

The impact of Epstein-Barr Virus genomic variations on clinical subtypes and prognosis in EBV-associated diseases

Xurong Li

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Jing Wu

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Shenglei Yu

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Zhe Zhou

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Wenhong Zhang

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Lingyun Shao

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Ning Jiang

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Qingluan Yang

10301010198@fudan.edu.cn

Shanghai key laboratory of infectious diseases and biosafety emergency response, national medical center of infectious diseases, huashan hospital, fudan university

Research Article

Keywords: Epstein-Barr virus, infectious mononucleosis, chronic active EB virus infection, NK/T lymphoma, hemophagocytic lymphohistiocytosis, genomic variations

Posted Date: March 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8753453/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

Epstein-Barr virus (EBV) -associated diseases comprise a spectrum of conditions, including infectious mononucleosis (IM) and chronic active EBV infections (CAEBV), as well as malignant lymphomas. The study aimed to compare genomic variations of EBV across different subtypes of EBV-associated diseases to investigate the impact of EBV mutations on disease progression and clinical outcomes.

Methods

A total of 36 patients with EBV-associated diseases were enrolled. Peripheral blood DNA samples (400µL) were collected and sequenced using Illumina sequencing technology. Following quality control with FastQC, Fastp, and TriTrimmomatic, 27 samples meeting the criteria of minimum effective coverage > 10 and coverage > 80% were retained. Single nucleotide polymorphisms (SNP) annotations were generated into VCF files, and mutation profiles were visualized using heatmaps. Mutational rates were compared using Fisher's exact test, based on disease subtype, presence of hemophagocytic lymphohistiocytosis (HLH), and clinical outcomes. Odds ratios (ORs) were calculated using the Haldane-Anscombe correction.

Results

Compared with the reference strain B95.8, a total of 2182 mutation sites were identified in 27 samples. The mutations located in coding region included 523 missense variants, 1 insertion mutations, and 2 deletion mutations, all of which only involved single amino acid changes in the encoded proteins. Although a large number of mutations were detected, no statistically significant differences were observed among the IM, CAEBV and lymphoma groups, or between patients with and without HLH. However, three mutation sites in BFLF1 were significantly more frequent in deceased patients compared to survivors, which detailedly refers to Ile35Thr, Asp44Glu and Val246Ile [$P = 0.0267$, OR = 15.00, 95%CI= (0.7436,301.5)].

Conclusion

Single amino acid changes in EBV proteins had no significant effect on the progression status of EBV-associated diseases, but the mutation in BFLF1 were associated with an increased risk of mortality.

Introduction

Epstein-Barr virus (EBV), first identified and named by Epstein and Barr in 1964, is a widespread human herpesvirus classified as type 4. Its genome is a linear double-stranded DNA (dsDNA) approximately

172kb in length, encoding 85 genes and flanked by four internal repeat sequences (IRs) and terminal repeat sequences (TRs) at both ends^[1]. Upon binding to lymphocyte receptors (CD21) through its membrane glycoproteins and entering target cells, EBV DNA forms circular structure that remain in the cytoplasm, enabling long-term latency. The life cycle of EBV consists of two phases: latent infection and lytic replication. During latency, it mainly expresses EB nuclear antigen (EBNA) and membrane proteins (LMP) while in the lytic phase, early antigen (EA), capsid antigen (VCA), and membrane antigen (MA) are produced to utilize in viral assembly and cell lysis^[2].

EBV primarily infects epithelial cells and B cells. Latent EBV infection may induce immortalization of target lymphocytes, which is associated with lymphoproliferative disorders. The main types of Epstein-Barr virus(EBV) -associated diseases include: IM, CAEBV, and lymphomas (B-cell Burkitt lymphoma, Hodgkin's lymphoma, extranasal NK/T-type lymphoma, aggressive NK cell leukemia, and non-specific peripheral T lymphoma)^[3]. Most individuals infected EBV during childhood because of its transmission through saliva, though typically presenting as asymptomatic subclinical infections. In rare cases, acute primary infection triggers immune responses and benign lymphoid hyperplasia in B cells, leading to IM symptoms such as fever, pharyngitis, lymphadenopathy, hepatosplenomegaly and rash. The course usually lasts about one month, with most patients achieving complete remission. Only a small number of patients experience persistent IM symptoms or relapse lasting over six months, developing into rare CAEBV characterized by persistent IM-like symptoms and high EBV viremia, including fever, hepatosplenomegaly, lymphadenopathy, liver dysfunction and leukopenia. CAEBV has a poor prognosis of lymphoma progression or other severe complications. Besides its direct association with hematologic malignancies like lymphoma, EBV is also linked to various solid tumors such as nasopharyngeal carcinoma and gastric cancer.

Hemophagocytic lymphohistiocytosis (HLH) is a group of diseases characterized by excessive inflammatory responses caused by genetic or acquired immunodeficiency, the feature of which involves abnormal proliferation of activated lymphocytes and histiocytes, triggering cytokine storm that induces systemic inflammation and often leading to multiple organ dysfunction syndrome (MODS) or other life-threatening conditions^[4]. EBV-associated HLH is often linked to reactivation of EBV, frequently coexisting with CAEBV and lymphoma. Compared to other types of HLH, it has a poorer prognosis and higher mortality rate.

To investigate whether the EBV strain itself contributes to different disease manifestations and clinical outcomes, we developed an EBV-specific probe library and utilized Illumina technology to amplify and fluorescently label target EBV DNA samples. By measuring fluorescence intensity, we generated corresponding EBV DNA sequences and then compared them with the reference sequence B95.8 (NC007605). This comparative analysis aimed to investigate how nucleotide polymorphisms in EBV DNA sequences influence disease progression and clinical outcomes.

Methods

Patients

Our study enrolled 36 patients diagnosed with EBV-associated disease at Huashan Hospital, Fudan University. The cohort comprised 5 cases of IM, 10 cases of CAEBV, and 21 cases of NK/T lymphoma. All participants were adults (aged ≥ 18 years) and excluded individuals with diabetes, autoimmune diseases, or other chronic infections (including HBV/HCV, CMV, HIV, and tuberculosis). Informed consent was obtained from all subjects and the study protocol was approved by the Ethics Committee of Huashan Hospital, Fudan University.

Diagnostic Criteria for IM^[5]: The patient must meet any three of the following clinical manifestations and one of laboratory evidences of primary EBV infection: (1) Clinical manifestations: $\boxtimes$ fever; $\boxtimes$ pharyngitis; $\boxtimes$ enlarged cervical lymph nodes; $\boxtimes$ hepatosplenomegaly; $\boxtimes$ rash; $\boxtimes$ eyelid edema; (2) Laboratory findings: $\boxtimes$ positive anti-EBV-CA-IgM and anti-EBV-CA-IgG with negative anti-EBV-NA-IgG; $\boxtimes$ positive single anti-EBV-CA-IgG antibody with low-affinity binding to EBV-CA-IgG; (3) Exclusion of acute infection caused by other pathogens.

Diagnostic criteria for CAEBV^[6]: (1) Persistent or recurrent infectious mononucleosis-like symptoms lasting over three months; (2) Increased EBV genomic load in peripheral blood and/or affected tissues (elevated EBV DNA or positive EBV EER); (3) EBV infection affecting T cells, NK cells, or multiple cell lineages (two or more types including T, B, or NK); (4) Exclusion of other chronic diseases that may cause similar symptoms. EBV-associated NK/T lymphoma can be diagnosed according to WHO-HAEM5 and increased EBV genome in peripheral blood and/or affected tissues (elevated EBV DNA or EBER positive)^[7]. EBV-associated HLH can be diagnosed according HLH-2024^[8]: Increased EBV genomic load in peripheral blood and/or affected tissues (elevated EBV DNA or positive EBV DNA test).

EBV genomic sequencing

Genomic DNA was isolated from 400 μ L of peripheral blood. DNA mass was quantified fluorometrically using Qubit (Thermo Fisher Scientific). The EBV DNA load was determined by absolute quantification qPCR using primers GCCGGTGTGTTTCGTATATGG/CAAAACCTCAGCAAATATATGAG. Threshold cycle (CT) values were compared to a standard curve generated from a linearized plasmid containing a 598-bp target sequence amplified from the EBV genome (primers: AGGCCATTTTTCCACCCTGT/CGTCAAAGCTGCACACAGTC) to calculate EBV genome copies per microliter. Samples exceeding 5.0×10^3 copies/ μ g DNA were selected for sequencing.

Genomic DNA (1 μ g) was fragmented to an average size of ~ 250 bp using the Covaris ME220 System (Covaris, Woburn, MA) according to the manufacturer's recommendations. Subsequent library preparation steps—including end repair, 3'-dA tailing, Illumina-compatible adapter ligation, and post-ligation cleanup—were performed according to the SureSelectXT Illumina Paired-End Sequencing Library protocol (Agilent Technologies). Custom biotinylated RNA baits targeting the EBV genome were designed using the SureDesign web tool (Agilent Technologies, Santa Clara, CA). Target enrichment was performed using the SureSelectXT Target Enrichment Kit (Agilent Technologies, USA) with hybridization

at 65°C, strictly following the manufacturer's protocol. Post-hybridization capture cleanup and enrichment PCR were conducted following the kit instructions. Final enriched libraries were quantified using the Kapa Library Quantification Kit (Kapa Biosystems, Wilmington, MA) on an ABI 9700HT Fast Real-Time PCR system (Thermo Fisher Scientific) and sequenced on the Illumina NovaSeq platform (2x150 bp paired-end reads).

Quality control of raw data

The 36 FastQC files generated by the platform were processed through FastP and Trimmomatic in Java programs, with results shown in supplementary figure 1. We observed that Per Base Sequence Content remained abnormal after processing by both FastP and Trimmomatic, indicating that the issue might not stem from low-quality sequences or adapter contamination, but rather from library construction or probe bias. Given the similar results from FastP and Trimmomatic, we selected the BAM file processed by FastP for subsequent analysis. We defined effective coverage as DP>10 and valid samples as coverage>80%, ultimately eliminating 9 invalid samples, with results shown in supplementary table 1. The remaining 27 valid samples were further analyzed, including 2 cases of IM, 9 cases of CAEBV, and 16 cases of NK/T lymphoma.

Mutation annotation

Using B95.8 (NC007605) as the reference sequence, we performed variant detection on 27 valid samples through VCF analysis, covering single nucleotide polymorphisms (SNPs), insertions, and deletions. As EBV is haploid with only one allele per locus, the VCF annotation directly uses 0 (reference allele) and 1 (variant allele) markers. Mutations within gene repeated regions were excluded from subsequent analyses. We also utilized bioinformatics tools to generate phylogenetic trees for visualizing genetic mutation patterns.

Structural prediction of mutant protein

We obtained relevant information of BFLF1 protein of B95.8 from the UniProt database, including its amino acid sequence and spatial structure. Based on the amino acid changes caused by DNA mutation, we generated the amino acid sequence of the mutant protein. Subsequently, both the wild-type and mutant protein sequences were respectively imported into AlphaFold3 to predict their quaternary structures. Finally, ChimeraX was used to visualize the structures.

Statistical analysis

The presence of mutations at a certain site was expressed by 0/1 and displayed in the form of heat map. The comparison of mutation rates between groups was expressed in the form of contingency table, and the statistical comparison was made by Fisher exact probability method. The 95% confidence interval (CI) for the odds ratio (OR) was calculated using the Haldane-Anscombe correction. $P<0.05$ was considered as statistical difference. GraphPad Prism9.5 was used to complete the analysis.

Results

1. Characteristic of enrolled patients

27 patients whose sequencing data meet the criteria of quality control showed no significant difference in any baseline characteristics except hemoglobin ($P=0.037$) among IM, CAEBV and NK/T lymphoma. NK/T lymphoma patients had a lower number of hemoglobin than other 2 groups indicating that proliferating lymphoma cells would lead to hematopoietic disorder. Notably, EBV-DNA in whole blood from CAEBV patients was higher than that in patients with IM and lymphoma. This discrepancy may reflect the systemic massive expansion of EBV in infected cells in CAEBV. More detailed baseline characteristics are presented in Table 1.

2. Phylogenetic map

The 27 patients in the final cohort were from six provinces: Anhui, Jiangsu, Jiangxi, Shanghai, Xinjiang, and Zhejiang, indicating considerable regional diversity. To assess whether disease type and geographic distribution were associated with viral genetic variation, we conducted a phylogenetic analysis to identify the EBV strain sources and their relationship. The results revealed that the submitted samples clustered across multiple evolutionary branches, with no significant pathological and geographical pattern observed among them (Figure 1).

3. Mutation between different groups

3.1 Landscape of mutation sites in analyzed samples

The sequences of the 27 analyzed samples were compared with the B95.8 sequence (GenBank accession: NC 007605) from the NCBI database and performed SNP annotation. 2182 single nucleotide variants (SNVs), 12 insertion mutations, and 11 deletion mutations were identified relative to the reference. Among these, only 523 were missense mutations, and only one insertion and two deletions resulted in amino acid changes. Mutations were distributed across 74 viral proteins, including BNRF1, BPLF1, BMRF2, etc. All identified mutations altered only a single codon, specifically modifying individual amino acids rather than causing broader disruptions such as reading frameshifts, premature termination, or promoter insertions (Table 2).

3.2 Comparison of mutation sites across EBV-associated disease subtypes

Patients were categorized into three groups based on their EBV-associated disease subtype: IM, CAEBV or EBV⁺NK/T lymphoma groups. Heatmaps were generated to visualize mutation profiles across groups (Figure 2), and comparative analysis of individual mutation sites revealed no statistically significant differences. These results indicate that single-amino-acid changes in EBV proteins are not a major determinant of disease subtype classification.

The incidence of concurrent HLH varied among the disease subtypes. No HLH cases were observed in the IM group, while the incidence rate was 2/7 in the CAEBV group. Notably, lymphoma patients exhibited a significantly higher rate of 7/16 (Table 1). To examine whether EBV sequence variation is associated with HLH development, mutation profiles were compared between patients with and without HLH (Figure 3). No statistically significant mutation sites were identified, indicating that single-amino-acid substitutions in EBV proteins are not linked to HLH occurrence in this cohort.

3.3 Comparison of mutation sites with clinical outcomes

Disease severity and clinical outcomes vary considerably among the three EBV-associated diseases included in this study: IM, CAEBV and EBV-positive NK/T-cell lymphoma. While IM typically follows a mild course with most patients recovering spontaneously, CAEBV and EBV-positive NK/T-cell lymphoma frequently involve severe disease progression. These conditions are often characterized by widespread infiltration of proliferating lymphocytes or tumor cells into multiple organs, leading to life-threatening organ dysfunction. Based on hospital records and follow-up data, all IM patients in our cohort achieved remission or were completely cured. In contrast, the remission rate was 0% in the CAEBV group and 18.8% in the lymphoma group, respectively. Mortality was notably high, reaching 22.2% in CAEBV patients and 43.8% in lymphoma patients.

To evaluate whether EBV sequence variation influences clinical outcomes, we compared genomic polymorphisms across the study population (Figure 4). After excluding two patients lost to follow-up, three specific mutations in the BFLF1 gene were found to be significantly associated with mortality: nucleotide changes at positions 104, 132, and 736, resulting in amino acid substitutions at protein positions 35, 44, and 246, respectively. These mutations were significantly more frequent in deceased patients than in survivors ($P=0.0267$) (Figure 5). The corresponding odds ratio (OR) was as high as 15.00 [95% CI=0.7436, 301.5].

4. Structural analysis of mutant BFLF1

Given that the three mutation sites in the BFLF1 gene, which were found to be significantly associated with clinical outcomes, co occurred in a single patient, we hypothesized that they represent a distinct mutational pattern. To evaluate the potential structural impact, we introduced all three single amino acid substitutions simultaneously into a mutant protein model. According to Uniprot, the native BFLF1 protein consists of two subunits, each assembled from five identical polypeptide chains. Due to the current limitation of AlphaFold3, we ultimately generated one of these two subunits. The structure was visualized using ChimeraX (Figure 6), with the three substituted amino acid positions marked in red. Comparison between the wild type and mutant models revealed no apparent structural alterations, suggesting that these amino acid changes have a negligible effect on the overall conformation of the BFLF1 protein.

Discussion

EBV-associated diseases encompass a broad spectrum ranging from benign proliferative disorders to highly malignant tumors. Their heterogeneity poses significant challenges for clinical diagnosis, treatment selection, and prognosis assessment. Our research focuses on analyzing how genetic polymorphisms in EBV sequence influence the complex pathogenesis and clinical outcomes of EBV-associated diseases. Tiny disparity was observed among patients with different types of EBV-associated diseases. Only 3 sites on BFLF1 encoding gene were found to be more frequent in EBV from patients with poor outcomes.

The development of EBV-associated diseases involves the transformation and proliferation of EBV-infected lymphocytes, coupled with an imbalanced host immune surveillance against viral-infected cells. Various proteins encoded by EBV exhibit immunosuppressive effects. EBNA1 inhibits Toll-like receptor (TLR) pathways, while BPLF1 suppresses retinoic acid-inducible gene I (RIG-I) and interferon (IFN) signaling pathways. These mechanisms enable EBV to evade immune defense and surveillance, achieving prolonged latency and short-term lytic replication in infected cells^[9]. Additionally, acquired immune deficiencies caused by human immunodeficiency virus (HIV) infection or immunosuppressive agents, along with inborn errors of immunity (IEIs), all increase susceptibility to EBV-associated diseases^[10].

EBV can exist in host cells through multiple latent modes, expressing distinct latent genes such as EBNA1, EBNA2, LMP1, and LMP2A/B. These genes play crucial roles in maintaining viral latency, promoting cell proliferation, inhibiting apoptosis and evading immune clearance. During the lysis and replication phase, EBV switches to different gene expression patterns, expressing transcription factors, DNA polymerase, and viral structural proteins (e.g. BZLF1, BALF5 and BGLF4). These proteins mediate biological processes including replication of genetic material, assembly of viral structures and lysis of infected cells to release viral particles, to complete the viral life cycle. The differences in EBV gene expression patterns, variations in host immune status, and diversity in affected lymphocyte types collectively shape the remarkable heterogeneity of EBV-associated diseases. Nguyen et al^[11] identified 1,453 EBV transcript isoforms through long-read sequencing technology, revealing the complexity of EBV genome expression. Tian et al^[12] revealed that EBV utilizes host CCCTC-binding factors and RNA polymerase II (RNAPII) to form distinct chromatin contact domains and RNAPII-associated interaction domains through structural analysis of the EBV genome and human lymphoblast chromatin. Notably, EBV's non-coding RNAs, particularly its encoded small RNAs (EBERs), target chromatin regions with low accessibility, inhibiting host gene expression by blocking RNAPII-associated chromatin loops, which demonstrates multifaceted genomic interactions between EBV and the host genome. By comparing EBV genomes derived from 27 patients with the reference sequence B95.8, we identified 2159 single nucleotide variants (SNVs), 12 insertion mutations, and 11 deletion mutations. Most (1636 SNVs, 11 insertions, and 9 deletions) were located in regulatory regions (e.g. upstream/downstream of coding frames or intergenic regions), with no direct amino acid alteration in protein. These findings indicate that EBV sequences in EBV-associated diseases patients remain relatively conserved, and genomic polymorphisms do not lead to protein structural diversity.

Despite the limited number of mutations detected in the reading frame, the 523 missense mutations, 1 insertion mutation, and 2 deletion mutation involve single amino acid alterations in 74 EBV structural proteins with multiple functions. For example, BNRF1, BPLF1 and BKRF4 are tegument proteins located between the capsid and envelope, playing crucial roles in nucleic acid transportation, virus assembly and immune evasion^[13]. Notably, the rest 73 proteins except BARF0 with detected amino acid changes were expressed during the lytic phase, suggesting a close relationship between structural changes in lytic phase proteins and EBV-associated diseases pathogenesis. Muruta et al^[14] noted that genome deletion causes overexpression of lytic phase genes, while EBV's failure to complete its lysis due to missing key genes leads to persistent infection in host cells, ultimately promoting the development of EBV-associated diseases such as CAEBV. EBV genome mutations not only affect EBV-associated diseases but also influence other EBV-related diseases. Xu et al^[15] identified two BALF2 mutation sites with higher risk of nasopharyngeal carcinoma through genomic sequencing of patients with nasopharyngeal carcinoma, other EBV-associated diseases and healthy individuals, providing a valuable tool for screening nasopharyngeal carcinoma. Borozan et al^[16], on the other hand, detected 13 novel sequences in the EBV genomes of patients with EBV-associated gastric adenocarcinoma. Their comparison with B95.8 revealed that mutation sites were predominantly concentrated in latency genes such as EBNA1 and LMP1, highlighting the role of these genes in carcinogenesis. Due to insufficient sample size and limited disease spectrum, we failed to identify statistically significant mutation sites from the discovered variants. Nevertheless, this finding suggests that EBV in EBV-associated diseases patients, which promotes pathological proliferation of lymphocytes rather than epithelial cells, possesses distinct genetic sequences compared to other EBV-associated epithelial malignancies.

Although not finding any difference between disease types and the risk of HLH development, our comparison of patients with different clinical outcomes revealed that mutations at three sites in the BFLF1 gene were significantly more prevalent in death patients than in survival patients. This suggests that changes in the BFLF1 protein structure may influence the mortality risk in EBV-associated diseases. The BFLF1 protein, also known as the DNA packaging protein UL32, plays a crucial role in the herpesviral life cycle by participating in the cleavage and packaging of the viral genome. Its main function is to encapsulate newly replicated viral DNA into the capsid, thereby forming mature viral particles^[17]. Pavlova et al^[18] successfully constructed an EBV mutant devoid of BFLF1 which produced virus-like particles (VLPs) free of EBV DNA that elicited a strong CD4⁺ T cell response. It highlights the essential role of BFLF1 in the assembly of EBV virion structure. While the role of UL32 has been extensively studied in other herpesviruses, there is a remarkable lack of research focusing on the specific function of BFLF1 in EBV-associated diseases. Our finding, therefore, assists to fill this gap in the literature.

Given the global prevalence of EBV, we aimed to investigate regional differences in EBV strains and their impact on EBV-associated diseases through phylogenetic analysis. EBV-associated tumors including nasopharyngeal carcinoma, gastric cancer, and lymphoma exhibit significant regional variations. A 2020 epidemiological study revealed that between 1990 and 2017, East Asia accounted for 43% of EBV-related tumor cases and 40% of related deaths, because of the high prevalence of nasopharyngeal and gastric

cancers in East Asia particularly China^[19]. Palser et al^[20] analyzed EBV DNA sequences from 71 healthy individuals and EBV-associated tumor patients across multiple countries and regions, covering East Asia, Western Europe, Central Africa, North America and Oceania. They identified genetic variations concentrated in latency genes such as EBNA2 and LMP1, which may be crucial for distinguishing asymptomatic EBV infections from pathogenic cases. Our study enrolled patients from five eastern Chinese provinces (Shanghai, Jiangsu, Zhejiang, Anhui, Jiangxi) and western Xinjiang. Through phylogenetic analysis, we found that the results did not demonstrate significant regional differences due to insufficient geographical diversity of the included patients. Lin et al^[21] detected EBV sequences in the peripheral blood of 8 Chinese patients with extranodal NK/T cell lymphoma (nasal type) (NKTCL). The phylogenetic tree also confirmed that their EBV sequences were similar to those of other Asian NKTCL patients, further indicating the relative conservation of EBV sequences in Asia.

Although this study failed to identify statistically significant mutations in the EBV DNA sequences of EBV-associated diseases patients from the perspective of disease subtypes and HLH incidence, our findings from other EBV-associated diseases provide indirect evidence of both internal homogeneity within EBV-associated diseases and heterogeneity across different conditions. BFLF1 mutations, which are associated with an increased mortality risk, may provide reliable prognostic markers for the clinical outcome of EBV-associated diseases. Our future research will focus on expanding the sample size, incorporating patients with other EBV-related diseases and populations from diverse regions to further investigate DNA sequence polymorphisms. Additionally, we will employ multi-omics approaches such as transcriptomics and proteomics to conduct comprehensive analysis of EBV-associated diseases.

Declarations

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are available in the GenomeSequence Archive (CRA038738). The data can be accessed now on <https://ngdc.cncb.ac.cn/gsa/browse/CRA038738>.

Competing interest

The authors declare that they have no competing interests.

Funding

This work was supported by research grants from the Shanghai Municipal Health Commission (20244Y0025, SHWSRS(2024)_070).

Ethics approval and consent to participate

All procedures involving human participants were performed in accordance with the ethical standards of the Declaration of Helsinki. This study was reviewed and approved by the Ethics Committee of Huashan Hospital (KY2020-1122).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Authors' contributions

X.L., J.W., S.Y. and Q.Y. contributed to the design of the study, directed the field work and collection of study data, analysis and interpretation of data, co-writing of the first draft, and approval of the manuscript. J.W. and Z.Z. contributed to enrolled the participants, follow up the participants, collection of data and approval of the manuscript. X.L. and N.J. contributed to analysis of the published data and approval of the manuscript. W.Z. and L.S. contributed to the design of the study, supervised the field trial, analysis of the published data, and writing and approval of the manuscript.

Acknowledgments

We thank all the participants of the study for their willingness to participate in this research. A version of the abstract related to this study was previously deposited as a preprint in the 4TU.ResearchData repository[22].

References

1. Tierney RJ, Kao KY, Nagra JK, Rickinson AB. Epstein-Barr virus BamHI W repeat number limits EBNA2/EBNA-LP coexpression in newly infected B cells and the efficiency of B-cell transformation: a rationale for the multiple W repeats in wild-type virus strains. *J Virol*. 2011 Dec;85(23):12362-75. doi: 10.1128/JVI.06059-11.
2. Münz C. Latency and lytic replication in Epstein-Barr virus-associated oncogenesis. *Nat Rev Microbiol*. 2019 Nov;17(11):691-700. doi: 10.1038/s41579-019-0249-7.
3. Quintanilla-Martinez L, Swerdlow SH, Tousseyn T, Barrionuevo C, Nakamura S, Jaffe ES. New concepts in EBV-associated B, T, and NK cell lymphoproliferative disorders. *Virchows Arch*. 2023 Jan;482(1):227-244. doi: 10.1007/s00428-022-03414-4.
4. Janka GE. Hemophagocytic syndromes. *Blood Rev*. 2007 Sep;21(5):245-53. doi: 10.1016/j.blre.2007.05.001.

5. Walther LE, Ilgner J, Oehme A, Schmidt P, Sellhaus B, Gudziol H, Beleites E, Westhofen M. Die infektiöse Mononukleose [Infectious mononucleosis]. *HNO*. 2005 Apr;53(4):383-92; quiz 393. German. doi: 10.1007/s00106-004-1210-1.
6. Kawada JI, Ito Y, Ohshima K, Yamada M, Kataoka S, Muramatsu H, Sawada A, Wada T, Imadome KI, Arai A, Iwatsuki K, Ohga S, Kimura H; Committee for Guidelines for the Management of Chronic Active EBV Disease, Related Disorders (the MHLW Research Team in Japan). Updated guidelines for chronic active Epstein-Barr virus disease. *Int J Hematol*. 2023 Nov;118(5):568-576. doi: 10.1007/s12185-023-03660-5.
7. Miranda RN, Amador C, Chan JKC, Guitart J, Rech KL, Medeiros LJ, Naresh KN; WHO fifth edition classification project. Fifth Edition of the World Health Organization Classification of Tumors of the Hematopoietic and Lymphoid Tissues: Mature T-Cell, NK-Cell, and Stroma-Derived Neoplasms of Lymphoid Tissues. *Mod Pathol*. 2024 Aug;37(8):100512. doi: 10.1016/j.modpat.2024.100512.
8. Henter JI, Horne A, Aricó M, Egeler RM, Filipovich AH, Imashuku S, Ladisch S, McClain K, Webb D, Winiarski J, Janka G. HLH-2004: Diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer*. 2007 Feb;48(2):124-31. doi: 10.1002/pbc.21039.
9. Yuan L, Zhong L, Krummenacher C, Zhao Q, Zhang X. Epstein-Barr virus-mediated immune evasion in tumor promotion. *Trends Immunol*. 2025 May;46(5):386-402. doi: 10.1016/j.it.2025.03.007.
10. Latour S. Human Immune Responses to Epstein-Barr Virus Highlighted by Immunodeficiencies. *Annu Rev Immunol*. 2025 Apr;43(1):723-749. doi: 10.1146/annurev-immunol-082323-035455.
11. Nguyen TD, Wang J, Lam MT, McFerrin H, O'Grady TM, Roberts C, Van Otterloo N, Nguyen TT, Baddoo M, Wyczechowska D, Morales M, Dong Y, Flemington EK. Comprehensive resolution and classification of the Epstein Barr virus transcriptome. *Nat Commun*. 2025 Jul 10;16(1):6381. doi: 10.1038/s41467-025-61870-3.
12. Tian SZ, Yang Y, Ning D, Yu T, Gao T, Deng Y, Fang K, Xu Y, Jing K, Huang G, Chen G, Yin P, Li Y, Zeng F, Tian R, Zheng M. Landscape of the Epstein-Barr virus-host chromatin interactome and gene regulation. *EMBO J*. 2025 Jul;44(13):3872-3915. doi: 10.1038/s44318-025-00466-5.
13. Murata T. Tegument proteins of Epstein-Barr virus: Diverse functions, complex networks, and oncogenesis. *Tumour Virus Res*. 2023 Jun;15:200260. doi: 10.1016/j.tvr.2023.200260.
14. Murata T, Okuno Y, Sato Y, Watanabe T, Kimura H. Oncogenesis of CAEBV revealed: Intragenic deletions in the viral genome and leaky expression of lytic genes. *Rev Med Virol*. 2020 Mar;30(2):e2095. doi: 10.1002/rmv.2095.
15. Xu M, Yao Y, Chen H, Zhang S, Cao SM, Zhang Z, Luo B, Liu Z, Li Z, Xiang T, He G, Feng QS, Chen LZ, Guo X, Jia WH, Chen MY, Zhang X, Xie SH, Peng R, Chang ET, Pedergrana V, Feng L, Bei JX, Xu RH, Zeng MS, Ye W, Adami HO, Lin X, Zhai W, Zeng YX, Liu J. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nat Genet*. 2019 Jul;51(7):1131-1136. doi: 10.1038/s41588-019-0436-5.
16. Borozan I, Zaparka M, Frappier L, Ferretti V. Analysis of Epstein-Barr Virus Genomes and Expression Profiles in Gastric Adenocarcinoma. *J Virol*. 2018 Jan 2;92(2):e01239-17. doi: 10.1128/JVI.01239-

17. Erratum in: J Virol. 2018 Apr 27;92(10):e00330-18. doi: 10.1128/JVI.00330-18.
17. Sugimoto A, Yamashita Y, Kanda T, Murata T, Tsurumi T. Epstein-Barr virus genome packaging factors accumulate in BMRF1-cores within viral replication compartments. PLoS One. 2019 Sep 13;14(9):e0222519. doi: 10.1371/journal.pone.0222519.
18. Pavlova S, Feederle R, Gärtner K, Fuchs W, Granzow H, Delecluse HJ. An Epstein-Barr virus mutant produces immunogenic defective particles devoid of viral DNA. J Virol. 2013 Feb;87(4):2011-22. doi: 10.1128/JVI.02533-12.
19. Khan G, Fitzmaurice C, Naghavi M, Ahmed LA. Global and regional incidence, mortality and disability-adjusted life-years for Epstein-Barr virus-attributable malignancies, 1990-2017. BMJ Open. 2020 Aug 30;10(8):e037505. doi: 10.1136/bmjopen-2020-037505.
20. Palser AL, Grayson NE, White RE, Corton C, Correia S, Ba Abdullah MM, Watson SJ, Cotten M, Arrand JR, Murray PG, Allday MJ, Rickinson AB, Young LS, Farrell PJ, Kellam P. Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. J Virol. 2015 May;89(10):5222-37. doi: 10.1128/JVI.03614-14.
21. Lin N, Ku W, Song Y, Zhu J, Lu Z. Genome-Wide Analysis of Epstein-Barr Virus Isolated from Extranodal NK/T-Cell Lymphoma, Nasal Type. Oncologist. 2019 Sep;24(9):e905-e913. doi: 10.1634/theoncologist.2017-0588.
22. Li XR, Wu J, Yu SL, Zhou Z, Zhang WH, Shao LY, Jiang N, Yang QL. The impact of Epstein-Barr Virus genomic variations on clinical subtypes and prognosis in EBV-associated diseases. Url: https://data.4tu.nl/private_datasets/Mcui8_q05h_qSV71balx0le6NI_RGPdGh7vTVfB9G_c.

Tables

Table 1. Basic information of EBV-LPDs patients

	IM	CAEBV	NK/T Lymphoma
Number	2	9	16
Gender [Male (%)]	1 (50%)	6 (66.7%)	11 (68.8%)
Age [M(Q1,Q3)]	20 (15,25)	30 (23,40)	38 (22,55)
Province [count (%)]			
Anhui	0	4 (44.4%)	5 (31.3%)
Jiangsu	1 (50%)	0	4 (25%)
Jiangxi	0	1 (11.1%)	2 (12.5%)
Shanghai	1 (50%)	1 (11.1%)	1 (6.25%)
Xinjiang	0	0	1 (6.25%)
Zhengjiang	0	1 (11.1%)	3 (18.8%)
Unknown	0	2 (22.2%)	0
Symptoms ¹ [count (%)]			
Fever	1 (100%)	7 (100%)	13 (86.7%)
Pharyngalgia	1 (100%)	0	1 (6.67%)
Rash	0	1 (14.3%)	5 (33.3%)
Lymphadenectasis	1 (100%)	2 (28.6%)	5 (33.3%)
Hepatomegaly	0	0	5 (33.3%)
Splenomegaly	1 (100%)	6 (85.7%)	12 (80%)
Comorbidity [count(%)]			
Hemophagocytic lymphohistiocytosis	0	2 (22.2%)	7 (43.8%)
Diabetes	0	0	1 (6.25%)
Chronic Nephropathy	0	0	0

	HIV infection	0	0	0
Blood routine ² [M(Q1,Q3)]				
	WBC count (x10 ⁹ /L)	11.0 (2.37,19.6)	3.60 (2.09,5.60)	2.67 (1.40,6.72)
	Neutrophil ratio (%)	34.1 (15.53,2)	53.5 (32.4,77.8)	59.5 (52.6,79.1)
	Lymphocyte ratio (%)	24.8 (18,31.6)	36.8 (15.0,48.1)	19.8 (11.5,35.3)
	Platelet (x10 ⁹ /L)	142 (121,163)	150 (80,215)	94 (36,143)
	Hemoglobin (g/L)	143 (120,166)	116 (109,142)	91 (81,119)
EBV DNA ³ [M(Q1,Q3)]				
	Whole blood (x10 ⁶ copies/L)	1.03 (0.1,1.95)	47.3 (3.25,93.5)	4.44 (0.1,27.7)
	Plasma (x10 ⁴ copies/L)	0.044 (0.04,0.048)	1.12 (0.78,38.1)	1.58 (1.11,12.9)
PBMC sorting ⁴ [M(Q1,Q3)]				
	CD3+ T cell (%)	78.8 (68.7,88.9)	48.4 (38.2,60.1)	49.5 (30.1,78.0)
	CD4+ T cell (%)	30.2 (7.57,52.8)	25.9 (18.8,38.7)	25.2 (18.6,46.8)
	CD8+ T cell (%)	45.5 (13.7,77.4)	14.8 (9.39,30.4)	18.3 (7.61,25.3)
	CD4/CD8 ratio	1.98 (0.1,3.85)	2.13 (0.62,3.36)	1.80 (1.21,2.69)
	CD19+ B cell (%)	7.30 (0.82,13.8)	8.17 (1.30,14.3)	5.53 (3.98,10.1)
	CD56+ NK cell (%)	13.7 (9.93,17.5)	38.6 (26.1,59.9)	31.3 (11.8,66.3)
Outcome [count (%)]				

Cure	1 (50%)	0	0
Improvement	1 (50%)	0	3 (18.8%)
Stable	0	5 (55.6%)	6 (37.5%)
Death	0	2 (22.2%)	7 (43.8%)
Loss of follow up	0	2 (22.2%)	0

¹ 1, 2 and 1 patients in IM group, CAEBV group and lymphoma group respectively failed to record their symptoms.

² 2 patients in CAEBV group refused to test blood routine and 3 patients in lymphoma group failed to categorize neutrophils and lymphocyte due to their insufficient peripheral blood cells.

³ 3 patients in CAEBV group refused to quantify EBV DNA both in whole blood and plasma while 2 patients in lymphoma group lost their results only in whole blood.

⁴ 3 patients respectively in CAEBV group and lymphoma group lost their results.

Table 2. Mutation site of EBV proteins

Protein	Period	Mutation site
BSLF2/BMLF1 (Mta)	Immediate early lytic phase (IE)	Arg9Gln
		Leu39Met
		Glu43Asp
		Ile70Val
		Arg73Gln
		Gly465Ala
BRRF1	Immediate early lytic phase (IE)	Ser4dup
		Asp42Asn
		Lys69Glu
		Asp101Ala
		Ile159Leu
		Ser236Ala
		His261Tyr
		Val300Ala
BRRF2	Immediate early lytic phase (IE)	Val33Asp
		Asp56Asn
		Asp81Asn
		Ala85Thr
		Leu184Pro
		Asp202Asn
BHRF1	Early lytic phase (E)	Thr24Ser
		Arg39Cys
		Val79Leu
		Leu88Val
BaRF1	Early lytic phase (E)	Ala119Thr
BMRF1	Early lytic phase (E)	Val232Ser
		Val232Ala

BMRF2	Early lytic phase (E)	Leu280Met
		Ala307Thr
		Asp201Asn
		Leu219Met
		Ala317Ser
BSLF1	Early lytic phase (E)	Ala3Thr
		His77Arg
		Arg97His
		Asp103His
		Gln125His
		Ala135Val
		Ala245Thr
		Met290Val
		Arg400Gln
		Thr461Met
		Ala511Pro
		Val577Ile
		Asn675Lys
BSRF1	Early lytic phase (E)	Gln809Glu
		Pro178Ala
BBLF4	Early lytic phase (E)	Thr218Met
		Arg27Cys
		Arg37Lys
		Gln41Leu
		Thr94Ser
		Leu143Ile
		Lys263Asn
		Leu322Met
		Ala339Thr

		Thr360Ala
		Ala374Thr
		Thr416Asn
		Leu540Met
		Arg543His
		His646Asn
		Lys675Glu
		Lys675Arg
		Ser729Thr
		Arg804Cys
BBLF2/BBLF3	Early lytic phase (E)	Ile126Leu
		Val153Gly
		Met160Val
		Asp231Asn
		Ala232Thr
		Arg236His
		Arg236His
		Pro496Thr
		Glu503Lys
		Arg522Trp
		Val534Met
		Gly548Ala
		His600Arg
		Ser640Pro
		Thr690Ala
		Ile691Met
		Thr696Met
BBLF1	Early lytic phase (E)	Ala42Val
BGLF5	Early lytic phase (E)	Phe46Tyr

		Lys79Gln
		Glu93Ala
		Pro377Arg
BGLF4	Early lytic phase (E)	Gln131His
		Thr292Ala
		Pro402Ser
BGLF3.5	Early lytic phase (E)	His54Arg
BGLF3	Early lytic phase (E)	Pro86Ser
		Gln150His
		Asn204Ser
		Val312Ile
BGRF1/BDRF1	Early lytic phase (E)	Ala17Val
		Val75Met
		Ala78Val
		Gln111Leu
		Gln344Lys
		Ser354Phe
		Val533Ala
		Ser616Ile
BDLF3.5	Early lytic phase (E)	Pro4Leu
		Gly15Val
		Glu21Lys
		Thr36Ala
		Val42Met
		Glu51Gln
BcRF1	Early lytic phase (E)	Thr33Ala
		Arg70Gln
		Met71Ile
		Asn133Ser

		Arg171Gln
		Asn181Thr
		Asn181Lys
		Thr231Ile
		Leu240Val
		Ser400Phe
		Arg548Gln
		Gln569Arg
		Cys745Arg
BTRF1	Early lytic phase (E)	Ala136Thr
		Ala193Thr
		Thr208Ala
		Gly218Arg
		Pro249Ala
		Pro285Gln
		Ile353Ser
BALF5	Early lytic phase (E)	Ser154Arg
		Arg227Gln
		Asp313Asn
		Arg334His
		Ser402Thr
		Ala481Val
		Glu746Lys
		Asp776Glu
		His861Gln
		Thr916Ala
Asp977Asn		
BALF2	Early lytic phase (E)	Ser181Ala
		Ser292Leu

		Val317Met
		Leu416Pro
		Ile613Val
		Val700Leu
		Ala1064Val
		Glu1065Gly
		Ala1066Thr
		Ala1067Thr
		Ser1093Gly
		Ser1099Pro
BALF1	Early lytic phase (E)	Arg91Gln
		Val90Ile
		Pro15Ser
BARF1	Early lytic phase (E)	Val29Ala
BNLF2b	Early lytic phase (E)	Pro3Gln
		Ala12Asp
		Leu14Val
		Arg19Thr
		Gln69Lys
		Phe71Cys
BNLF2a	Early lytic phase (E)	Ser29Asn
BNRF1	Late lytic phase (L)	Val12Ile
		Glu36Gln
		Leu61Gln
		Ala65Thr
		Gln134Glu
		Gly200Arg
		Leu270Trp
		Thr280Ala

		Gly456Arg
		Ala485Ser
		Arg486Gly
		Ser497Gly
		Gly549Ser
		Phe573Ile
		Pro694His
		Cys736Phe
		Asn760Thr
		Ala762Ser
		Asn797Ser
		Met944Ile
		Arg1071His
		Glu1072Asp
		Phe1110Ser
		Ala1116Ser
		Val1222Ile
		Ala1289Thr
BCRF1	Late lytic phase (L)	Val6Met
		Gly23Ser
		Phe67Leu
BFLF2	Late lytic phase (L)	Asp7Glu
		Ala20Ser
		Ala20Thr
		Ser63Arg
		Ile215Leu
		Ile228Thr
BFLF1	Late lytic phase (L)	Gly16Arg
		Ile35Thr

		Ala36Thr
		Asp44Glu
		Val246Ile
		Leu250Gln
		Leu45Phe
BFRF1A	Late lytic phase (L)	Leu59Ile
		Pro126Ser
BFRF1	Late lytic phase (L)	Asp145Asn
		Pro212His
		Gly213Ser
		His224Leu
		Ala313Thr
		Phe332Val
BFRF2	Late lytic phase (L)	Arg7His
		Cys17Gly
		Pro24Leu
		Val26Gly
		Ser27Phe
		Gln34Pro
		Gly40Glu
		Val129Ala
		Val130Phe
		Asp169Ala
		Arg364Trp
		Cys448Phe
		Arg462Gln
		Arg479Cys
		Ala536Val
		His541Tyr

BFRF3	Late lytic phase (L)	His575Tyr
		His165Gln
		Pro169Leu
BPLF1	Late lytic phase (L)	Thr12Pro
		Val18MetI
		Val26Met
		Gly285Asp
		Ser654Ala
		Val657Ile
		Ala681Val
		Glu688Asp
		Ser796Ala
		Lys847Glu
		Thr862Met
		Val866Ile
		Asp885Gly
		Leu891Met
		Ala909Gly
		Gly910Glu
		Glu911Gly
		Asn919His
		Glu936Asp
		Asp1168Asn
		Ile1203Val
		Ile1315Val
		Thr1337Met
		Gln1428Arg
		Asp1473Glu
		Arg1512Gln

Thr1534Met
Asp1535Glu
Pro1548His
Asp1579Asn
Asp1579Gly
Ala1626Ser
Ile1634Thr
Ser1674Leu
Arg1752Cys
Ala1799Thr
Ala1823Val
Gly1824Arg
Val1879Phe
Ser1937Leu
Tyr1942Phe
Ala2127Ser
Arg2167Leu
Met2177Ile
Ala2179Thr
Asn2246Asp
Glu2249Lys
Arg2256Cys
Arg2390Lys
Leu2451Gln
Asp2526Asn
His2551Asn
Glu2552Lys
Pro2557Gln
Arg2602Trp

		Lys2616Glu
		Ala2640Ser
		Arg2655Leu
		Val2657Ala
		Gly2671Val
		Arg2673Gln
		Ser2720Arg
		Ser2763Thr
		Pro2765Ser
		Gly2778Asp
		Ala2932Thr
		Glu2983Asp
		Thr2996Ala
		Arg3000Gln
		Gln3036Glu
		Gly3085Ser
BOLF1	Late lytic phase (L)	Ser10Gly
		Glu144Asp
		Ala151Thr
		Ala151Glu
		Ala203Val
		Ala480Thr
		Glu686Asp
		Glu688Asp
		Ala691Ser
		Glu697Asp
		Arg795Lys
		Arg809Gly
		Ser833Tyr

		His896Arg
		Ile968Thr
		Pro975Leu
		Leu991Met
		Ser996Arg
		Pro1022Ser
		Pro1035Ser
		Val1040Asp
		Ala1083Val
		Ala1236Thr
BORF1	Late lytic phase (L)	Thr36Ala
		Val247Gly
		Asn339Ile
BORF2	Late lytic phase (L)	Asp317Ala
		Ala420Ser
		Val450Ile
		Asp734Asn
		Pro763Leu
		Ala773Thr
BLLF3	Late lytic phase (L)	Gly189Ser
		Leu235Ile
		Thr247Ala
		Ser253Gly
BLRF1	Late lytic phase (L)	Ile81Val
BLRF2	Late lytic phase (L)	Pro134Leu
		Ser137Pro
BLLF1	Late lytic phase (L)	Glu201Gln
		Asp215Glu
		Ala252Val

		Leu434Ser
		Ala440Asp
		Asp441Ala
		Ile710Thr
		Ser711Asn
		Pro742Leu
		Gln754Lys
		Gly805Asp
		Pro812Thr
		Ser825Pro
BLLF2	Late lytic phase (L)	Lys827Glu
		Arg887His
		His8Arg
		Arg29Trp
		Val92Met
		Asp98Glu
BKRF2	Late lytic phase (L)	Arg103Gln
		Gln134Leu
		Ala3Thr
		Ile10Thr
BKRF3	Late lytic phase (L)	Gly102Ser
		Thr113Ser
		Thr63Ile
BKRF4	Late lytic phase (L)	Val105Ile
		Ala120Val
		Ser20Leu
		Val47Ala
		Glu61del
		Asp88Asn

		Ser89Ala
		Pro146Leu
		Gln162Pro
		Gly169Val
		His171Asn
BBRF1	Late lytic phase (L)	Asp95Asn
		Pro157His
		Thr189Met
		Lys196Arg
		Ile199Met
		Cys229Gly
		Ser558Gly
		Ala577Val
		Thr594Ala
BBRF2	Late lytic phase (L)	Ile56Val
		Val76Ile
BBRF3	Late lytic phase (L)	Val15Phe
		Thr17Met
		Val204Ile
		Ile249Val
		Leu266Met
		Ile312Met
		Val389Ile
BGLF2	Late lytic phase (L)	Val46Ala
		Arg224His
		Pro271His
BGLF1	Late lytic phase (L)	Asp85Glu
		Arg125Ser
		Ile149Met

		Val171Ala
		Ala200Val
		Ser226Pro
		Gly255Ser
		Leu264Pro
		Thr407del
		Pro442Ser
BDLF4	Late lytic phase (L)	Arg22Cys
		Asp56Asn
		Tyr78Asp
		Lys114Gln
		Glu135Lys
		Glu141Asp
		Ala166Gly
BDLF3	Late lytic phase (L)	Met12Leu
		Ala209Thr
BDLF2	Late lytic phase (L)	Pro74Ser
		Pro79Ala
		Thr105Ala
		His269Leu
		Ala277Val
		Arg311Gln
		Cys357Ser
BDLF1	Late lytic phase (L)	Leu51Ile
BcLF1	Late lytic phase (L)	Val99Ile
		Phe207Val
		Pro353Thr
		Val515Met
		Ser651Ala

		Ala897Thr
		Val906Ile
		Pro912Gln
		Ala915Thr
		Asn974Lys
		Val1013Ala
		Arg1150His
BXLF2	Late lytic phase (L)	Ser34Leu
		Leu160Met
		Val197Ile
		Thr487Lys
		Thr500Met
		Ala505Val
BXLF1	Late lytic phase (L)	Asp21Val
		Arg133Leu
		Arg194His
		Met266Thr
		Arg320Leu
		Thr431Ala
		Ser464Arg
BXRF1	Late lytic phase (L)	Tyr87Ser
BVRF1	Late lytic phase (L)	Ala332Ser
		Gly338Val
		Gly341Cys
		Met416Val
		Pro426Ser
		Ala450Val
		Gln516Glu
BVLF1	Late lytic phase (L)	Ala209Val

		Glu247Asp
BVRF2	Late lytic phase (L)	Ala65Val
		Asp77Val
		Val105Leu
		Ala337Ser
		Ser482Pro
		Ala540Thr
		Pro545Leu
		His560Pro
		Lys564Gln
BdRF1	Late lytic phase (L)	Ala337Ser
		Ser482Pro
		Ala540Thr
		Pro545Leu
		His560Pro
		Lys564Gln
BILF2	Late lytic phase (L)	Val28Ala
		Val68Leu
		Thr75Pro
		Leu119Met
		His142Gln
		Ala143Cys
		Ser168Phe
LF3	Late lytic phase (L)	Lys915Thr
		Phe924Leu
LF2	Late lytic phase (L)	Ser36Leu
		Met38Thr
		Val324Ile
		Val336Ile

LF1	Late lytic phase (L)	Thr428Ser
		Asp6Gly
		Gln8His
		Ala120Ser
		Gln147Lys
		Ser183Tyr
		Asn185Lys
		Val206Ala
		Gly234Glu
		Ala241Thr
		Arg243Gln
		Ser275Thr
		Arg291Lys
		Pro306Leu
		Ser314Phe
		Leu324Met
		Pro446Leu
BILF1	Late lytic phase (L)	Ala6Val
		Asn22Asp
BALF4	Late lytic phase (L)	Thr40Ala
		Asp128Glu
		Thr153Ala
		Lys241Glu
		Lys241Thr
		Ala743Val
		Thr792Ala
BALF3	Late lytic phase (L)	Thr23Ala
		Glu193Val
		Gly317Trp

		Glu399Asp
		Ser405Arg
BARF0	Latent phase and lytic phase	Ala5Thr
		Cys30Arg
		Gly34Arg
		Gly61Arg
		Ala162Thr
		Leu198Val
		Gln214Glu
		Pro217Ser
		Arg257Cys
		Gly263Ser
		Val264Ile
		Ala299Ser
		Ala313Thr
		Leu323Phe
		Arg375Ser
		Ala377Pro
		Ala377Ser
		Glu193Val

Figures

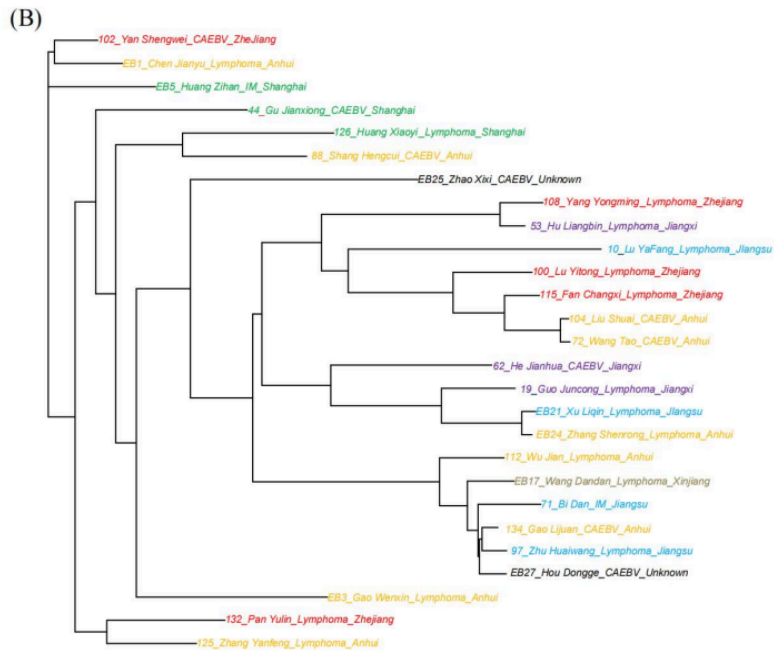
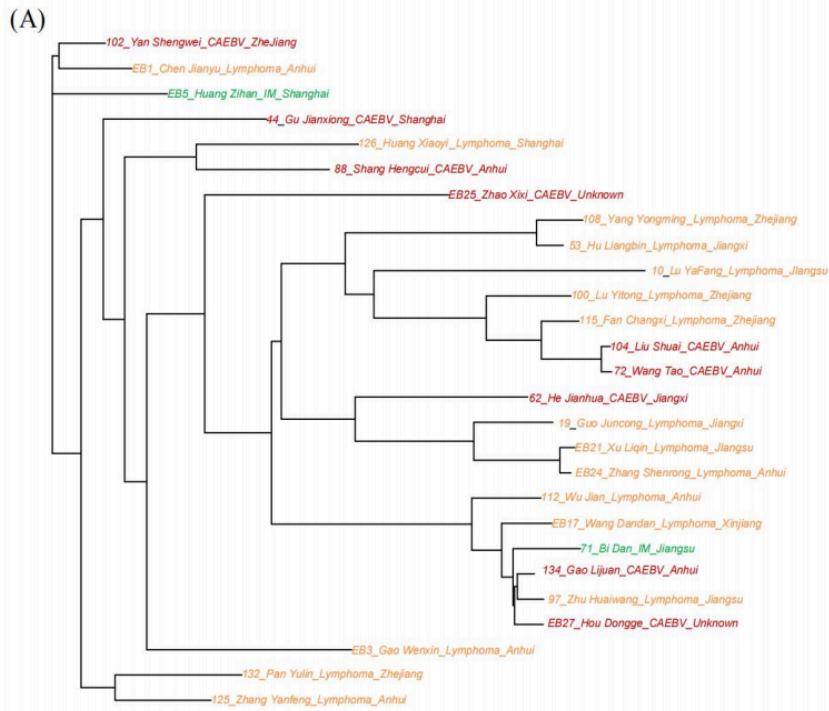


Figure 1

Phylogenetic map of patients with different types of EBV-associated diseases (A) and from different regions (B).

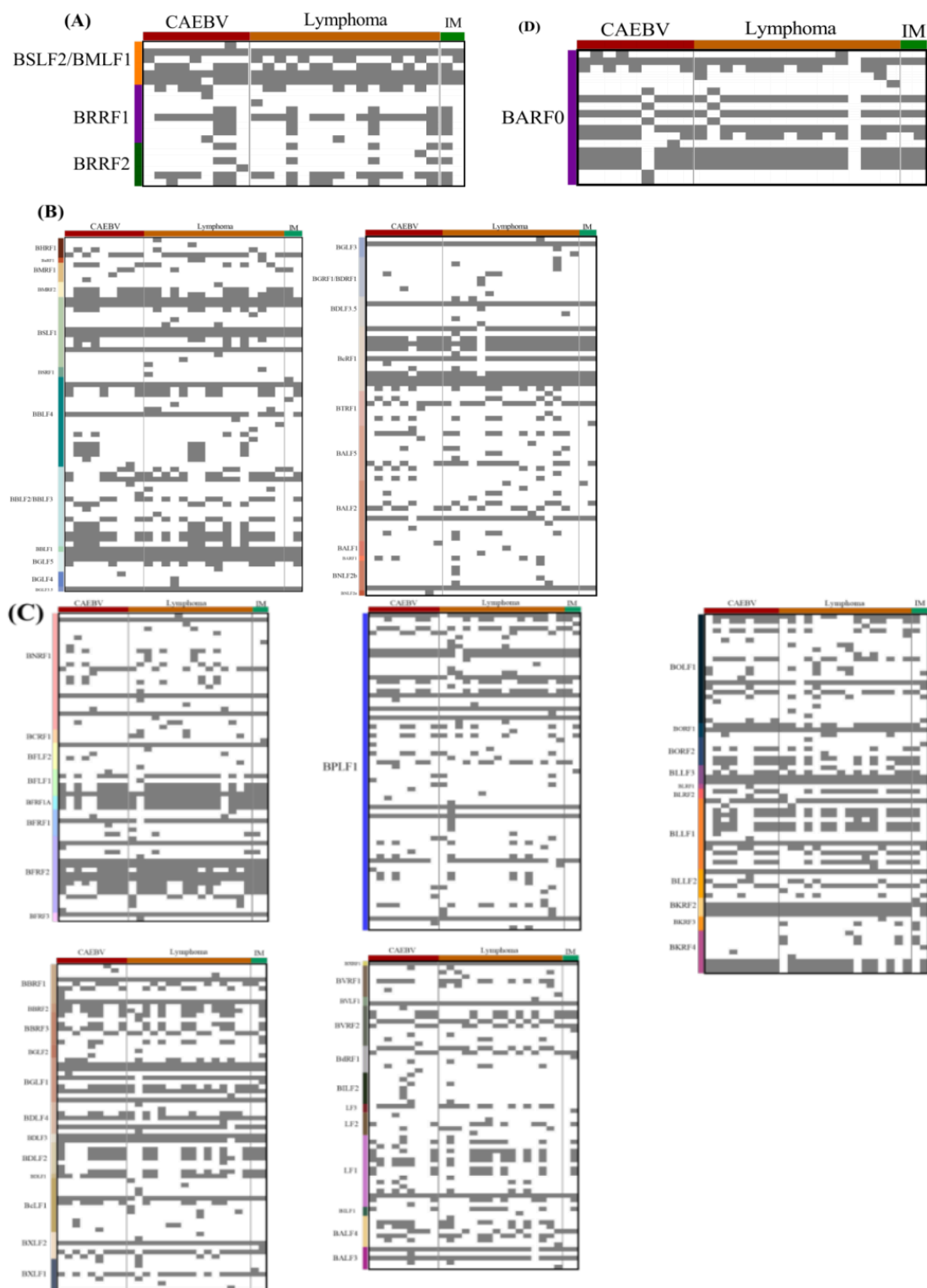


Figure 2

Mutation sites of EBV proteins based on EBV-LPDs subtypes. (A) proteins expressed in immediately early phase; (B) proteins expressed in early phase; (C) proteins expressed in late phase; (D) proteins expressed in both early and late phase.

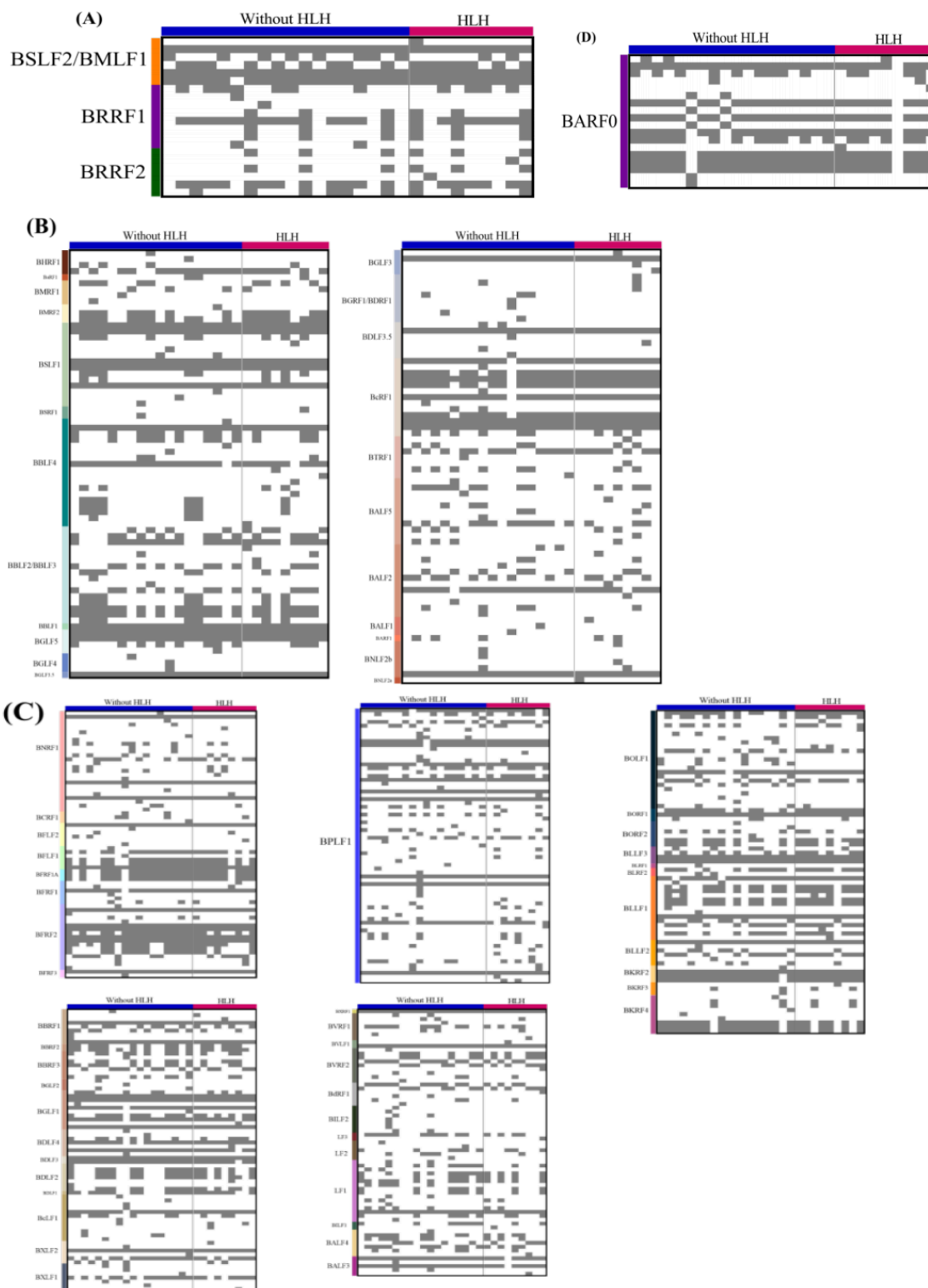


Figure 3

Mutation sites of EBV proteins based on comorbidity with HLH. (A) proteins expressed in immediately early phase; (B) proteins expressed in early phase; (C) proteins expressed in late phase; (D) proteins expressed in both early and late phase.

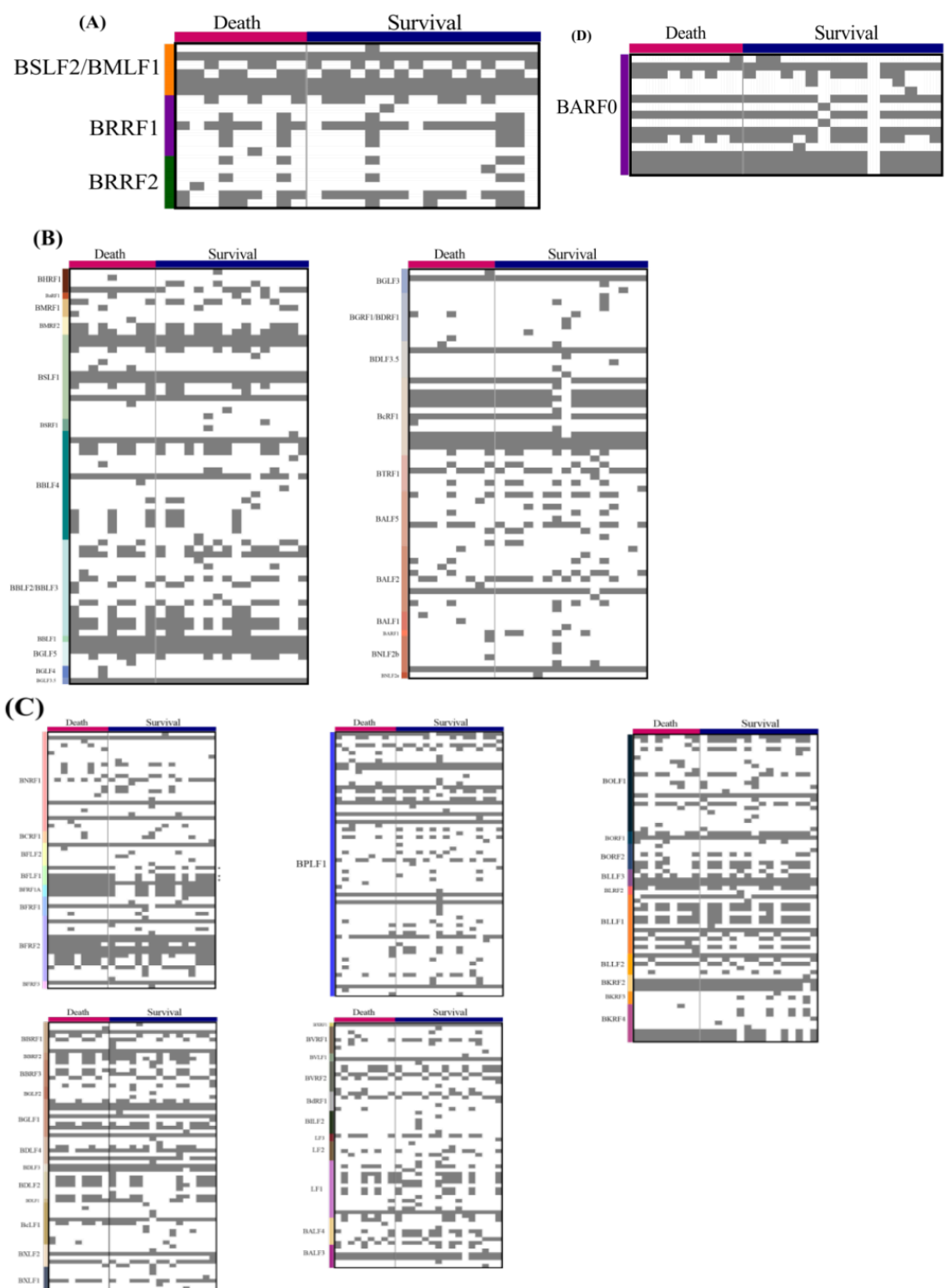


Figure 4

Mutation sites of EBV proteins based on clinical outcomes. (A) proteins expressed in immediately early phase; (B) proteins expressed in early phase; (C) proteins expressed in late phase; (D) proteins expressed in both early and late phase.

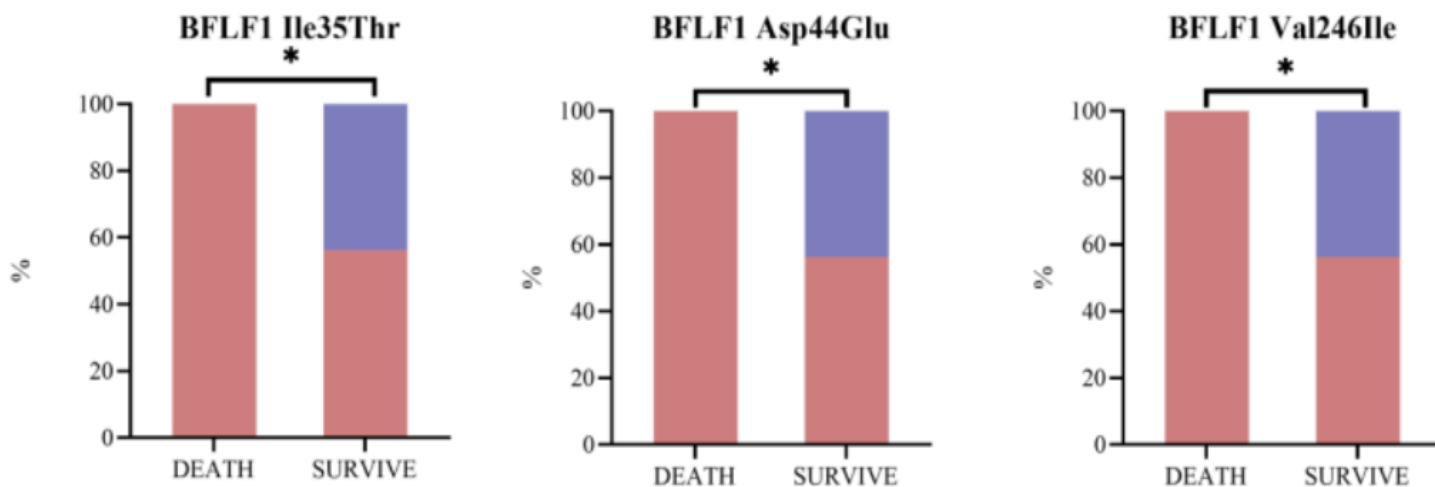


Figure 5

The frequency of mutant amino acid in BFLF1 between dead and live patients.

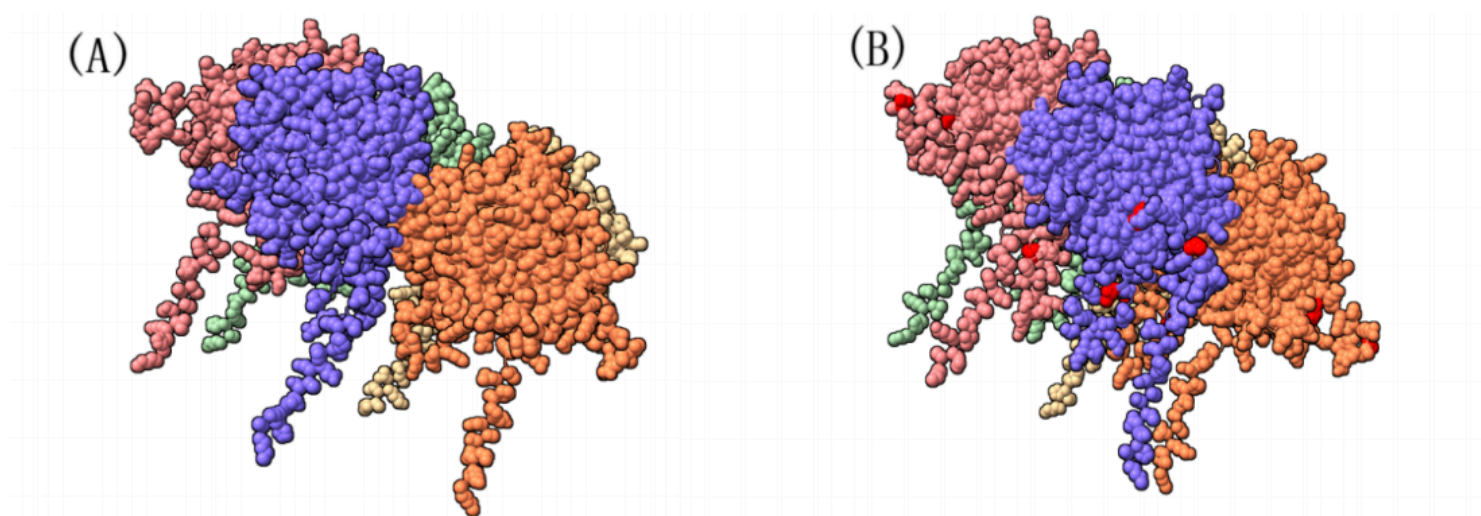


Figure 6

Structure of BFLF1 predicted by AlphaFold3. (A) wildtype BFLF1 subunit; (B) mutant BFLF1 subunit. *Red spots highlight the mutant amino acids.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementdata.docx](#)