

# Genome-wide analysis of Epstein-Barr virus (EBV) isolated from EBV-associated gastric carcinoma (EBVaGC)

Ying Liu<sup>1,\*</sup>, Wenjun Yang<sup>2</sup>, Yaqi Pan<sup>1</sup>, Jiafu Ji<sup>3</sup>, Zheming Lu<sup>1,\*</sup>, Yang Ke<sup>1</sup>

<sup>1</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Laboratory of Genetics, Peking University Cancer Hospital and Institute, Haidian, Beijing, China

<sup>2</sup>Key Laboratory of Reproduction and Heredity of Ningxia Region, Medical Oncology Department of General Hospital, Ningxia Medical University, Yinchuan, Ningxia, China

<sup>3</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Gastrointestinal Surgery, Peking University Cancer Hospital and Institute, Haidian, Beijing, China

\*These authors contributed equally to this work

**Correspondence to:** Zheming Lu, e-mail: zheminglu@163.com  
Yang Ke, e-mail: keyang@bjmu.edu.cn

**Keywords:** Epstein-Barr virus, gastric carcinoma, EBVaGC, next-generation sequencing

**Received:** July 29, 2015

**Accepted:** November 26, 2015

**Published:** December 24, 2015

## ABSTRACT

Epstein-Barr virus (EBV) is linked to the development of a variety of malignancies, including EBV-associated gastric carcinoma (EBVaGC). In this study, EBVaGC was detected in 15 (7.3%) of 206 GC cases. To identify the EBV genomic variation, EBV genomic sequences isolated from 9 EBVaGC biopsy specimens were successfully retrieved, designated EBVaGC1 to EBVaGC9. By comparative analysis of these strains with another 6 completely sequenced EBV strains, EBV-wild type, B95-8, AG876, GD1, GD2, and HKNPC1, it was demonstrated that EBVaGC1 to 9 were most closely related to the GD1 strain. Phylogenetic analysis of the GC biopsy specimen-derived EBV (GC-EBV) genomes was subsequently performed to assess their genomic diversity and it exhibited the greatest divergence from the type 2 strain, AG876. Compared with the reference EBV strain GD1, they harbored 961 variations in total, including 919 substitutions, 23 insertions, and 19 deletions. Single nucleotide polymorphism (SNP) density varied substantially across all known open reading frames and was highest in latency-associated genes. Moreover, we identified 2 interstrain recombinants at the *EBNA1* locus, which provided a further mechanism for the generation of diversity. Some T-cell epitope sequences in *EBNA1* and *LMP2A* genes showed extensive variation across strains, which implied their importance in the development of vaccines and T-cell therapy. In conclusion, we reported the first genome-wide view of sequence variation of EBV isolated from primary EBVaGC biopsy specimens, which might serve as an effective method for further understanding the genomic variations contribute to EBVaGC carcinogenesis and treatment.

## INTRODUCTION

Epstein-Barr virus (EBV)-associated gastric carcinoma (EBVaGC) is a distinct subset of gastric carcinoma, accounting for about 10% of total gastric carcinomas [1–4]. The monoclonal presence of the virus was uniformly distributed in malignant cells of EBV-positive tumors but not observed in the surrounding normal epithelial cells and dysplastic cells, providing strong evidence to support the role of EBV as an etiologic agent [5]. Furthermore, the presence of EBV in tumor

cells represents a potential “tumor-specific” target for therapeutic approaches. However, the exact role of EBV in the development and progression of this specific type of gastric carcinoma is not yet clear. Hence, understanding the pattern of EBV sequence variation is important for knowing whether there is a disease-related strain-specific or geographic regional variation of EBV strain.

Progress has been made in understanding EBV genome polymorphisms in EBVaGC, but studies conducted to date have predominantly targeted specific fragments of the virus, or have focused on different regions

of EBV, thus limiting our ability to understand the full spectrum of diversity existent within the EBV genome. Up to now, 19 complete or partial EBV genomes, EBV-wild type, B95-8, AG876, Akata, Mutu, K4413-Mi, K4123-Mi, C666-1, GD1, GD2, and HKNPC1 to 9, have been reported. The prototypic type 1 EBV strain B95-8 was the first complete genome sequenced from an individual with infectious mononucleosis [6]. A more representative type 1 EBV reference genome, human herpesvirus 4 complete wild type genome, was constructed by using B95-8 as the backbone with an 11-kb deletion segment provided by the Raji sequences [7]. AG876 was the unique complete type 2 EBV sequence from a Ghanaian case of Burkitt's lymphoma [8]. Akata and Mutu were sequenced from Burkitt's lymphoma cell lines from a Japanese patient and a Kenyan patient, respectively [9]. K4413-Mi and K4123-Mi were sequenced from immortalized human B lymphocyte cell lines [10]. C666-1 was derived from a nasopharyngeal carcinoma (NPC) xenograft cell line of southern Chinese origin [11, 12]. GD1, GD2, and HKNPC1 to 9 were isolated from NPC patients [13-16].

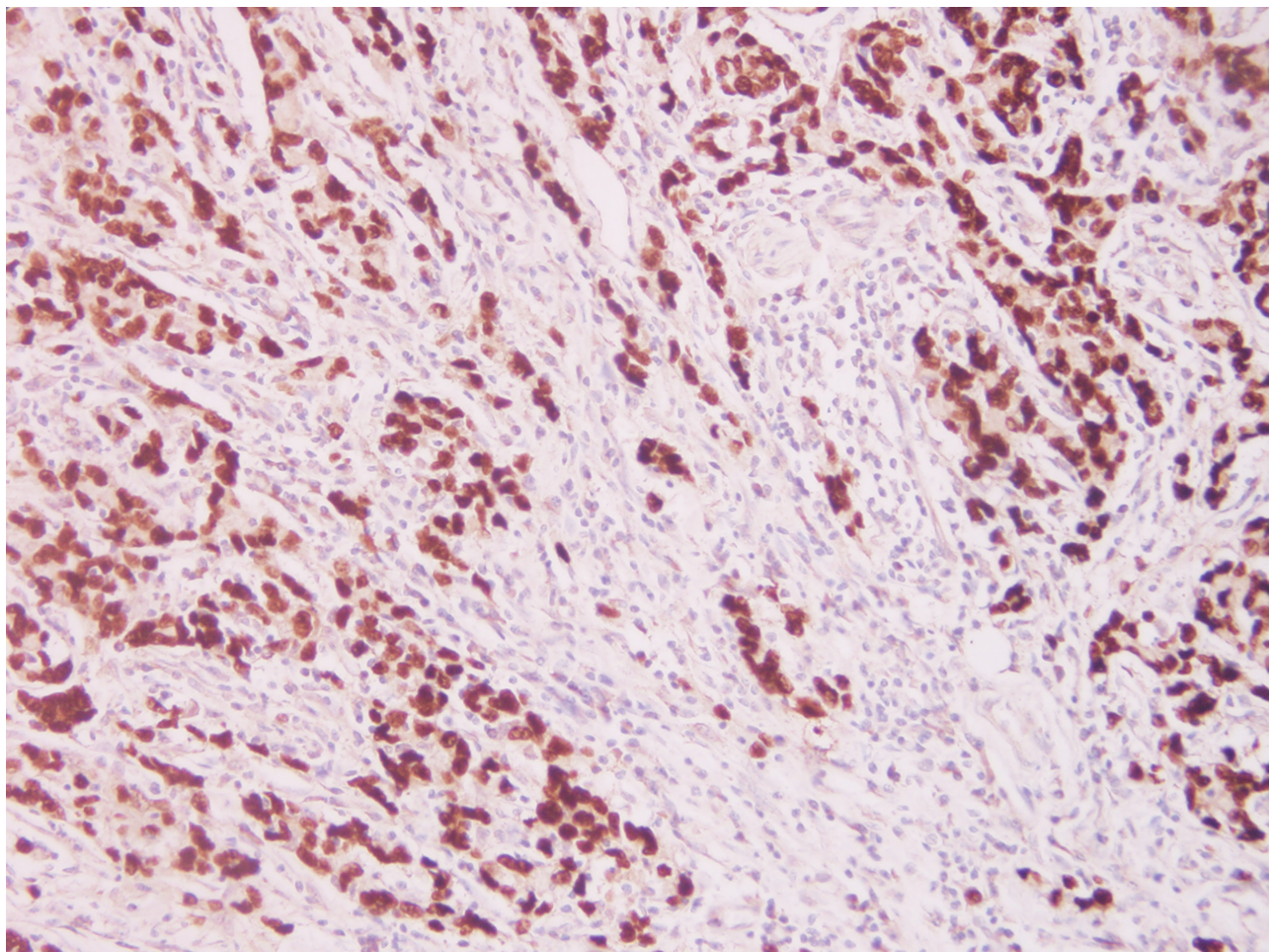
However, no complete EBV genome sequence derived from gastric carcinoma is yet available.

In this study, we used the method of target enrichment of EBV DNA by hybridization, followed by next-generation sequencing, *de novo* assembly, and joining of contigs to yield complete EBV genomes. Genomic diversity of EBV genomes isolated from primary gastric carcinoma biopsy specimens was evaluated by mutation and phylogenetic analysis.

## RESULTS

### Prevalence of EBVaGC and association with clinicopathological characteristics

To determine the prevalence of EBVaGC, 206 GC specimens were assessed using EBER *in situ* hybridization (ISH). Overall, 15 (7.3%) out of the 206 cases exhibited EBER ISH positive signals (Figure 1) restricted only to the nuclei of carcinoma cells and were considered as EBVaGC. Expression of EBER was not detected in



**Figure 1: Photomicrographs of Epstein-Barr virus (EBV) expression in gastric carcinoma.** EBV-encoded RNA 1 (EBER1) *in situ* hybridization in a gastric carcinoma section, reveals specific EBER1 transcripts (dark brown) in the nuclei of the tumor cells (Original magnification  $\times 20$ ).

**Table 1: Correlation between EBVaGC and clinicopathological parameters**

| Clinical Characteristics | N   | EBER expression |              | P            |
|--------------------------|-----|-----------------|--------------|--------------|
|                          |     | Negative (%)    | Positive (%) |              |
| <b>Gender</b>            |     |                 |              | 0.541        |
| Male                     | 153 | 143 (93.5)      | 10(6.5)      |              |
| Female                   | 53  | 48 (90.6)       | 5 (9.4)      |              |
| <b>Age</b>               |     |                 |              | <b>0.013</b> |
| ≤ 60                     | 111 | 98 (88.3)       | 13 (11.7)    |              |
| > 60                     | 95  | 93 (97.9)       | 2 (2.1)      |              |
| <b>Differentiation</b>   |     |                 |              | 1.000        |
| Poor                     | 158 | 146 (92.4)      | 12 (7.6)     |              |
| Well+Moderate            | 48  | 45 (93.7)       | 3 (6.3)      |              |
| <b>TNM</b>               |     |                 |              | 0.785        |
| I-II                     | 82  | 77 (93.9)       | 5 (6.1)      |              |
| III-IV                   | 124 | 114 (91.9)      | 10 (8.1)     |              |
| <b>Lymph node</b>        |     |                 |              | 0.203        |
| 0                        | 43  | 42 (97.7)       | 1 (2.3)      |              |
| 1                        | 163 | 149 (91.4)      | 14 (8.6)     |              |
| <b>Anatomical site</b>   |     |                 |              | 0.535        |
| Cardia                   | 43  | 39 (90.7)       | 4 (9.3)      |              |
| Body                     | 33  | 29 (87.9)       | 4 (12.1)     |              |
| Antrum                   | 128 | 121 (94.5)      | 7 (5.5)      |              |
| Gastric stump            | 2   | 2 (100.0)       | 0 (0.0)      |              |
| <b>Venous invasion</b>   |     |                 |              | 0.176        |
| Negative                 | 77  | 74 (96.1)       | 3 (3.9)      |              |
| Positive                 | 129 | 117 (90.7)      | 12 (9.3)     |              |

Abbreviations: EBVaGC, Epstein-Barr virus associated gastric carcinoma; TNM, the tumor, node, and metastasis classification.

corresponding non-neoplastic gastric mucosal, or stromal cells (endothelial cells and fibroblasts), or infiltrating inflammatory cells within the tumor sections.

Correlation between EBER expression and the clinicopathologic features of the study subjects were summarized in Table 1. We found a significant association between EBER expression and age at diagnosis ( $P = 0.013$ ). 13 out of the 15 patients with EBVaGC were 60 years or younger. No significant difference was found in gender, anatomical site, histology, TNM stage, nodal status, or venous invasion of the tumor.

### Summary of the sequencing data

In the current study, 9 EBVaGC cases were successfully performed for genome-wide analysis of EBV. The DNA sequences of these 9 GC-EBV genomes

were compared against 6 available EBV genomes, including EBV-wild type, B95-8, AG876, GD1, GD2, and HKNPC1. According to the value of coverage of the target region, all the DNA sequence generated from the GC-EBV strains resembled that of GD1. Thus, to facilitate studies on the whole viral genome, GD1 was used as the reference sequence for most of the subsequent analyses in this study. The coverage of target region of EBVaGC1 to 9 genomes mapped to the reference EBV-GD1 ranged from 93.2% to 98.4%. The fraction of effective bases on target, which was calculated by dividing the effective sequence on target by the total effective yield, varied from 22.3% (66.39/297.51Mb, EBVaGC6) to 37.9% (77.72/204.93Mb, EBVaGC5). The average sequencing depth on target of these 9 GC-EBV genomes was 383-fold, ranging from 212 to 656-fold. Details of the sequencing data are listed in Table S1 in the Supplementary Materials.

## Assembly of EBVaGC1 to 9 genomes

*De novo* assembly was performed for 9 sequenced GC-EBV strains. The number of contigs ranged from 17 (EBVaGC3) to 57 (EBVaGC1). N50 sizes of contigs ranged from 7770 bp (EBVaGC1) to 19,812 bp (EBVaGC8). The longest contigs were approximately 46 kb in length for the majority of the 9 samples. GC contents were all approximately 57%. Details of the contig data are listed in Table S2 in the Supplementary Materials. The gaps between the contigs were further linked up either by Sanger sequencing or tracts of “N” with length estimated based on the EBV reference GD1 (AY961628.3).

All 9 GC-EBV genomes were successfully sequenced. The sequences of these 9 GC-EBV genomes, designated EBVaGC1 to EBVaGC9, were then determined. The genome sizes, estimated based on the reference EBV GD1 sequence, were as follows: EBVaGC1 (171,957 bp), EBVaGC2 (171,759 bp), EBVaGC3 (171,615 bp), EBVaGC4 (171,661 bp), EBVaGC5 (171,759 bp), EBVaGC6 (171,612 bp), EBVaGC7 (171,673 bp), EBVaGC8 (171,664 bp), and EBVaGC9 (171,858 bp).

## Mutation analysis of the GC-EBV genomes

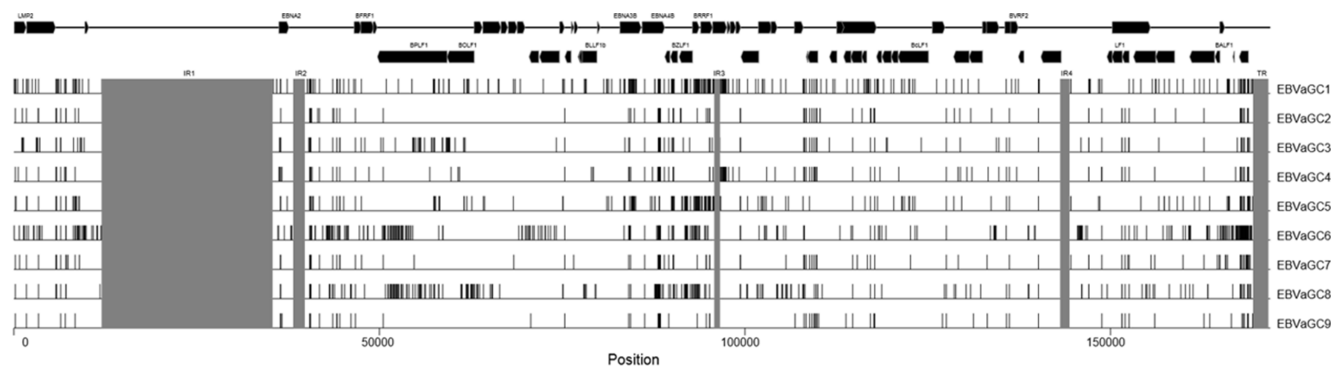
961 variations were observed in the EBVaGC1 to 9 genomes in comparison to the reference EBV GD1, including 919 substitutions, 23 insertions, and 19 deletions. A total of 961 variations, 663 substitutions, 9 insertions, and 7 deletions were located in the coding regions of the genomes, while the remaining variations were found in the noncoding regions. The detailed data for mutations in individual genomes are listed in Table S3 in the Supplementary Materials. The number of variations and their percentages, in relation to the size of the genomes, were as follows: EBVaGC1 (324, 0.188%), EBVaGC2 (89, 0.052%), EBVaGC3 (147, 0.085%), EBVaGC4 (135, 0.079%), EBVaGC5 (202, 0.118%), EBVaGC6 (354, 0.206%), EBVaGC7 (103, 0.059%), EBVaGC8 (241, 0.140%), and EBVaGC9 (90, 0.052%).

Genetic variations for all 9 GC-EBV genomes relative to the reference EBV GD1 strain are illustrated in Figure 2. The pattern of distribution of variations among the genomes of EBVaGC2, 4, 7 and 9 was rather similar, showing relative fewer variations. Whereas the genomes of EBVaGC1, 3, 5, 6, and 8 showed many more variations, indicating that they might harbor more mutations.

Overall, EBVaGC1 and 6 had higher numbers of nonsynonymous mutations (128 and 100, respectively), followed by EBVaGC5 (74), 8 (72), 4 (69), 3 (51), 7 (36), 2 (35), and 9 (31) (Figure 3 and Table S6). Nonsynonymous mutations in the 9 GC-EBV genomes were classified into 9 categories, according to the function of EBV-encoded proteins defined by Tarbouriech *et al.* [17]. Both latent genes and genes encoding tegument proteins in all 9 GC-EBV genomes were found to harbor the majority of nonsynonymous mutations, accounting for 58.4% (EBVaGC8) to 84.3% (EBVaGC3) of all nonsynonymous mutations detected for each genome. The remaining nonsynonymous mutations were located in membrane glycoproteins, capsid, and proteins for replication, transcription, nucleotide metabolism, or packaging, or in proteins of unknown function.

## Phylogenetic analysis of the GC-EBV genomes

The phylogenetic analysis was conducted based on multiple nucleotide sequence alignments of whole genomes and *EBNA1*, *EBNA2*, *LMP1*, *LMP2A*, and *BZLF1* genes, all published 19 EBV genomes and the newly available 9 GC-EBV genomes in this study included (see Figure 4 and Figure S1 in the Supplementary Materials). Analyses on whole EBV genomes indicated that the 9 GC-EBV strains were closely related to all Asian-derived EBV strains, including HKNPC1 to 9, GD1 and 2, C666-1 (Chinese strain), and Akata (Japanese strain), distant to the non-Asian strains, AG876, B95-8, Mutu, K4413-Mi, and K4123-Mi. Similar results were observed when the sequences of *LMP1* gene were compared. Additionally, analysis on *LMP-1* genes of EBVaGC1 to 9 revealed that



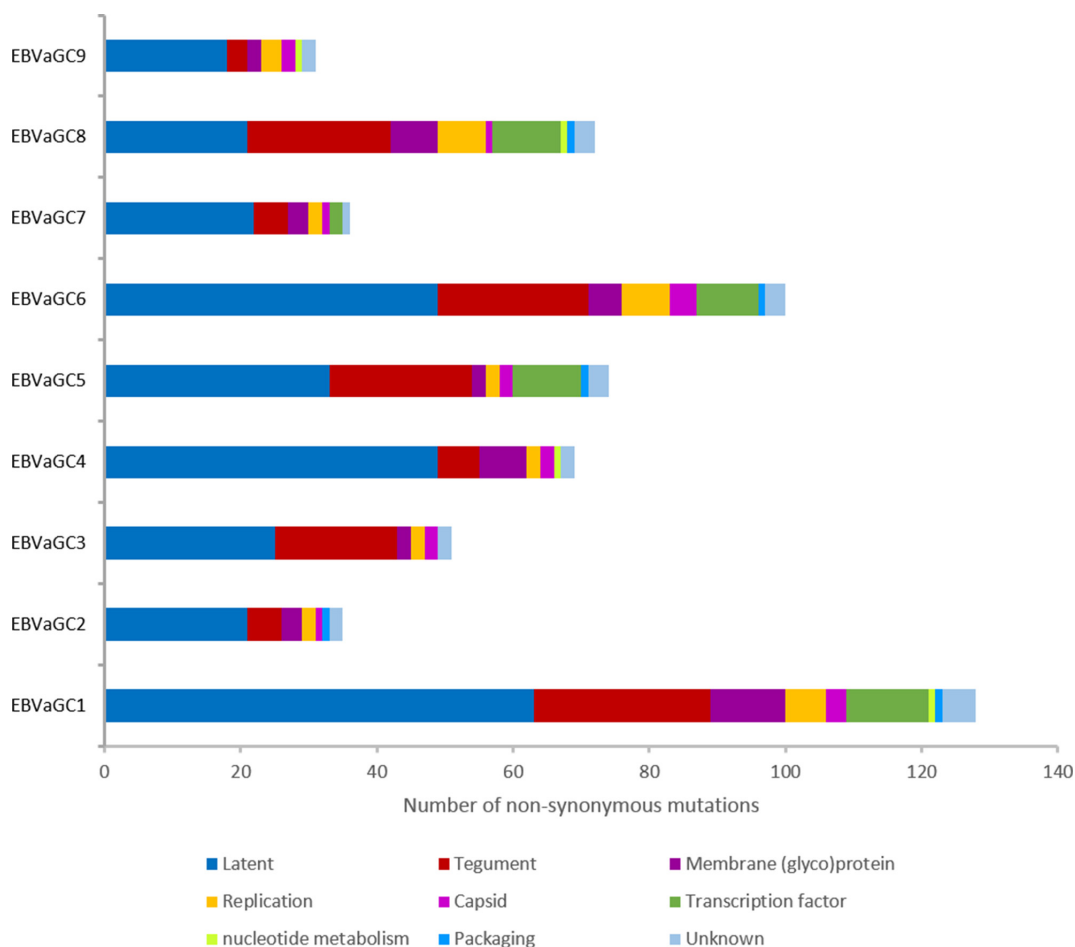
**Figure 2: Genetic variations among GC-EBV strains.** Mutations of the GC-EBV strains relative to the reference EBV strain (GD1). Mutations in internal repeats and terminal repeats are disregarded, and the regions are shaded in gray. Rightward and leftward open reading frames of EBV are overlaid on top of the mutations. Abbreviations: GC-EBV, gastric carcinoma biopsy specimen-derived Epstein-Barr virus.

EBVaGC6 was distant to other GC-EBV strains, indicating that EBVaGC6 might harbor many mutations that were not present in the other GC-EBV strains. Neighbor-joining trees derived from the sequences of *EBNA2* gene showed that all 9 GC-EBV genomes are type 1 viruses, clustered in a branch with other type 1 EBV strains, distant to the only type 2 EBV strain, AG876. Analyses of the nucleotide sequences of the *BZLF1* gene illustrated that EBVaGC2, 3, 4, 6, 7, and 9 were more closely related. Phylogenetic analysis of nucleotide sequences of the *EBNA1* gene showed that EBVaGC5 was located in a branch distinct from that of the EBVaGC1 and 4 and the other GC-EBV strains.

### Nonsynonymous mutations in EBV nuclear antigen 1 (EBNA1) and its interstrain recombinant

The sequences of *EBNA1* gene in EBVaGC1 to 9 ranged from 1794 bp to 2058 bp in length, mostly due to variation in the length of the gly-ala repeat, which was disregarded in subsequent mutation analysis.

Figure S2 in the Supplementary Materials illustrates the nonsynonymous mutations that resulted in amino acid changes in EBNA1 in at least 1 of the 9 GC-EBV strains relative to the reference GD1 EBNA1 sequence. Based on the signature codon 487 as well as particular amino acid alterations in other sites in comparison to that of the B95.8-derived virus, as suggested by Bhatia *et al.* and Gutierrez *et al.* [18, 19], 5 EBNA1 subtypes were determined, including P-ala, P-thr, V-pro, V-leu, and V-val. In the current study, in accordance with most studies, we also used AA 487 as the signature residue combined with common changes across codons 377–641 to classify the sequence variation patterns. The sequences with identical consensus mutations were arranged into 1 group. As a result, 2 patterns of variations were observed. The most common pattern was V-val subtype which was observed in EBVaGC2, 3, 5, 6, 7, 8, and 9, and also in GD1 strain. In the V-val group, all the sequence changes were consistent across codons 377–641. The other pattern was found in EBVaGC1 and 4. The consensus sequence in this pattern carried more substitutions at codons 377 to 641 than the P-thr subtype, thus we defined this pattern as P-thr variant



**Figure 3: Number of nonsynonymous mutations contained in the 9 categories of EBV-encoded proteins.** The majority of the amino acid changes are located in latent proteins (blue) and tegument proteins (red) in all 9 GC-EBV strains. Abbreviations: GC-EBV, gastric carcinoma biopsy specimen-derived Epstein-Barr virus.

(P-thrV). Unlike the V-val group, EBVaGC1 carried 2 additional coding mutations at residue 479 (Glu→Gln) and 584 (Met→Ile), apart from the consensus mutations shared with EBVaGC4.

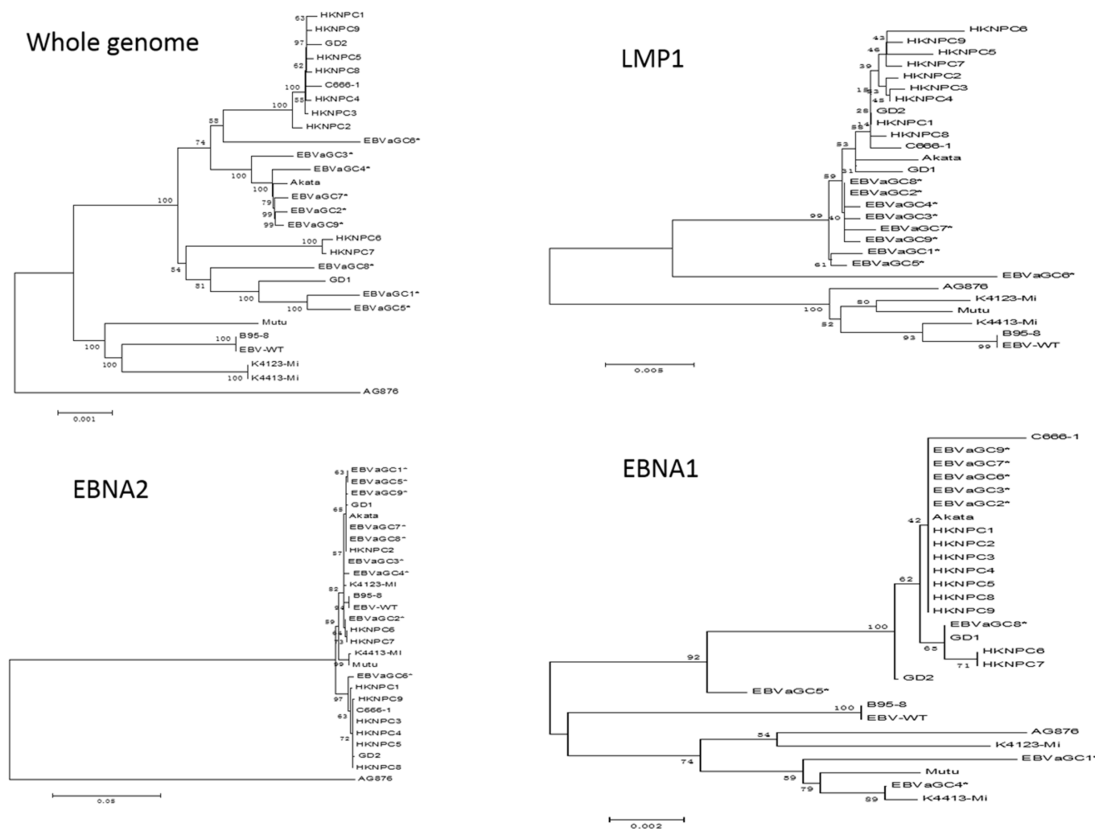
In addition to changes in the C-terminus, EBNA-1 has variation in the N-terminus at codons 16, 18, 20, 24, 27 and 85 (Figure S2). EBNA-1 N-terminus changes, though generally linked to variations in the C-terminus, have revealed additional variants that were not evident by evaluating the C-terminus alone. Of interest, EBVaGC5 from the V-val subtype and EBVaGC1 from the P-thrV subtype had the consensus sequences at codons 16 to 364, identical to that of the B95-8 strain. Though EBVaGC1 and EBVaGC4 were defined as the P-thrV subtype, the sequences at codons 16 to 85 of EBVaGC1 were different from that of EBVaGC4, which was identical to that of AG876. Thus, it is necessary to identify both N- and C-terminal sequences to gain an authentic picture of EBNA1 subtype polymorphism.

Full length *EBNA1* sequences obtained from EBVaGC1 to 9 were aligned with genotype GD1, B95-8, AG876 strains in Figure S2. The data showed that the most predominant EBNA1 subtypes among the isolates were the EBVaGC2, 3, 6, 7, 8, and 9. However, a small number of isolates displayed a different combination, for

example, EBVaGC5 may arise from recombination of EBVaGC1 and EBVaGC2, and EBVaGC4 may arise from recombination of AG876 and EBVaGC1 (Figure 5 and Figure S3).

### Amino acid changes in CD4<sup>+</sup> and CD8<sup>+</sup> T-cell epitopes of EBNA1 and LMP2A

EBVaGC shows EBV type I latency neoplasm, in which EBNA1 is expressed in 100% and LMP2A in about half of EBVaGC cases, respectively [20–22]. The extensive genetic diversity within EBNA1 and LMP2A can have a significant impact on immune recognition of these antigens. A comprehensive investigation into the immunological impact of EBNA1 and LMP2A sequence polymorphism is essential for optimizing the adoptive cell therapy approaches and vaccine development strategies. According to the epitopes specific for both CD4<sup>+</sup> and CD8<sup>+</sup> T cells defined and reviewed in previous publications [23–25], amino acid changes were found in 5 CD8<sup>+</sup> epitopes and 28 CD4<sup>+</sup> epitopes of EBNA1, 14 CD8<sup>+</sup> epitopes and 8 CD4<sup>+</sup> epitopes of LMP2A. Some of the nonsynonymous mutations were affecting multiple epitopes. For example, a G-to-A substitution at coordinate 97120 (NC\_007605) resulted in the change of residue



**Figure 4: Phylogenetic trees of the whole EBV genomes and nucleotide sequences of *EBNA1*, *EBNA2*, and *LMP1* genes.** Phylogenetic analyses were conducted using the neighbor-joining (NJ) algorithm implemented in MEGA software (version 6). Bootstrap analysis of 1000 replicates was performed on each tree to determine the confidence. EBVaGC1 to 9 in the phylogenetic trees were marked with an asterisk. Abbreviations: EBV, Epstein-Barr virus.

487 (A→T) in EBNA1 in EBVaGC1 and EBVaGC4, and a C-to-T substitution at coordinate 97121 resulted in the change of residue 487 (A→V) in EBNA1 in the other 7 GC-EBV strains, where CD4<sup>+</sup> epitopes SNP, NPK, ENI, IAE, and LRA were located. A C-to-T substitution at coordinate 97232 resulted in the change of residue 524 (T→I) in EBNA1 in all 9 GC-EBV strains, where CD4<sup>+</sup> epitopes VYG, TSL, YNL, NLR, and EEG were located. Another G-to-T substitution at coordinate 97250 resulted in the change of residue 563 (M→I) in EBNA1 in only EBVaGC1 and EBVaGC4, where CD4<sup>+</sup> epitopes APG, PQP, PGP, LRE, YFM, and MVF were located. The positions of the nonsynonymous changes located in the epitopes are illustrated in Figure 6 and tabulated in Tables S4 and S5 in the Supplementary Materials.

## DISCUSSION

Epstein-Barr virus (EBV)-associated gastric carcinoma (EBVaGC) has been recognized as a distinct entity of gastric carcinoma. In the present study, 15 out of

206 cases of gastric carcinoma were identified as EBVaGC cases using EBER ISH. The frequency was 7.3%, which was similar to previously reported results in cases from Asia (8.3%) [4]. In terms of the clinicopathological features of EBVaGC, strong association with age at diagnosis was observed ( $P = 0.013$ , Table 1). The majority of patients with EBVaGC were 60 years old or younger, which is in line with previous studies [26–28]. It has been observed that EBVaGC has distinct clinicopathological features, such as male predominance, and predisposition to the proximal stomach [4, 20, 29–34]. However, no significant differences were observed in gender, differentiation, TNM stage, anatomical site, or venous invasion in our study, which may ascribe to the low number of positive cases.

EBV, a ubiquitous human herpesvirus, has a double stranded DNA genome comprised of approximately 170 kilobases with more than 85 genes. Previous studies conducted to date have targeted specific gene regions of the virus and no genome-wide sequence of EBV in EBVaGC has been completed so far. In the current study,

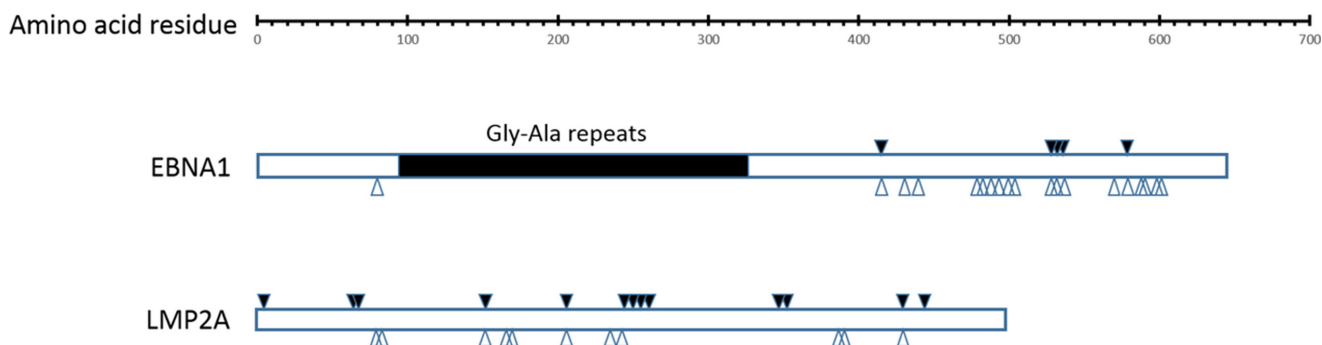
A

|         | 16 | 18 | 20 | 85 | 364 | 411 | 418 | 429 | 439 | 476 | 479 | 487 | 492 | 499 | 502 | 528 | 533 | 563 | 574 | 584 | 585 | 595 | 621-622 |
|---------|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| EBVaGC1 | E  | G  | T  | T  | G   | E   | H   | M   | A   | Q   | Q   | T   | C   | D   | T   | I   | L   | I   | G   | I   | P   | A   | DDG     |
| EBVaGC5 | E  | G  | T  | T  | G   | D   | L   | V   | T   | P   | E   | V   | S   | E   | N   | V   | I   | M   | V   | M   | T   | V   |         |
| EBVaGC2 | Q  | E  | S  | A  | R   | D   | L   | V   | T   | P   | E   | V   | S   | E   | N   | V   | I   | M   | V   | M   | T   | V   |         |

B

|         | 16 | 18 | 24 | 27 | 85 | 487 | 499 | 502 | 595 | 621-622 |
|---------|----|----|----|----|----|-----|-----|-----|-----|---------|
| AG876   | Q  | E  | D  | S  | A  | L   | E   | N   | V   |         |
| EBVaGC4 | Q  | E  | D  | S  | A  | T   | D   | T   | A   | DDG     |
| EBVaGC1 | E  | G  | E  | G  | T  | T   | D   | T   | A   | DDG     |

**Figure 5: Interstrain recombination analysis in the EBNA1 gene.** Numbers across the top correspond to the amino acid positions of EBNA1 under which the GD1 amino acid is listed. (A) Recombination analysis in the EBNA1 gene of EBVaGC5. Only Amino acid changes in at least 1 of the 3 strains, including EBVaGC1 (red), 2 (black), and 5 are indicated. The amino acids of EBVaGC5 are composed of parts of EBVaGC1 (red) and parts of EBVaGC2 (black), indicating a recombination event. (B) Recombination analysis in the EBNA1 gene of EBVaGC4. Only Amino acid changes in at least 1 of the 3 strains, including AG876 (red), EBVaGC1 (black), and EBVaGC4 are indicated. The amino acids of EBVaGC4 are composed of parts of AG876 (red) and parts of EBVaGC1 (black), indicating a recombination event. The accession numbers of relevant strains are given in Materials and Methods.



**Figure 6: Amino acid changes in CD8<sup>+</sup> and CD4<sup>+</sup> specific T cell epitopes in EBNA1 and LMP2A.** Amino acid changes in at least 1 of the 9 GC-EBV strains at CD8<sup>+</sup> and CD4<sup>+</sup> specific T cell epitopes are marked with solid (CD8<sup>+</sup>) and hollow arrows (CD4<sup>+</sup>) arrows, respectively. Abbreviations: GC-EBV, gastric carcinoma biopsy specimen-derived Epstein-Barr virus.

we used the method of target enrichment of EBV DNA by hybridization, followed by next-generation sequencing. EBV probes were designed according to full-length genome of 6 available EBV strains, including EBV-wild type, B95–8, AG876, GD1, GD2, and HKNPC1. According to the value of coverage of the target region, all DNA sequence generated from GC-EBV strains most resembled GD1. Thus, GD1 was used as the reference EBV genome in our study. *De novo* assembly was performed for 9 sequenced GC-EBV strains. The gaps between the contigs were linked up either by Sanger sequencing or tracts of “N” with length estimated based on the EBV GD1 reference. Finally, 9 EBVaGC genomes were successfully sequenced, designated EBVaGC1 to 9.

The variability in terms of the number of variations as a proportion of the total number of bases of the GC-EBV genomes, ranged from 0.052% (EBVaGC2) to 0.206% (EBVaGC6), and was much lower than the interstrain variability of ca. 0.5% for viruses of the same type, indicating that the variability is unlikely to be the result of coinfection of different viral strains. Consistent with previous reports [16], there is clear evidence (Figure 3) for a higher frequency of SNPs in latent genes, followed by the genes encoding tegument and membrane glycoproteins.

EBNA1 is essential for maintenance of the EBV episome in latently infected cells and is the only EBV antigen that is consistently expressed in all EBV associated malignancies [35]. V-val and P-thrV were observed in our study based on the amino acid changes at position 487 in the COOH-terminal region in EBNA1 relative to B95–8 (P-ala). V-val was the most common subtype, accounting for 77.7% of all 9 GC-EBV strains, followed by P-thrV, accounting for 22.3%. These findings are similar to those of previous studies on Asia populations, but different from those on western cases [18, 19, 36–44]. Multiple results showed that V-val is the dominant subtype in Asian regions studies, not only in EBVaGC but also in NPC and healthy donors, while V-val subtype was rarely found in Africa, Europe, and America irrespective of source (lymphoma, NPC, EBVaGC, or healthy donors) [18, 19, 36–44], indicating that polymorphism of EBNA1 subtypes has geographic differences but is not tumor-specific. Apart from changes in the C-terminus, EBNA1 has variations in the N-terminus. Surprisingly, EBVaGC5 from the V-val subtype and EBVaGC1 from the P-thrV subtype defined by the codon 487 had the consensus sequences at codons 16 to 364 identical to that of the B95–8 strain, and the sequences at codons 16 to 85 of EBVaGC1 and EBVaGC4 from the same P-thrV subtype defined by the codon 487 were different, and the sequences at codons 16 to 85 of EBVaGC4 were identical to that of the AG876 strain, indicating a recombination event. Accordingly, the phylogenetic analysis of nucleotide sequences of the *EBNA1* gene showed that EBVaGC1, 4 and 5 are located in different branches distant to EBVaGC2, 3, 6,

7, 8, and 9 that are clustered in a branch. Thus, EBNA1 N-terminus changes have revealed additional variants that were not simply classified based on the signature amino acid residue 487 in the C-terminus as widely used previously. The N-terminus changes reinforce the need to evaluate the EBV genome more comprehensively in order to characterize the full extent of EBV genetic diversity. A comprehensive investigation into the functional and immunological impact of the naturally occurred *EBNA1* sequence variations and interstrain recombinants is required to evaluate their possible significance, which may also be helpful for clarifying the association of EBNA1 subtypes and EBVaGC.

Recent studies show that EBNA1, as well as LMP2A, can be presented to both CD4<sup>+</sup> and CD8<sup>+</sup> T cells, highlighting its potential importance in the development of therapeutic strategies against EBV-associated malignancies [45–47]. There is some clear evidence for sequence variation affecting immune recognition of EBNA1 and potential epitope selection for vaccine development [46]. So far, most research on the EBNA1 protein has been focused exclusively on the B95–8 strain alone [46, 47]. In this study, sequence analysis of the gene encoding EBNA1 in EBV isolates from 9 EBVaGC specimens has revealed considerable *EBNA1* sequence divergence from the B95–8 strain. Importantly, T cell recognition of EBNA1 epitope might be greatly influenced by this sequence polymorphism as adoptive transfer of EBNA1-targeted T cells has a potential use in immunotherapy of EBV associated carcinomas.

Phylogenetic trees based on the LMP1 gene and whole EBV genomes indicated that the 9 GC-EBV strains were closely related to all Asian-derived EBV strains and distant to the non-Asian strains, suggesting that the LMP1 gene can serve as a geographical marker. This is in line with the previous results from the NPC-EBV genomes [16].

In summary, we reported 9 newly sequenced EBV genomes isolated from primary GC biopsy specimens and demonstrated the sequence diversity on a whole-genome level. In the future, studies should be conducted to assess the role of EBV genomic variations in gastric carcinogenesis.

## MATERIALS AND METHODS

### GC patients

206 fresh tissue specimens were collected from patients diagnosed with gastric carcinoma who underwent surgeries at Peking University Cancer Hospital, Beijing, China, from 2008 to 2014. This study was approved by the Institutional Review Board of the hospital for the purpose of EBV assays. All patients signed the informed consent form. None of the patients had received pre-operative chemotherapy or radiotherapy. Clinicopathological



parameters were collected by reviewing medical charts and pathological records, including gender, age, histological type, pathological stage, anatomical site, and venous invasion. At the time of surgery, the median age of the patients was 60 years, ranging from 32 to 82 years. The anatomical site of the tumor was determined on the basis of the predominant location of the lesion as cardia ( $n = 43$ ), body ( $n = 33$ ), antrum ( $n = 128$ ), and gastric stump ( $n = 2$ ). The histological subtypes contained poorly differentiated ( $n = 158$ ), moderately differentiated ( $n = 46$ ), and well-differentiated ( $n = 2$ ) according to the criteria of the World Health Organization. The clinical stage of the disease was determined based on the tumor, node, and metastasis (TNM) classification of the American Joint Committee on Cancer and the Union for International Cancer Control, including stage I or II ( $n = 82$ ), stage III or IV ( $n = 124$ ).

### ***In situ* hybridization**

EBV was identified by the expression of EBV-encoded small RNA (EBER). Briefly, according to the manufacturer's instructions, 4  $\mu\text{m}$  paraffin-block sections were hybridized with EBV oligonucleotide probes complementary to the EBER (Leica Biosystems Newcastle Ltd, Newcastle Upon Tyne, United Kingdom). The hybridization signals were detected using diaminobenzidine (DAB) system, and dark brown staining within tumor cell nuclei under light microscopy was defined as a positive signal. 15 cases (7.3%) of a total of 206 cases of gastric carcinomas were found to be EBV-positive.

### **Sample DNA preparation**

DNA was extracted from the tumor biopsy specimens in the EBV positivity group using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA concentration was determined by the use of a Nano-Drop (Thermo Fish Scientific, Wilmington, Delaware, USA).

### **Complete workflow for EBV DNA capture and sequencing**

#### **Prepare DNA whole-genomic library**

In order to establish the whole genomic library with the targeted gene, 2–3  $\mu\text{g}$  genomic DNA was sheared to about 150 bp fragments by Covaris S2 (Covaris, MA, USA). These fragments were subsequently end-blunted, "A"-tailed, adaptor-ligated, and amplified 7 cycles by PCR. Out of the prepared library samples, 1  $\mu\text{l}$  was used for quantification using Nanodrop 2000, and 3  $\mu\text{l}$  was used for identification in 2% agarose gel. The final size of the fragments was around 300–500 bp.

### **Capture targeted gene regions and sequencing**

EBV probes were designed according to 6 available full-length EBV reference genomes by MyGenostics (MyGenostics, Beijing, China). The overall experiment was conducted according to the manufacturer's protocol. In brief, the whole-genomic libraries were hybridized with EBV probes (MyGenostics GenCap Technology, Beijing, China), adsorbed onto the magnetic beads via biotin and streptavidin, and the uncaptured DNA fragments were removed by washing. The fragments eluted from the beads containing the targeted gene were then enriched by 14 cycles of PCR to generate libraries for sequencing. Libraries were quantified and sequenced for paired-end 125 bp using the Illumina HiSeq 2500 sequencer (Illumina Inc., San Diego, CA, USA).

### **Bioinformatics analysis method**

The 125 bp paired-end reads were used for bioinformatics analysis. For quality control, we first filtered low quality reads using the Trim Galore program. Then, 3'/5' adaptors were trimmed using the Cutadapt program implemented in Trim Galore, thereby rendering high quality clean reads, whose quality value exceeds 20 with read length greater than 80 bp. High quality reads were obtained.

Illumina clean reads were mapped to human (NCBI build 37, HG19) and each EBV reference genome using the Burrow Wheeler Aligner (BWA) program and quality scores were recalibrated and realigned to reference. Reads that perfectly paired-end aligned to the human genome were removed, while the remaining EBV clean reads were reserved. 6 available EBV reference genomes included EBV-wild type (NC\_007605.1), AG876 (DQ279927.1), B95-8 (V01555.2), GD1 (AY961628.3), GD2 (HQ020558.1), and HKNPC1 (JQ009376). According to the value of coverage of the target region, all DNA sequences generated from the GC samples most resembled that of GD1. Therefore, GD1 (AY961628.3) served as the reference EBV genome for the alignment and mutation analysis.

### **De novo assembly of EBV genomes**

EBV clean reads were assembled using the Velvet 1.2.10 (with parameters `-ins_length 180, -exp_cov auto`). The settings were optimized for each sample using k-mer lengths of 59–73, and the minimum k-mer coverage was 35 to 70. Then the contigs were blasted to the NCBI nt database and the contigs mapped to the human genome were removed.

The location and orientation of contigs were evaluated by pairwise alignment of the contigs to the reference EBV GD1 genome. The gaps between the contigs were linked up using either Sanger sequencing

or tracts of “N” with length estimated based on the EBV reference, AY961628.3.

In this study, EBV genomic sequences isolated from 9 EBVaGC biopsy specimens were successfully retrieved, designated EBVaGC1 to EBVaGC9. Sequence data for the 9 GC-EBV genomes have been submitted to GenBank. Raw sequencing data also have been submitted to Sequence Read Archive. EBV genomes of other 6 EBVaGC biopsy specimens failed to be assembled, mainly owing to the low level of the average sequencing depth on target and N50 sizes of contigs.

## Mutation analysis

Duplicated reads were removed using Sequence Alignment/Map tools (SAMtools)<sup>3</sup> and only uniquely mapped reads were used for variation detection. Single nucleotide variants (SNVs) were detected and genotyped with the GATK UnifiedGenotyper in single-sample mode (with parameters `-im ALL -mbq 20 -mmq 20 -mm42 3 -deletions 0.05`). Variants were filtered with GATK VariantFiltration module (with filters “`QUAL < 50.0 & QD < 5.0 & HRun > 10 & DP < 4`” and parameters `-cluster 3 -window 10`). Insertions and deletions (indels) were detected with GATKIndelGenotyperV2 (with parameters `-im ALL`) and filtered with a custom python module that removed sites with `amax_cons_av ≥ 1.9` (maximum average number of mismatches across reads supporting the indel) or `max_cons_nqs_av_mm ≥ 0.2` (maximum average mismatch rate in the 5-bp NQS window around the indel, across indel-supporting reads). Substitutions, insertions, and deletions were all considered as variations.

The variations identified in the major repeats, including internal repeats (IR) 1 to 4 and terminal repeats (TR), were disregarded. Positions marked with “N” were also ignored in the mutation analysis. Classification of nonsynonymous mutations was based on the 9 categories of EBV-encoded proteins, as defined by Tarbouriech *et al.* [17]. More attention was paid to those that resulted in amino acid changes in CD8<sup>+</sup> and CD4<sup>+</sup> T-cell epitopes of EBNA1 and LMP2A.

## Phylogenetic analysis of EBV genomes

Phylogenetic analysis was performed using Molecular Evolutionary Genetics Analysis version 6 (MEGA6) and the neighbor-joining algorithm. This approach was based on multiple sequence alignments of whole genomes or nucleotide sequences of individual genes of all the sequenced EBV genomes in this study. *EBNA1*, *EBNA2*, *LMP1*, *LMP2A*, and *BZLF1* genes were selected. Bootstrap analysis of 1000 replicates was performed on each tree to determine the confidence.

## Accession numbers

Sequence data for the 9 GC biopsy specimen-derived EBV (GC-EBV) genomes were submitted to GenBank under accession numbers KT273942 (EBVaGC1), KT273943 (EBVaGC2), KT254013 (EBVaGC3), KT273944 (EBVaGC4), KT273945 (EBVaGC5), KT273946 (EBVaGC6), KT273947 (EBVaGC7), KT273948 (EBVaGC8), and KT273949 (EBVaGC9). Raw sequencing data were submitted to the Sequence Read Archive (study accession number SRP060585).

## Statistical analysis

Pearson Chi-Square or Fisher’s exact test was used for statistical analysis in the present study. *P*-value of less than 0.05 was considered as statistically significant. All the *P*-values presented are two-sided.

## CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

## GRANT SUPPORT

This work was supported by grants from Natural Science Foundation of China (81321003), and Beijing Municipal Science and Technology (Z151100001615022).

## REFERENCES

1. Shibata D, Weiss LM. Epstein-Barr virus-associated gastric adenocarcinoma. *Am J Pathol.* 1992; 140:769–774.
2. Tokunaga M, Land CE, Uemura Y, Tokudome T, Tanaka S, Sato E. Epstein-Barr virus in gastric carcinoma. *Am J Pathol.* 1993; 143:1250–1254.
3. Takada K. Epstein-Barr virus and gastric carcinoma. *Mol Pathol.* 2000; 53:255–261.
4. Murphy G, Pfeiffer R, Camargo MC, Rabkin CS. Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology.* 2009; 137:824–833.
5. Akiba S, Koriyama C, Herrera-Goepfert R, Eizuru Y. Epstein-Barr virus associated gastric carcinoma: epidemiological and clinicopathological features. *Cancer Sci.* 2008; 99:195–201.
6. Farrell PJ. Epstein-Barr virus. The B95–8 strain map. *Methods Mol Biol.* 2001; 174:3–12.
7. de Jesus O, Smith PR, Spender LC, Elgueta Karstegl C, Niller HH, Huang D, Farrell PJ. Updated Epstein-Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV. *J Gen Virol.* 2003; 84:1443–1450.

8. Dolan A, Addison C, Gatherer D, Davison AJ, McGeoch DJ. The genome of Epstein-Barr virus type 2 strain AG876. *Virology*. 2006; 350:164–170.
9. Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, Baribault C, Fewell C, Hulme W, Hedges D, Taylor CM, Flemington EK. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *J Virol*. 2013; 87:1172–1182.
10. Lei H, Li T, Hung GC, Li B, Tsai S, Lo SC. Identification and characterization of EBV genomes in spontaneously immortalized human peripheral blood B lymphocytes by NGS technology. *BMC genomics*. 2013; 14:804.
11. Cheung ST, Huang DP, Hui AB, Lo KW, Ko CW, Tsang YS, Wong N, Whitney BM, Lee JC. Nasopharyngeal carcinoma cell line (C666–1) consistently harbouring Epstein-Barr virus. *Int J Cancer*. 1999; 83:121–126.
12. Tso KK, Yip KY, Mak CK, Chung GT, Lee SD, Cheung ST, To KF, Lo KW. Complete genomic sequence of Epstein-Barr virus in nasopharyngeal carcinoma cell line C666–1. *Infect Agent Cancer*. 2013; 8:29.
13. Zeng MS, Li DJ, Liu QL, Song LB, Li MZ, Zhang RH, Yu XJ, Wang HM, Ernberg I, Zeng YX. Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. *J Virol*. 2005; 79:15323–15330.
14. Liu P, Fang X, Feng Z, Guo YM, Peng RJ, Liu T, Huang Z, Feng Y, Sun X, Xiong Z, Guo X, Pang SS, Wang B, et al. Direct sequencing and characterization of a clinical isolate of Epstein-Barr virus from nasopharyngeal carcinoma tissue by using next-generation sequencing technology. *J Virol*. 2011; 85:11291–11299.
15. Kwok H, Tong AH, Lin CH, Lok S, Farrell PJ, Kwong DL, Chiang AK. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. *PLoS One*. 2012; 7:e36939.
16. Kwok H, Wu CW, Palser AL, Kellam P, Sham PC, Kwong DL, Chiang AK. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. *J Virol*. 2014; 88:10662–10672.
17. Tarbouriech N, Buisson M, Geoui T, Daenke S, Cusack S, Burmeister WP. Structural genomics of the Epstein-Barr virus. *Acta crystallographica Section D, Biological crystallography*. 2006; 62:1276–1285.
18. Bhatia K, Raj A, Guitierrez MI, Judde JG, Spangler G, Venkatesh H, Magrath IT. Variation in the sequence of Epstein Barr virus nuclear antigen 1 in normal peripheral blood lymphocytes and in Burkitt's lymphomas. *Oncogene*. 1996; 13:177–181.
19. Gutierrez MI, Raj A, Spangler G, Sharma A, Hussain A, Judde JG, Tsao SW, Yuen PW, Joab I, Magrath IT, Bhatia K. Sequence variations in EBNA-1 may dictate restriction of tissue distribution of Epstein-Barr virus in normal and tumour cells. *J Gen Virol*. 1997; 78:1663–1670.
20. Chen JN, Ding YG, Feng ZY, Li HG, He D, Du H, Wu B, Shao CK. Association of distinctive Epstein-Barr virus variants with gastric carcinoma in Guangzhou, southern China. *J Med Virol*. 2010; 82:658–667.
21. Sugiura M, Imai S, Tokunaga M, Koizumi S, Uchizawa M, Okamoto K, Osato T. Transcriptional analysis of Epstein-Barr virus gene expression in EBV-positive gastric carcinoma: unique viral latency in the tumour cells. *Br J Cancer*. 1996; 74:625–631.
22. Luo B, Wang Y, Wang XF, Liang H, Yan LP, Huang BH, Zhao P. Expression of Epstein-Barr virus genes in EBV-associated gastric carcinomas. *World J Gastroenterol*. 2005; 11:629–633.
23. Long HM, Leese AM, Chagoury OL, Connerty SR, Quarcoopome J, Quinn LL, Shannon-Lowe C, Rickinson AB. Cytotoxic CD4<sup>+</sup> T cell responses to EBV contrast with CD8 responses in breadth of lytic cycle antigen choice and in lytic cycle recognition. *J Immunol*. 2011; 187:92–101.
24. Sitompul LS, Widodo N, Djati MS, Utomo DH. Epitope mapping of gp350/220 conserved domain of Epstein Barr virus to develop nasopharyngeal carcinoma (npc) vaccine. *Bioinformation*. 2012; 8:479–482.
25. Hislop AD, Taylor GS, Sauce D, Rickinson AB. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. *Annual review of immunology*. 2007; 25:587–617.
26. Herrera-Goepfert R, Akiba S, Koriyama C, Ding S, Reyes E, Itoh T, Minakami Y, Eizuru Y. Epstein-Barr virus-associated gastric carcinoma: Evidence of age-dependence among a Mexican population. *World J Gastroenterol*. 2005; 11:6096–6103.
27. Trimeche M, Ksiaa F, Ziadi S, Mestiri S, Hachana M, Gacem RB, Sriha B, Korbi S. Prevalence and characteristics of Epstein-Barr virus-associated gastric carcinomas in Tunisia. *Eur J Gastroenterol Hepatol*. 2009; 21:1001–1007.
28. BenAyed-Guerfali D, Ayadi W, Miladi-Abdennadher I, Khabir A, Sellami-Boudawara T, Gargouri A, Mokdad-Gargouri R. Characteristics of Epstein Barr virus variants associated with gastric carcinoma in Southern Tunisia. *Virology*. 2011; 8:500.
29. Uozaki H, Fukayama M. Epstein-Barr virus and gastric carcinoma—viral carcinogenesis through epigenetic mechanisms. *International journal of clinical and experimental pathology*. 2008; 1:198–216.
30. Alipov G, Nakayama T, Nakashima M, Wen CY, Niino D, Kondo H, Pruglo Y, Sekine I. Epstein-Barr virus-associated gastric carcinoma in Kazakhstan. *World J Gastroenterol*. 2005; 11:27–30.
31. Chang MS, Lee HS, Kim CW, Kim YI, Kim WH. Clinicopathologic characteristics of Epstein-Barr virus-incorporated gastric cancers in Korea. *Pathology, research and practice*. 2001; 197:395–400.
32. Galetsky SA, Tsvetnov VV, Land CE, Afanasieva TA, Petrovichev NN, Gurtsevitch VE, Tokunaga M. Epstein-

- Barr-virus-associated gastric cancer in Russia. *Int J Cancer*. 1997; 73:786–789.
33. Camargo MC, Murphy G, Koriyama C, Pfeiffer RM, Kim WH, Herrera-Goepfert R, Corvalan AH, Carrascal E, Abdirad A, Anwar M, Hao Z, Kattoor J, Yoshiwara-Wakabayashi E, et al. Determinants of Epstein-Barr virus-positive gastric cancer: an international pooled analysis. *Br J Cancer*. 2011; 105:38–43.
  34. Fukayama M, Ushiku T. Epstein-Barr virus-associated gastric carcinoma. *Pathology, research and practice*. 2011; 207:529–537.
  35. Young LS, Murray PG. Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene*. 2003; 22:5108–5121.
  36. Chang KL, Chen YY, Chen WG, Hayashi K, Bacchi C, Bacchi M, Weiss LM. EBNA-1 gene sequences in Brazilian and American patients with Hodgkin's disease. *Blood*. 1999; 94:244–250.
  37. Chen YY, Chang KL, Chen WG, Shibata D, Hayashi K, Weiss LM. Epstein-Barr virus-associated nuclear antigen-1 carboxy-terminal gene sequences in Japanese and American patients with gastric carcinoma. *Laboratory investigation*. 1998; 78:877–882.
  38. Fassone L, Cingolani A, Martini M, Migliaretti G, Oreste PL, Capello D, Gloghini A, Vivenza D, Dolcetti R, Carbone A, Antinori A, Gaidano G, Larocca LM. Characterization of Epstein-Barr virus genotype in AIDS-related non-Hodgkin's lymphoma. *AIDS research and human retroviruses*. 2002; 18:19–26.
  39. Greiner TC, Abou-Elella AA, Smir BN, Orazi A, Hinrichs S, Anderson J, Gross T, Bierman P, Hauke R. Molecular epidemiology of EBNA-1 substrains of Epstein-Barr virus in posttransplant lymphoproliferative disorders which have infrequent p53 mutations. *Leukemia & lymphoma*. 2000; 38:563–576.
  40. Gutierrez MI, Spangler G, Kingma D, Raffeld M, Guerrero I, Misad O, Jaffe ES, Magrath IT, Bhatia K. Epstein-Barr virus in nasal lymphomas contains multiple ongoing mutations in the EBNA-1 gene. *Blood*. 1998; 92:600–606.
  41. Habeshaw G, Yao QY, Bell AI, Morton D, Rickinson AB. Epstein-barr virus nuclear antigen 1 sequences in endemic and sporadic Burkitt's lymphoma reflect virus strains prevalent in different geographic areas. *J Virol*. 1999; 73:965–975.
  42. MacKenzie J, Gray D, Pinto-Paes R, Barrezueta LF, Armstrong AA, Alexander FA, McGeoch DJ, Jarrett RF. Analysis of Epstein-Barr virus (EBV) nuclear antigen 1 subtypes in EBV-associated lymphomas from Brazil and the United Kingdom. *J Gen Virol*. 1999; 80:2741–2745.
  43. 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. *J Pathol*. 2000; 191:127–131.
  44. Wang Y, Liu X, Xing X, Cui Y, Zhao C, Luo B. Variations of Epstein-Barr virus nuclear antigen 1 gene in gastric carcinomas and nasopharyngeal carcinomas from Northern China. *Virus Res*. 2010; 147:258–264.
  45. Rooney CM, Leen AM, Vera JF, Heslop HE. T lymphocytes targeting native receptors. *Immunological reviews*. 2014; 257:39–55.
  46. Bell MJ, Brennan R, Miles JJ, Moss DJ, Burrows JM, Burrows SR. Widespread sequence variation in Epstein-Barr virus nuclear antigen 1 influences the antiviral T cell response. *J Infect Dis*. 2008; 197:1594–1597.
  47. Icheva V, Kayser S, Wolff D, Tuve S, Kyzirakos C, Bethge W, Greil J, Albert MH, Schwinger W, Nathrath M, Schumm M, Stevanovic S, Handgretinger R, et al. Adoptive transfer of Epstein-Barr virus (EBV) nuclear antigen 1-specific T cells as treatment for EBV reactivation and lymphoproliferative disorders after allogeneic stem-cell transplantation. *J Clin Oncol*. 2013; 31:39–48.