journal.pone.0125420.

[9] Rickinson AB, Young LS, Rowe M. Influence of the Epstein-Barr virus nuclear antigen EBNA 2 on the growth phenotype of virus-transformed B cells. Journal of Virology. 1987; 61: 1310–1317. https://doi.org/10.1128/JVI.61.5.1310-1317.1987.

[10] Edwards RH, Seillier-Moiseiwitsch F, Raab-Traub N. Signature amino acid changes in latent membrane protein 1 distinguish Epstein-Barr virus strains. Virology. 1999; 261: 79–95. https://doi.org/10.1006/viro.1999.9855.

[11] Hulse M, Johnson SM, Boyle S, Caruso LB, Tempera I. Epstein-Barr Virus-Encoded Latent Membrane Protein 1 and B-Cell Growth Transformation Induce Lipogenesis through Fatty Acid Synthase. Journal of Virology. 2021; 95: e01857–20. https://doi.org/10.1128/JVI.01857-20.

[12] Banko A, Miljanovic D, Lazarevic I, Cirkovic A. A Systematic Review of Epstein-Barr Virus Latent Membrane Protein 1 (LMP1) Gene Variants in Nasopharyngeal Carcinoma. Pathogens (Basel, Switzerland). 2021; 10: 1057. https://doi.org/10.3390/pathogens10081057.

[13] Thuan LD, Kha ND, Minh NT, Thuy L. Novel Patterns of the Epstein-Barr Nuclear Antigen (EBNA-1) V-Val Subtype in EBV-associated Nasopharyngeal Carcinoma from Vietnam. Balkan Journal of Medical Genetics: BJMG. 2019; 22: 61–68. https://doi.org/10.2478/bjmg-2019-0011.

[14] Sun L, Zhao Z, Liu S, Liu X, Sun Z, Luo B. Sequence Variation Analysis of Epstein-Barr Virus Nuclear Antigen 1 Gene in the Virus Associated Lymphomas of Northern China. PloS One. 2015; 10: e0140529. https://doi.org/10.1371/journal.pone.0140529.

[15] Hui KF, Chan TF, Yang W, Shen JJ, Lam KP, Kwok H, *et al*. High risk Epstein-Barr virus variants characterized by distinct polymorphisms in the EBER locus are strongly associated with nasopharyngeal carcinoma. International Journal of Cancer. 2019; 144: 3031–3042. https://doi.org/10.1002/ijc.32049.

[16] Zanella L, Riquelme I, Buchegger K, Abanto M, Ili C, Brebi P. A reliable Epstein-Barr Virus classification based on phylogenomic and population analyses. Scientific Reports. 2019; 9: 9829. https://doi.org/10.1038/s41598-019-45986-3.

[17] Begić V, Korać P, Gašparov S, Rozman M, Simicic P, Zidovec-Lepej S. Molecular Characterisation of Epstein-Barr Virus in Classical Hodgkin Lymphoma. International Journal of Molecular Sciences. 2022; 23: 15635. https://doi.org/10.3390/ijms232415635.

[18] Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, *et al*. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016; 127: 2375–2390. https://doi.org/10.1182/blood-2016-01-643569.

[19] de Leval L, Rickman DS, Thielen C, Reynies AD, Huang YL, Delsol G, *et al*. The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. Blood. 2007; 109: 4952–4963. https://doi.org/10.1182/blood-2006-10-055145.

[20] Vose J, Armitage J, Weisenburger D, International T-Cell Lymphoma Project. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. Journal of Clinical Oncology. 2008; 26: 4124–4130. https://doi.org/10.1200/JCO.2008.16.4558.

[21] de Leval L, Gisselbrecht C, Gaulard P. Advances in the understanding and management of angioimmunoblastic T-cell lymphoma. British Journal of Haematology. 2010; 148: 673–689. https://doi.org/10.1111/j.1365-2141.2009.08003.x.

[22] Kim TY, Min GJ, Jeon YW, Park SS, Park S, Shin SH, *et al*. Impact of Epstein-Barr Virus on Peripheral T-Cell Lymphoma Not Otherwise Specified and Angioimmunoblastic T-Cell Lymphoma. Frontiers in Oncology. 2022; 11: 797028. https://doi.org/10.3389/fonc.2021.797028.

[23] Bahri R, Boyer F, Halabi MA, Chaunavel A, Feuillard J, Jaccard A, *et al*. Epstein-Barr Virus (EBV) Is Mostly Latent and Clonal in Angioimmunoblastic T Cell Lymphoma (AITL). Cancers. 2022; 14: 2899. https://doi.org/10.3390/cancers14122899.

[24] Lachenal F, Berger F, Ghesquières H, Biron P, Hot A, Callet-Bauchu E, *et al*. Angioimmunoblastic T-cell lymphoma: clinical and laboratory features at diagnosis in 77 patients. Medicine. 2007; 86: 282–292. https://doi.org/10.1097/MD.0b013e3181573059.

[25] Mourad N, Mounier N, Brière J, Raffoux E, Delmer A, Feller A, *et al*. Clinical, biologic, and pathologic features in 157 patients with angioimmunoblastic T-cell lymphoma treated within the Groupe d'Etude des Lymphomes de l'Adulte (GELA) trials. Blood. 2008; 111: 4463–4470. https://doi.org/10.1182/blood-2007-08-105759.

[26] Zettl A, Lee SS, Rüdiger T, Starostik P, Marino M, Kirchner T, *et al*. Epstein-Barr virus-associated B-cell lymphoproliferative disorders in angloimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified. American Journal of Clinical Pathology. 2002; 117: 368–379. https://doi.org/10.1309/6UTX-GVC0-12ND-JJEU.

[27] Willenbrock K, Bräuninger A, Hansmann ML. Frequent occurrence of B-cell lymphomas in angioimmunoblastic T-cell lymphoma and proliferation of Epstein-Barr virus-infected cells in early cases. British Journal of Haematology. 2007; 138: 733–739. https://doi.org/10.1111/j.1365-2141.2007.06725.x.

[28] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: a Cancer Journal for Clinicians. 2011; 61: 69–90. https://doi.org/10.3322/caac.20107.

[29] Okekpa SI, S M N Mydin RB, Mangantig E, Azmi NSA, Zahari SNS, Kaur G, *et al*. Nasopharyngeal Carcinoma (NPC) Risk Factors: A Systematic Review and Meta-Analysis of the Association with Lifestyle, Diets, Socioeconomic and Sociodemographic in Asian Region. Asian Pacific Journal of Cancer Prevention: APJCP. 2019; 20: 3505–3514. https://doi.org/10.31557/APJCP.2019.20.11.3505.

[30] Li X, Fasano R, Wang E, Yao KT, Marincola FM. HLA associations with nasopharyngeal carcinoma. Current Molecular Medicine. 2009; 9: 751–765. https://doi.org/10.2174/156652409788970698.

[31] Sun L, Wang Y, Shi J, Zhu W, Wang X. Association of Plasma Epstein-Barr Virus LMP1 and EBER1 with Circulating Tumor Cells and the Metastasis of Nasopharyngeal Carcinoma. Pathology Oncology Research: POR. 2020; 26: 1893–1901. https://doi.org/10.1007/s12253-019-00777-z.

[32] Wang YJ, Ba Y, Chen QY, Han YQ. Clinical Significance of Peripheral Blood EBV-DNA Determination and Genotyping in lymphoma Patients. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2021; 29: 1802–1806. https://doi.org/10.19746/j.cnki.issn.1009-2137.2021.06.019. (In Chinese)

[33] Hau PM, Lung HL, Wu M, Tsang CM, Wong KL, Mak NK, *et al*. Targeting Epstein-Barr Virus in Nasopharyngeal Carcinoma. Frontiers in Oncology. 2020; 10: 600. https://doi.org/10.3389/fonc.2020.00600.

[34] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al*. Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England). 2007; 23: 2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

[35] Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics (Oxford, England). 2014; 30: 3276–3278. https://doi.org/10.1093/bioinformatics/btu531.

[36] Sample J, Young L, Martin B, Chatman T, Kieff E, Rickinson A, *et al*. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. Journal of Virology. 1990; 64: 4084–4092. https://doi.org/10.1128/JVI.64.9.4084-4092.1990.

[37] Gutiérrez MI, Raj A, Spangler G, Sharma A, Hussain A, Judde JG, *et al*. Sequence variations in EBNA-1 may dictate restriction of tissue distribution of Epstein-Barr virus in normal and tumour cells. The Journal of General Virology. 1997; 78 (Pt 7): 1663–1670. https://doi.org/10.1099/0022-1317-78-7-1663.

[38] MacKenzie J, Gray D, Pinto-Paes R, Barrezueta LFM, Arm-

strong AA, Alexander FA, *et al*. Analysis of Epstein-Barr virus (EBV) nuclear antigen 1 subtypes in EBV-associated lymphomas from Brazil and the United Kingdom. The Journal of General Virology. 1999; 80 (Pt 10): 2741–2745. https://doi.org/10.1099/0022-1317-80-10-2741.

[39] Fassone L, Bhatia K, Gutierrez M, Capello D, Gloghini A, Dolcetti R, *et al*. Molecular profile of Epstein-Barr virus infection in HHV-8-positive primary effusion lymphoma. Leukemia. 2000; 14: 271–277. https://doi.org/10.1038/sj.leu.2401651.

[40] Sandvej K, Zhou XG, Hamilton-Dutoit S. EBNA-1 sequence variation in Danish and Chinese EBV-associated tumours: evidence for geographical polymorphism but not for tumour-specific subtype restriction. The Journal of Pathology. 2000; 191: 127–131. https://doi.org/10.1002/(SICI)1096-9896(200006)191:2<127::AID-PATH614>3.0.CO;2-E.

[41] Do NV, Ingemar E, Phi PTP, Jenny A, Chinh TT, Zeng Y, *et al*. A major EBNA1 variant from Asian EBV isolates shows enhanced transcriptional activity compared to prototype B95.8. Virus Research. 2008; 132: 15–24. https://doi.org/10.1016/j.virusres.2007.10.020.

[42] Chang CM, Yu KJ, Mbulaiteye SM, Hildesheim A, Bhatia K. The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: a need for reappraisal. Virus Research. 2009; 143: 209–221. https://doi.org/10.1016/j.virusres.2009.07.005.

[43] Lorenzetti MA, Altcheh J, Moroni S, Moscatelli G, Chabay PA, Preciado MV. EBNA1 sequences in Argentinean pediatric acute and latent Epstein-Barr virus infection reflect circulation of novel South American variants. Journal of Medical Virology. 2010; 82: 1730–1738. https://doi.org/10.1002/jmv.21871.

[44] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution. 2018; 35: 1547–1549. https://doi.org/10.1093/molbev/msy096.

[45] Coleman CB, Daud II, Ogolla SO, Ritchie JA, Smith NA, Sumba PO, *et al*. Epstein-Barr Virus Type 2 Infects T Cells in Healthy Kenyan Children. The Journal of Infectious Diseases. 2017; 216: 670–677. https://doi.org/10.1093/infdis/jix363.

[46] Zidovec-Lepej S, Batovic M, Rozman M, Bodulić K, Prtorić L, Šokota A, *et al*. Distribution of Epstein-Barr Virus LMP1 Variants in Patients with Infectious Mononucleosis and Association with Selected Biochemical and Hematological Parameters. Pathogens (Basel, Switzerland). 2023; 12: 915. https://doi.org/10.3390/pathogens12070915.

[47] Banko A, Lazarevic I, Cupic M, Stevanovic G, Boricic I, Jovanovic T. Carboxy-terminal sequence variation of LMP1 gene in Epstein-Barr-virus-associated mononucleosis and tumors from Serbian patients. Journal of Medical Virology. 2012; 84: 632–642. https://doi.org/10.1002/jmv.23217.

[48] Dirnhofer S, Angeles-Angeles A, Ortiz-Hidalgo C, Reyes E, Gredler E, Krugmann J, *et al*. High prevalence of a 30-base pair deletion in the Epstein-Barr virus (EBV) latent membrane protein 1 gene and of strain type B EBV in Mexican classical Hodgkin's disease and reactive lymphoid tissue. Human Pathology. 1999; 30: 781–787. https://doi.org/10.1016/s0046-8177(99)90138-7.

[49] van Baarle D, Hovenkamp E, Dukers NH, Renwick N, Kersten MJ, Goudsmit J, *et al*. High prevalence of Epstein-Barr virus type 2 among homosexual men is caused by sexual transmission. The Journal of Infectious Diseases. 2000; 181: 2045–2049. https://doi.org/10.1086/315521.

[50] Crawford DH, Macsween KF, Higgins CD, Thomas R, McAulay K, Williams H, *et al*. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. Clinical Infectious Diseases. 2006; 43: 276–282. https://doi.org/10.1086/505400.

[51] Higgins CD, Swerdlow AJ, Macsween KF, Harrison N, Williams H, McAulay K, *et al*. A study of risk factors for acquisition of Epstein-Barr virus and its subtypes. The Journal of Infectious Diseases. 2007; 195: 474–482. https://doi.org/10.1086/510854.

[52] Palma I, Sánchez AE, Jiménez-Hernández E, Alvarez-Rodríguez F, Nava-Frias M, Valencia-Mayoral P, *et al*. Detection of Epstein-Barr virus and genotyping based on EBNA2 protein in Mexican patients with hodgkin lymphoma: a comparative study in children and adults. Clinical Lymphoma, Myeloma & Leukemia. 2013; 13: 266–272. https://doi.org/10.1016/j.clml.2012.11.010.

[53] Tabibzadeh A, Karbalaie Niya MH, Esghaei M, Bokharaei-Salim F, Ataei-Pirkooh A, Kiani SJ, *et al*. Molecular Epidemiology of Epstein-Barr virus (EBV) in Patients with Hematologic Malignancies. Asian Pacific Journal of Cancer Prevention: APJCP. 2020; 21: 693–698. https://doi.org/10.31557/APJCP.2020.21.3.693.

[54] Zhang XS, Wang HH, Hu LF, Li A, Zhang RH, Mai HQ, *et al*. V-val subtype of Epstein-Barr virus nuclear antigen 1 preferentially exists in biopsies of nasopharyngeal carcinoma. Cancer Letters. 2004; 211: 11–18. https://doi.org/10.1016/j.canlet.2004.01.035.

[55] Vafapour Z, Tabatabaie FH, Hosseini SY, Haghighat S, Hashemi SMA, Moattari A, *et al*. Sequence variation of the Epstein-Barr virus nuclear antigen 1 (EBNA1) gene in chronic lymphocytic leukemia and healthy volunteer subjects. Archives of Virology. 2023; 169: 1. https://doi.org/10.1007/s00705-023-05933-0.

[56] Xu M, Yao Y, Chen H, Zhang S, Cao SM, Zhang Z, *et al*. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. Nature Genetics. 2019; 51: 1131–1136. https://doi.org/10.1038/s41588-019-0436-5.

[57] Wong KW, Hui KF, Lam KP, Kwong DLW, Lung ML, Yang W, *et al*. Meta-analysis of Epstein-Barr virus genomes in Southern Chinese identifies genetic variants and high risk viral lineage associated with nasopharyngeal carcinoma. PLoS Pathogens. 2024; 20: e1012263. https://doi.org/10.1371/journal.ppat.1012263.

[58] Young LS. A novel Epstein-Barr virus subtype associated with nasopharyngeal carcinoma found in South China. Cancer Communications (London, England). 2020; 40: 60–62. https://doi.org/10.1002/cac2.12006.

[59] Yajima M, Kanda T, Takada K. Critical role of Epstein-Barr Virus (EBV)-encoded RNA in efficient EBV-induced B-lymphocyte growth transformation. Journal of Virology. 2005; 79: 4298–4307. https://doi.org/10.1128/JVI.79.7.4298-4307.2005.

[60] Jin Y, Xie Z, Lu G, Yang S, Shen K. Characterization of variants in the promoter of BZLF1 gene of EBV in nonmalignant EBV-associated diseases in Chinese children. Virology Journal. 2010; 7: 92. https://doi.org/10.1186/1743-422X-7-92.

[61] Li J, Zhang Y. Relationship between promoter sequence and its strength in gene expression. The European Physical Journal. E, Soft Matter. 2014; 37: 44. https://doi.org/10.1140/epje/i2014-14086-1.

[62] Kim H, Burassakarn A, Kang Y, Iizasa H, Yoshiyama H. A single nucleotide polymorphism in the BART promoter region of Epstein-Barr virus isolated from nasopharyngeal cancer cells. Biochemical and Biophysical Research Communications. 2019; 520: 373–378. https://doi.org/10.1016/j.bbrc.2019.10.028.

[63] Rozman M, Planinić A, Bodulić K, Škorić D, Gršković P, Marković AS, *et al*. Detection of Epstein-Barr virus coinfections based on LMP-1 gene diversity obtained by the high-throughput sequencing. Gene Reports. 2025; 41: 102369. https://doi.org/10.1016/j.genrep.2025.102369.