We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Genome-Wide Profiling of Epstein-Barr Virus (EBV) Isolated from EBV-Related Malignancies

Ying Liu, Zheming Lu and Hongying Huang

Abstract

Epstein–Barr virus (EBV) is the cause of certain cancers, such as Burkitt lymphoma, Hodgkin lymphoma, NK/T cell lymphoma, nasopharyngeal carcinoma, and a subset of gastric carcinomas. The genome-wide characteristics of EBV are essential to understand the diversity of strains isolated from EBV-related malignancies, provide the first opportunity to test the general validity of the EBV genetic map and explore recombination, geographic variation, and the major features of variation in this virus. Moreover, understanding more about EBV sequence variations isolated from EBV-related malignancies might give important implications for the development of effective prophylactic and therapeutic vaccine approaches targeting the personalized or geographic-specific EBV antigens in these aggressive diseases. In this chapter, we will mainly focus on the EBV genome-wide profiling in three common EBV-related cancers in Asia, including nasopharyngeal carcinoma, EBV-associated gastric carcinoma, and NK/T-cell lymphoma.

Keywords: Epstein–Barr virus (EBV), next-generation sequencing (NGS), nasopharyngeal carcinoma (NPC), EBV-associated gastric carcinoma (EBVaGC), NK/T-cell lymphoma (NKTCL)

1. Introduction

Epstein–Barr virus (EBV), a ubiquitous human herpesvirus discovered in 1964 is classified as a group I carcinogen by the International Agency for Research on Cancer (IARC), since the latent infection by EBV has been estimated to be responsible for 200,000 cancer cases worldwide [1], including Burkitt lymphoma, Hodgkin lymphoma, NK/T cell lymphoma (NKTCL), nasopharyngeal carcinoma (NPC), and a subset of gastric carcinomas. It has been shown that viruses can contribute to the biology of multistep oncogenesis and are implicated in many of the hallmarks of cancer [2]. Notably, the discovery of links between viral infection and cancer types has provided actionable opportunities, such as the use of human papilloma virus (HPV) vaccines as a preventive measure, to reduce the global impact of cancer. However, until now, approved vaccines for EBV have not been available.

EBV has a double stranded DNA genome comprised of approximately 172 kilobases. The expression products cover at least 86 proteins and 46 functional small-untranslated RNAs [3–5]. EBV has two distinct life cycles: latency and lytic

replication. During latency, viral genomes only express a limited number of latent proteins (EBV-determined nuclear antigen 1 (EBNA1), 2, 3A, 3B, and 3C and EBNA leader protein (EBNA-LP); latent membrane protein 1 (LMP1) and LMP2 (which encodes two isoforms, LMP2A and LMP2B)), noncoding EBV-encoded RNAs (EBER1 and EBER2), and viral miRNAs (BHRF1-miRNA and BART-miRNA). EBV latency is categorized as three latency types (latency I–III). EBV genomes in type-I latency are known to express EBNA1 and EBER. EBV genomes in type-II latency are known to express more genes such as *EBNA-LP*, *LMP1*, *LMP2A*, and *LMP2B*. EBV genomes in type-III latency are known to express most restricted latent genes including *EBNA2*, *EBNA3A*, *EBNA3B*, and *EBNA3C*. Lytic genes encode viral transcription factors (e.g., BZLF1), a viral DNA polymerase (BALF5) and associated factors, and viral glycoproteins (e.g., gp350/220 and gp110) and structural proteins (capsid and tegument proteins).

Southern blot of restriction fragment length polymorphisms was first used to detect EBV strain variation, and Sanger sequencing of certain specific viral genes (e.g., EBNA1 and LMP1) was later developed to detect sequence diversity. Now, on the basis of high-throughput sequencing, genome-wide analysis is becoming possible.

Prior to 2013, EBV whole genome sequences available from GenBank were limited to less than 10 strains (B95-8, EBV-WT, GD1, AG876, GD2, HKNPC1, Akata, and Mutu). The prototypic type 1 EBV strain B95-8 was the first complete genome sequenced from an individual with infectious mononucleosis using a conventional strategy (i.e., subcloning followed by Sanger sequencing) [6]. Subsequently, 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 (named EBV-WT) [7]. AG876 was the unique complete type 2 EBV sequence from a Ghanaian case of Burkitt lymphoma [8]. Akata and Mutu were sequenced from Burkitt lymphoma cell lines from a Japanese patient and a Kenyan patient, respectively [9]. GD1 [10], GD2 [11] and HKNPC1 [12] were isolated from NPC patients.

Since 2014, a new technology named Hybrid Capture (**Figure 1**), has marked a new era of EBV genome sequencing. Using the method of target enrichment of EBV DNA by hybridization, followed by next-generation sequencing, de novo assembly, and joining of contigs can yield complete EBV genomes. The development of high-throughput sequencing technologies enabled sequencing of EBV genomes derived from a wide variety of clinical samples, such as tumor biopsy samples [13]. The number of available EBV sequences is increasing exponentially and up to now, more than 500 EBV genomes have been sequenced from a variety of human malignancies, including NPC, lymphoma, gastric cancer, and lung cancer, as well as from healthy carriers [14–25]. Progress has made it possible that the population-based case–control studies of EBV strain variation in EBV-related cancer patients as compared with the healthy population and a comprehensive survey of EBV integration in a variety of human malignancies can be effectively conducted [20, 25–27]. These developments have revealed that various EBV strains are differentially distributed throughout the world, and that the behavior of cancer-derived EBV strains is different from that of the prototype EBV strain of noncancerous origin.

Hence, the genome-wide characteristics of EBV are essential to assess the diversity of strains isolated from EBV-related malignancies. Meanwhile, 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, and might provide important implications for the development of effective prophylactic and therapeutic vaccine approaches targeting the personalized or geographic-specific EBV antigens in these aggressive diseases.

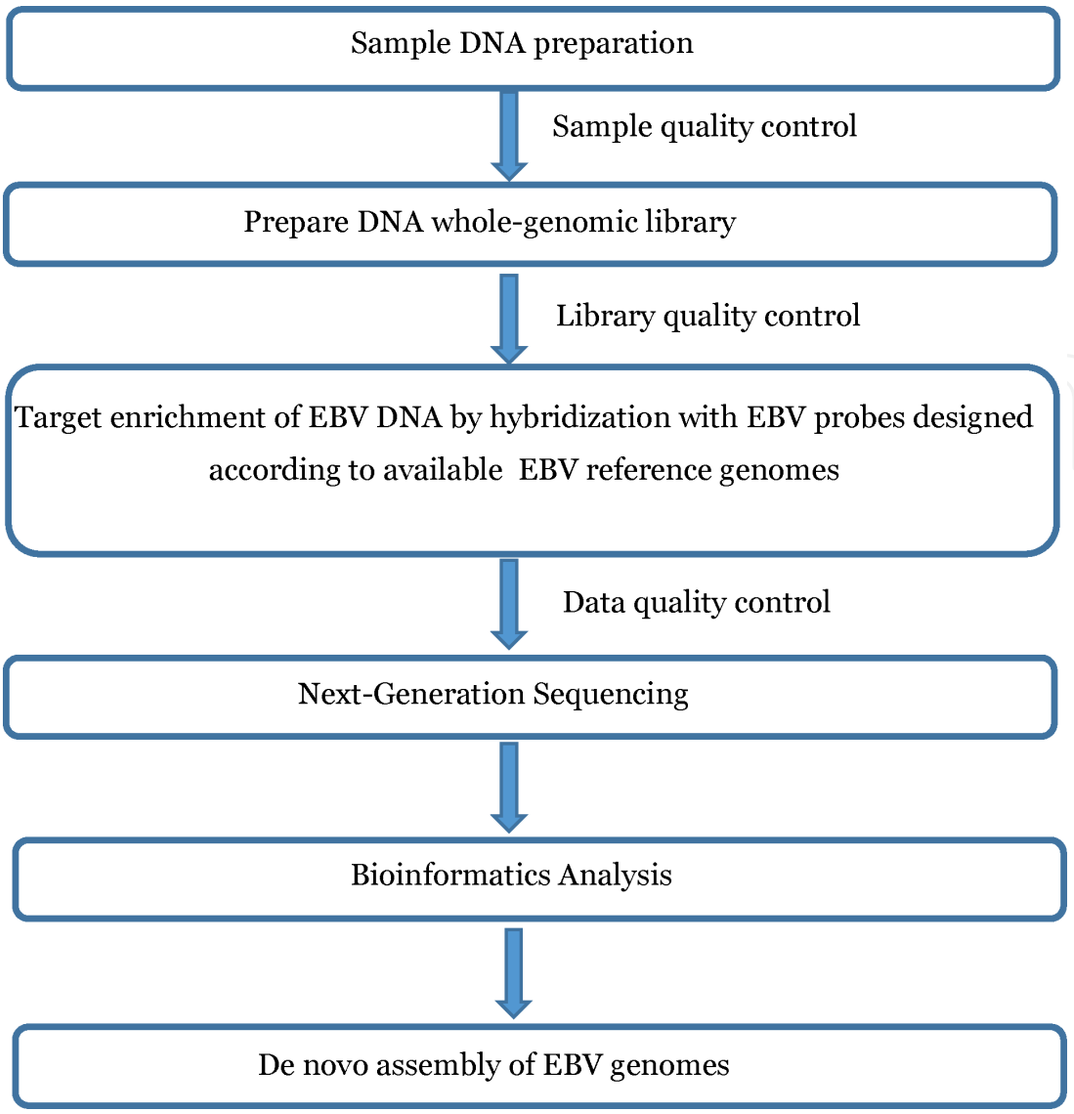


Figure 1.
Complete workflow for EBV DNA capture and sequencing.

EBV strain	Origin	Disease	Year	Refs
GD1	Guangdong, China	NPC	2005	[10]
GD2	Guangdong, China	NPC	2011	[11]
HKNPC1	Hong Kong, China	NPC	2012	[12]
HKNPC2-HKNPC9	Hong Kong, China	NPC	2014	[14]
EBVaGC1-EBVaGC9	Beijing, China	EBVaGC	2016	[17]
GDGC1-GDGC2	Guangdong, China	EBVaGC	2018	[21]
NKTCL-EBV1-NKTCL-EBV8	Beijing, China	NKTCL	2019	[23]
NKTCL-SC01-NKTCL-SC15	15 from Southern China, 12 from Singapore	NKTCL	2019	[24]
NKTCL-SG01-NKTCL-SG12				

Table 1.
EBV genomes reviewed in this chapter

In this chapter, EBV genomes reviewed are from three common EBV-related cancers in Asia, including NPC, EBV-associated gastric carcinoma (EBVaGC), and NKTCL. The EBV strains include GD1 [10], GD2 [11], HKNPC1 [12],

HKNPC2-HKNPC9 [14], EBVaGC1-EBVaGC9 [17], GDGC1-GDGC2 [21], NKTCL-EBV1-NKTCL-EBV8 [23], NKTCL-SC01-NKTCL-SC15 and NKTCL-SG01-NKTCL-SG12 [24] (**Table 1**).

2. Genomic diversity of EBV-related malignances

2.1 NPC

NPC, an EBV-associated epithelial carcinoma, has a unique geographical distribution [28]. A recent World Health Organization (WHO) report estimated that there were around 130,000 new NPC cases worldwide in 2018 [29]. Rare in most of the world, NPC is particularly prevalent in South China and Southeast Asia [30]. In Hong Kong and Guangdong in South China, NPC incidence is as high as 12.8–25.0/100,000 per year [28, 29]. The cause of NPC endemicity remains unknown.

Many studies have shown that EBV genome is present in almost all endemic NPC tumors with a unique pattern of virus latent gene expression, suggesting that EBV plays an important role in the tumorigenesis of NPC [31]. Whole genome sequencing is useful for us to understand genomic characterization and divergence. Here, we mainly focus on 11 mostly available full-length genomes of NPC.

2.1.1 GD1

GD1 (Guangdong strain 1), the first NPC-derived EBV strain with full-length sequences determined using PCR amplification and sub-cloning followed by conventional Sanger sequencing technology, was analyzed from established a lymphoblastoid cell line (LCL) from umbilical cord blood mononuclear cells transformed by saliva virus from a Cantonese NPC patient in 2005 [10]. The entire GD1 sequence is 171,656 bp in length and GD1 belongs to type 1 strain. Many sequence variations in GD1 compared to prototypical strain B95-8 were detected, including 43 deletion sites, 44 insertion sites, and 1413 point mutations. Furthermore, the frequency of some GD1 mutations in Cantonese NPC patients was evaluated, such as a 30-bp deletion in the C terminus of LMP1, and the V-Val subtypes of EBNA1. The results suggested that GD1 is highly representative of the EBV strains isolated from NPC patients in Guangdong, China, an area with the highest incidence of NPC in the world.

2.1.2 GD2

With the invention of next-generation sequencing (NGS) systems, it is possible to determine genome-wide sequences and the viral clonality of EBV strains by direct sequencing of EBV genomes in clinical tumors in a time- and cost-effective manner. GD2 with 164,701 bp long was directly sequenced using the Illumina (Solexa) platform, and successfully assembled from an NPC tumor of a patient in Guangdong province, a region in China by the same group who determined GD1 [11]. GD2 was closely related to GD1 by sequence and phylogenetic analyses. The sequence similarities between GD2 and GD1 were 98.76%. GD2 and GD1 shared 505 common single-nucleotide variations (SNVs), including most SNVs in the coding regions (348 [68.91%] SNVs) and seven insertion and deletions (indels). From a comparison with the EBV-WT reference genome, a total of 927 SNVs and 160 indels with genome-wide distribution were found in the GD2 genome. The results

revealed that NGS allows the characterization of genome-wide variations of EBV in clinical tumors and provides evidence of monoclonal expansion of EBV *in vivo*.

2.1.3 HKNPC1

Because of the relatively small quantity of viral DNA present in the tumor sample, next-generation sequencing total cellular and viral DNA in a sample is costly and inefficient, and may limit the generation of the high read depth necessary to make high confident base calls of the viral genome. Using target enrichment technology could increase the relative amount of viral DNA. Kwok et al. reported an approach of PCR enrichment (Amplicon Sequencing) followed by sequencing the amplified products on the Illumina Genome Analyzer IIx platform to determine the genome sequence of an EBV isolate from NPC tumor of a Chinese patient in Hong Kong, designated as HKNPC1 [12]. HKNPC1 is approximately 171,549 bp, and contains 1589 SNVs and 132 indels in comparison to the reference EBV-WT sequence. Non-synonymous SNVs were mainly found in the latent, tegument and glycoprotein genes. The same point mutations were found in glycoprotein (BLLF1 and BALF4) genes of GD1, GD2 and HKNPC1 strains and might affect cell type specific binding. The results showed that whole genome sequencing of EBV in NPC may facilitate discovery of previously unknown variations of pathogenic significance.

2.1.4 HKNPC2-9

The group of Kwok and colleagues established a complete sequencing workflow comprising target enrichment of EBV DNA by hybridization, followed by next-generation sequencing, *de novo* assembly, and joining of contigs by Sanger sequencing to yield whole EBV genomes. The sequences of eight NPC biopsy specimen-derived EBV (NPC-EBV) genomes, designated HKNPC2 to HKNPC9, were then determined in the same geographic location in order to reveal their sequence diversity [14]. The eight NPC-EBV genome sizes estimated based on the reference EBV-WT sequence ranged from 170,062 bp (HKNPC2) to 171,556 bp (HKNPC3 and -6). A total of 1736 variations were found, including 1601 substitutions, 64 insertions, and 71 deletions, compared to the reference EBV-WT genome. Furthermore, genes encoding latent, early lytic, and tegument proteins and glycoproteins were found to contain nonsynonymous mutations of potential biological significance. Thus, much greater sequence diversity among EBV isolates derived from NPC biopsy specimens is demonstrated on a whole-genome level through a complete sequencing workflow.

Obtaining whole-genome sequence information for more clinical EBV isolates, with good representation of the EBV repertoire in tumors, could help to address that hypothesis and uncover the pathogenic subtypes of EBV in NPC tumorigenesis. A case–control (62 NPC patients and 142 population carriers) study of NPC in Hong Kong has identified high-risk EBV subtypes with polymorphisms in the EBV-encoded small RNA (EBER) locus [26]. A recent study published in Nature Genetics entitled 'Genome sequencing analysis identifies high-risk Epstein–Barr virus subtypes for nasopharyngeal carcinoma' by Xu et al. used large-scale EBV whole-genome sequencing to examine EBV subtypes in an attempt to explain the unique NPC endemicity in South China [25]. Through EBV genomes from 156 NPC cases and 47 controls and two-stage association study, they identified two non-synonymous EBV variants within the BALF2 gene (BamHIA leftward reading frame 2 encoding a single strand DNA binding protein associated with EBV replication) strongly associated with the risk of NPC (odds ratio [OR] = 8.69 for SNP162476_C and OR = 6.14 for SNP163364_T). The cumulative effects of these variants contribute to 83% of

the overall risk of NPC in southern China. These studies confirmed the critical role of EBV infection in the pathogenesis of NPC and provided an explanation for the striking epidemiological distribution of this tumor in South China.

2.2 EBVaGC

EBVaGC has been recognized as a distinct subset of gastric carcinoma, accounting for about 10% of total gastric carcinomas [32–35]. 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, providing strong evidence to support the role of EBV as an etiologic agent [32, 33]. However, the exact role of EBV in the development and progression of this specific type of gastric carcinoma is not yet clear.

Progress has been made in understanding the full spectrum of diversity existent within the EBV genome from EBVaGC clinical tumor samples, since the NGS technology has been developed. Here, 11 EBV strains from primary EBVaGC biopsy samples were included.

2.2.1 EBVaGC1-EBVaGC9

Our group reported the first genome-wide view of sequence variation of EBV isolated from primary EBVaGC biopsy specimens in 2016 [17]. 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 six available EBV strains, including EBV-WT, 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 nine sequenced GC-EBV strains. Finally, nine EBVaGC genomes were successfully sequenced, designated EBVaGC1 to EBVaGC9. The genome sizes, estimated based on the reference GD1 sequence, ranged from 171,612 bp (EBVaGC6) to 171,957 bp (EBVaGC1).

Whole-genome sequencing of EBV enabled the comparison and thus the determination of EBV variations at the genome level. In our study, 961 variations were observed in the EBVaGC1 to 9 genomes in comparison to the reference GD1, including 919 substitutions, 23 insertions, and 19 deletions. Both latent genes and genes encoding tegument proteins in nine 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.

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 [36]. 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 nine GC-EBV strains, followed by P-thrV, accounting for 22.3%. 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) [37–39], 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. Interestingly, we identified two interstrain recombinants at the EBNA1 locus, which provided a further mechanism for the generation of diversity. 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.

2.2.2 GDGC1 and GDGC2

In 2018, NGS was employed to determine the EBV genomes from two EBVaGC specimens, designated as GDGC1 and GDGC2, from Guangdong, China, an endemic area of NPC [21]. Due to the presence of the much more abundant cellular genomic DNA in the DNA preparations, the number of reads belonging to EBV was low, accounting for only 0.02–0.23% of the total reads. However, since the original data were sufficient, the average sequencing depth for genomes GDGC1 and GDGC2 was ~73x and ~24x, respectively, which was sufficient for further analysis. The genome sizes, estimated based on the reference EBV-WT genome sequence, were as follows: GDGC1 (169,611 bp) and GDGC2 (171,299 bp).

The authors reported that a total of 1815 SNPs (146 indels) and 1519 SNPs (106 indels) were found in GDGC1 and GDGC2, respectively, compared with the reference EBV-WT genome. Among these, 1229 SNPs (66 indels) and 1076 SNPs (54 indels) were located in the coding regions for GDGC1 and GDGC2, respectively, while the remaining variations were found in the non-coding regions. Consistent with previous reports [17], there is clear evidence for a higher frequency of SNPs in latent genes, followed by the genes encoding tegument and membrane glycoproteins. In contrast to the frequent mutations that occurred in latent genes, the sequences of promoters and ncRNAs were investigated to be strictly conserved. A few point mutations were found in the sequences of Cp, Qp, Fp and LMP2Ap, and only scattered mutations could be identified in certain ncRNA sequences. Promoters and EBV-generated ncRNAs play important roles in regulating viral processes and in mediating host-virus interactions. Thus, a detailed EBV genome-wide analysis of EBVaGC from Guangdong was performed, which would be helpful for further understanding of the relationship between EBV genomic variation and EBVaGC carcinogenesis.

The features of the disease and geographically associated EBV genetic variation as well as the roles that the variation plays in carcinogenesis and evolution remain unclear. A recent study sequenced 95 geographically distinct EBV isolates from EBVaGC biopsies (n = 41) and saliva of healthy donors (n = 54) to detect variants and genes associated with gastric carcinoma from a genome-wide spectrum [20]. BRLF1, BBRF3, and BBLF2/BBLF3 genes had significant associations with gastric carcinoma. LMP1 and BNLF2a genes were strongly geographically associated genes in EBV. The results provided insights into the genetic basis of oncogenic EBV for gastric carcinoma, and the genetic variants associated with gastric carcinoma could serve as biomarkers for oncogenic EBV.

2.3 NKTCL

Extranodal NKTCL, a rare type of non-Hodgkin lymphoma, is characterized by the presence of EBV in virtually all cases, irrespective of their ethnicity or geographical origin. NKTCL is an aggressive malignancy, predominantly occurs in the nasal, paranasal, and oropharyngeal sites, and is much more prevalent in East Asia and Latin America than in Western countries [40].

Although the association of this B lymphotropic virus with malignancies of T and NK cell origin was quite unexpected, both the presence of virus sequences in tumor cells and the virus's oncogenic potency have led to the hypothesis that whether particular EBV strains are preferentially selected in NKTCL. Pathogenesis and genotype analyses of NKTCL have mainly focused on genetic variations in a small fraction of EBV genes before, which is limited to define the spectrum of diversity within the whole genome of EBV. The genome-wide characteristics of EBV are essential to understand the diversity of strains isolated from NKTCL. In 2019, for the first time, 35 NKTCL-derived EBV genomic landscapes at genome-wide level were simultaneously systematically characterized by two groups.

2.3.1 NKTCL-EBV1-NKTCL-EBV8

Our group directly sequenced EBV-captured DNA from eight primary NKTCL biopsy samples from China using Illumina HiSeq 2500 sequencer platform and presented the eight EBV sequences, designated NKTCL-EBV1-NKTCL-EBV8 [23]. Aiming at knowing the detail of subtype, the obtained DNA sequences were compared with six mostly referenced sequences, including AG876, B95-8, EBV WT, GD1, GD2, and HKNPC1. The GD1 coverage percentages are higher than the rest. The genome sizes, estimated based on the reference GD1 sequence, ranged from 171,590 bp (NKTCL-EBV8) to 172,059 bp (NKTCL-EBV1).

Whole-genome sequence alignments revealed extensive nucleotide variation in the eight NKTCL-EBV genomes. In comparison with the most similar GD1 strain, the NKTCL-EBV1 to NKTCL-EBV8 harbored 2072 variations in total, including 1938 substitutions, 58 insertions, and 76 deletions. Among them, 1218 substitutions, 15 insertions, and 26 deletions were located in the coding regions. The number of the nonsynonymous mutations is highest in the gene regions encoding latent proteins in each of the NKTCL-EBV genomes, followed by genes encoding the tegument protein and membrane glycoproteins.

EBNA1 and LMP1 are the most frequently studied regions to date. Based on the amino acid changes in certain residues of LMP1 and EBNA1, eight NKTCL-EBVs were sorted to China 1 and V-val subtype, respectively. Of interest, EBNA1 of NKTCL-EBV3 sequence showed clustered away from the other seven NKTCL-EBV strains. Analysis of amino acid sequences of EBNA1 supported that EBNA1 of NKTCL-EBV3 may arise from recombination of GD1 and B95-8. Other two commonly classification systems for LMP-1 gene polymorphisms include a 30-bp deletion in the C terminus and the loss of restriction site Xho I in the N terminus of the gene. LMP1 is a key latent protein with abilities to promote cell proliferation and inhibit cell apoptosis in NKTCL. In our study, the LMP1 strain in NKTCL-EBV1-NKTCL-EBV7, but not NKTCL-EBV8, harbored the 30-bp deletion. The variant of 30-bp deletion of LMP1 has been demonstrated that it is associated with poor prognosis of patients with NKTCL, which might serve as a potential marker to monitor treatment [41]. In addition, eight NKTCL-EBV strains had Xho I restriction site loss at exon 1 of the LMP1 gene.

2.3.2 NKTCL-SC01-NKTCL-SC15 and NKTCL-SG01-NKTCL-SG12

The other group assembled 27 NKTCL-derived EBV genome sequences retrieved from whole-genome sequencing data using the Hiseq sequencer (Illumina), including 15 EBV-positive NKTCL tumor samples from Southern China and 12 samples from Singapore [24]. The average percentage of EBV sequences in WGS data is 0.45% (0.03–1.06%), and the coverage depth is 222.2X in average (26.7X–612.8X). As ~34 kb of 172 kb of EBV genome are repeat regions, which could not be properly

assembled with short-reads sequencing technology, the groups assigned "N" for these regions and subsequently joined the scaffolds, resulting in EBV genomes with ~172 kb in length.

The authors reported that among the 27 NKTCL samples, in average 1152 EBV SNVs for each sample were determined by aligning the viral reads against the reference EBV-WT genome. The most frequent tumor-specific non-synonymous mutations in NKTCL-derived EBV were located at BPLF1 gene (position 49,790–59,239 bp). An average of 44.8 small indels (<50 bp) of EBV were found in each NKTCL sample, and the 30-bp deletion of LMP1 was commonly found in the samples (21/27), with a frequency consistent with the previous study revealed by using Sanger sequencing [42]. Large deletions of EBV (>1 kb) were found in 10 of 27 NKTCL samples, without any sequencing coverage in the deleted regions. The findings provided insights into the understanding of EBV's role the etiology of NKTCL.

A genome-wide association study of 189 patients with extranodal NKTCL, nasal type and 957 controls from Guangdong province, Southern China was performed to identify common genetic variants affecting individual risk of NKTCL [43]. All cases were genotyped with Illumina Human OmniExpress ZhongHua-8 BeadChip, and population controls were scanned by Illumina OmniHumanExpress-24 V1.0 (both Illumina, San Diego, CA, USA). The findings were validated in four independent case–control series. The SNP with the strongest association was rs9277378 (OR 2.65 [95% CI 2.08–3.37]) located with HLA-DPB1, indicating the importance of HLA-DP antigen presentation in the pathogenesis of NKTCL. The pathogenic subtypes of EBV in NKTCL tumorigenesis should be further explored.

3. Phylogenetic analysis of the EBV genomes

Phylogenetic analysis of EBV genomes could demonstrate detailed overall genomic differences in EBV genome within or beyond subtypes of EBV-associated diseases, thus, EBV genomic similarity is likely to better infer the phylogenetic relatedness among EBV genomes.

Traditionally, EBV has two distinct subtypes, type 1 and type 2. Type 1 EBV (e.g., B95-8, GD1 and Akata) is the main EBV strain prevalent worldwide, while type 2 EBV (e.g., AG876) is abundant only in parts of Africa and New Guinea. Type 1 and type 2 EBV encode different *EBNA2* genes, with only 54% amino acid sequence identity. A recent whole genome sequencing study confirmed that *EBNA2* and *EBNA3* are the only genes that can distinguish type 1 and type 2 EBV strains [16]. Technologies for genome sequencing were currently developed with tools for genome analysis. High-throughput sequencing technology such as illumine dye sequencing was introduced to successfully sequence viral genomes. As exemplary tools for genome analysis, Molecular Evolutionary Genetics Analysis (MEGA) is used for both conducting statistical analysis of molecular evolution and constructing phylogenetic trees [44].

The NPC genomes from Asian EBV strains, including GD1, GD2, and HKNPC1-HKNPC9, are type 1 viruses and were clustered in a branch distant to the non-Asian strains AG876, B95-8 [14]. Analysis of LMP1 and -2 showed a phylogenetic relationship corresponding to the geographical origin of the viral genomes instead of the type 1 and 2 dichotomy, indicating that LMP1 and -2 genes can serve as geographical markers. GD1 seemed to harbor many mutations that were not present in the other Chinese strains. HKNPC6 and -7 genomes, which were isolated from tumor biopsy specimens of advanced metastatic NPC cases, were distinct from the

other NPC-EBV genomes. Future work should investigate the relationship between the distinct lineage of EBV and the clinical stages of NPC.

GC-EBV strains, EBVaGC1-EBVaGC9 and GDGC1-GDGC2 involved here, were closely related to all Asian-derived EBV strains, distant to the non-Asian strains, and also showed that the EBV sequences generally clustered in a manner consistent with geographical location [17, 21]. Neighbor-joining trees derived from the sequences of *EBNA2* gene showed that all the 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. Phylogenetic trees based on the LMP1 gene and whole EBV genomes indicated that the nine EBVaGC 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 [17]. This is in line with the previous results from the NPC-EBV genomes [14]. In addition, phylogenetic analyses on GDGC1 and GDGC2 derived from specific EBV-encoded gene suggested the presence of at least two parental lineages of EBV, as GDGC1 and GD2 clustered closely, while GDGC2 and GD1 clustered closely [21].

In our recent study, the phylogenetic trees were conducted based on alignment of eight full-length NKTCL-EBVs and previously published 28 strains [23]. Of note, eight NKTCL-EBVs genomes clearly sort into type 1, based on differences in whole genome and especially EBNA2. Eight NKTCL-EBVs were related to other Asian EBV strains, including EBVaGC1–9, HKNPC1–9, GD1, and GD2 obtained from China, and Akata from Japan, whereas none of the specimens was clustered

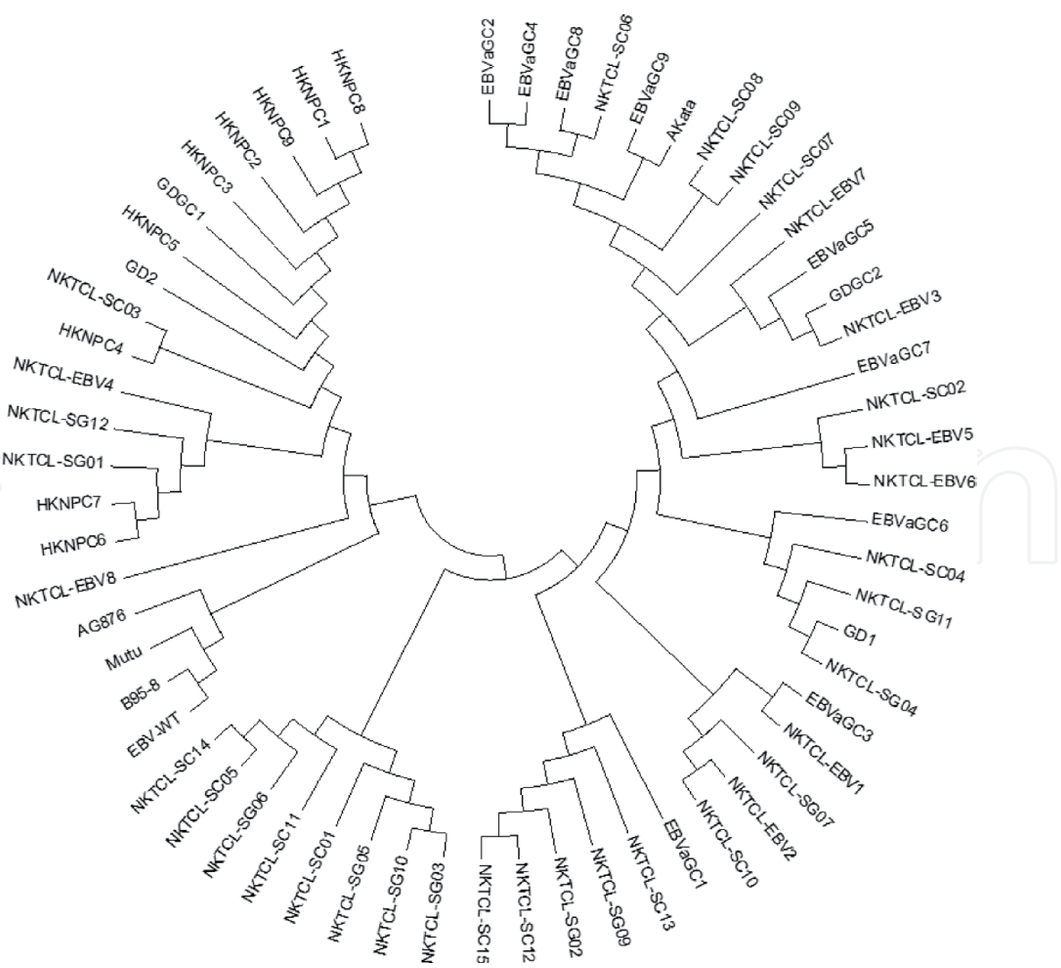


Figure 2.
Phylogenetic trees of EBV genomes. Phylogenetic analyses were conducted using the neighbor-joining (NJ) algorithm implemented in MEGA software (version 6). Bootstrap analysis of 1000 replicates was performed to determine the confidence.

in a branch of non-Asian strains AG876, B95-8, and Mutu. Other group compared the sequences between 27 NKTCL-derived EBV and 164 EBV genome sequences from public database to determine the sequence diversity of EBV [24]. Phylogenetic analysis revealed clear clustering of EBV isolates firstly according to their respective geographic origin; moreover, EBV isolates derived from NKTCL samples tend to cluster closely, apart from clusters by other diseases, supporting the hypothesis of the existence of disease-specific EBV. However, whether the unique EBV has been driving the development of NKTCL or simply adapted to the niche of NKTCL as bystander await further investigations.

In this chapter, phylogenetic analysis was conducted on full-length EBV genomes, including 11 NPC-EBV strains (GD1, GD2, HKNPC1-HKNPC9), 11 GC-EBV strains (EBVaGC1-EBVaGC9, GDGC1-GDGC2), 35 NKTCL-EBV strains (NKTCL-EBV1-NKTCL-EBV8, NKTCL-SC01-NKTCL-SC15, NKTCL-SG01-NKTCL-SG12), B95-8, EBV-WT, Mutu, Akata, and AG876 (**Figure 2**). The result of phylogenetic tree supports the conclusion that EBV infections are more likely affected by different geographic regions rather than particular EBV-associated malignancies.

4. Amino acid changes in $CD4^+$ and $CD8^+$ T-cell epitopes

Sequence variations of EBV genes also result in amino acid epitope exchanges, which should have a significant impact on EBV-specific T-cell immunity.

Among the shared non-synonymous SNVs of the Chinese derived GD1, GD2 and HKNPC1 isolates, 34 are associated with known EBV-specific epitopes; 19 and 15 are found in $CD8^+$ and $CD4^+$ epitopes, respectively [12]. HKNPC2-9 genomes harbored nonsynonymous mutations in epitopes specific for both $CD4^+$ and $CD8^+$ T cells [14]. Amino acid changes were found in seven $CD8^+$ epitopes of LMP2, five epitopes of EBNA3A, and three or fewer in other proteins. Thirteen $CD4^+$ epitopes of EBNA1, six in LMP1, six in LMP2, five in EBNA2, and three or fewer in other proteins contained amino acid changes. Some of the nonsynonymous mutations were affecting multiple epitopes.

EBVaGC shows EBV type I latency neoplasm, in which EBNA1 is expressed in 100% and LMP2A in about half of EBVaGC cases, respectively [45]. Recent studies show that EBNA1, as well as LMP2A, can be presented to both $CD4^+$ and $CD8^+$ T cells, highlighting its potential importance in the development of therapeutic strategies against EBV-associated malignancies [46, 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]. Sequence analysis of the gene encoding EBNA1 in EBV isolates from nine EBVaGC specimens has revealed considerable *EBNA1* sequence divergence from the B95-8 strain [17]. 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.

NKTCL is associated with type II EBV latency, in which only restricted EBV antigens, namely EBNA1, and LMP1 and 2, are expressed [48]. These EBV encoded proteins might be the targets of immune recognition during its persistent infection, and their nonsynonymous variations in $CD4^+$ and $CD8^+$ T-cell epitopes may affect the efficacy for a cytotoxic T lymphocyte (CTL)-based therapy. Many epitopes were defined and were mapped in EBV antigens and correlated with major histocompatibility complex type in previous studies. In our study, we mainly investigated the amino acid changes in $CD4^+$ and $CD8^+$ T-cell epitopes of

EBNA1, LMP1, and LMP2A. Compared with B95-8, amino acids changes were found in 3 $CD8^+$ epitopes of EBNA1, 8 epitopes of LMP1, and 12 epitopes of LMP2A. Eleven $CD4^+$ epitopes of EBNA1, 13 in LMP1, and 9 in LMP2A contained amino acids. Some of the nonsynonymous mutations were affecting multiple epitopes [23]. In another study, alterations of the known T-cell epitopes were examined in EBV sequences derived from NKTCL [24]. Alterations of T-cell epitopes were detected in EBV derived from NKTCL samples. Notably, 21 of these epitopes with significant enrichment in NKTCL samples were restricted to six EBV genes, including EBNA3A (G373D, F325L, I333K, L406P, S412R, H464R, M466R, T585I, and A588P), EBNA3B (A399S, V400L, V417L, K424T, Y662D, and K663E), EBNA3C (P916S), BARF1 (V29A), BCRF1 (V6M), and BNRF1 (G456R, S497G, and A1289T).

Therefore, these data have implications for the development of effective prophylactic and therapeutic vaccine approaches targeting the personalized EBV antigens in these aggressive diseases. Adoptive transfer of cytotoxic T cells (CTLs) specific for EBV antigens has proved safe and effective as prophylaxis and treatment for EBV-associated lymphoproliferative disease. Some patients with advanced stage or relapsed EBV-associated malignancies achieved complete remission after treatment with autologous LMP1/2- and EBNA1-specific CTLs or activated by peptides derived from LMP1/2 [49, 50]. Nonetheless, some cases still did not respond to LMP-CTL therapy, and this failure was usually attributed to immune escapes by antigen loss. It is worth noting that all these previous studies used prototype EBV sequence, B95-8, to design full-length LMP epitopes. Therefore, recent work gives an alternative explanation for the lack of tumor response. Whether changes in such epitopes confer immune evasion of the tumor cells may constitute another hypothesis for future testing.

5. Genomic integration of viral sequences

Viral integration into the host genome has been shown to be a causal mechanism that can lead to the development of cancer [51]. Not surprisingly, known tumor-associated viruses, such as EBV, HBV, HPV16 and HPV18, were among the most frequently detected targets [52]. Notably, the approach of WGS is sensitive to detect viruses. This is particularly true for the common integration verified for HBV, HPV16 and HPV18 in a variety of studies [53–55]. The known causal role of HPV16 and HPV18 in several tumor entities, which triggered one of the largest measures in cancer prevention, has been the motivation for extensive elucidation of the pathogenetic processes involved. Integration events with high confidence were demonstrated for HBV (liver cancer), HPV16 and HPV18 (in both cervical and head-and neck carcinoma), however, low-confidence integration events were detected for EBV (gastric cancer and malignant lymphoma) [56].

Comprehensive analyses of WGS datasets may reveal some novel findings on EBV integration. Recently, a comprehensive survey of EBV integration in a variety of human malignancies, including NPC, EBVaGC, and NKTCL was conducted, using EBV genome capture combined with ultra-deep sequencing, which could efficiently detect integrated EBV sequences from background "noise" introduced by nuclear EBV episomes [27]. The EBV integration rates were 25.6% (10/39), 16.0% (4/25), 9.6% (17/177) in the EBVaGC, NKTCL, and NPC tumors, respectively, which were lower than HPV integration in cervical cancer (76.3%) and head and neck squamous cell carcinoma (60.7%), and HBV in hepatocellular carcinoma (92.6%) [54, 57–59]. They found that EBV

integrations into the introns could decrease the expression of the inflammation-related genes, TNFAIP3, PARK2, and CDK15, in NPC tumors [27]. The EBV integration breakpoints were frequently at oriP or terminal repeats, and were surrounded by microhomology sequences, consistent with a mechanism for integration involving viral genome replication and microhomology-mediated recombination, which has an important role in the integration of other tumorigenic viruses, HBV and HPV [54, 59]. Meanwhile, researchers also observed integrations of short EBV fragments into human chromosomes, coincident with episomal EBV genomes in NKTCL, and showed that 31 EBV-host integration sites were detected from eight NKTCL samples, and enriched in the repeat regions of human genome, such as SINE, LINE, and satellite [24].

However, there are still few studies on EBV integration based on WGS technology. In addition, authors only selected some potential breakpoints to perform PCR and Sanger sequencing for validating. For example, Xu et al. randomly select 12 integrations from 197 breakpoints identified from NPC and other EBV-associated malignancies, and only 10 breakpoints were successfully validated [27]. As integration of EBV sequence into the host genome and the consequent disruption of the important host genes might represent a novel tumorigenesis mechanism in EBV associated malignancies, all the potential EBV integration breakpoints should be validated and biological function of host genes involved should be further conducted.

6. Summary

In conclusion, full-length EBV genomes isolated from primary NPC, EBVaGC, and NKTCL biopsy specimens have been successfully sequenced and the sequence diversity on a whole-genome level has been analyzed, although their pathogenesis remains to be clarified. Phylogenetic analysis has shown that all aforementioned NPC, GC, and NKTCL-EBV strains are type 1 EBV and close to other Asian subtypes, leading to the conclusion that EBV infections are more likely affected by different geographic regions rather than particular EBV-associated malignancies. In addition, sequence variations of EBV genes also result in amino acid epitope exchanges, which should have a significant impact on EBV-specific T-cell immunity. Recent data have provided optimization proposal for selecting EBV genome for treatment from individual patients or at least predominant strains prevalent in geographical regions instead of commonly used B95-8 genome. We acknowledge that further characterizations of the molecular events would provide more information on the exact mechanisms underlying their pathogenic potentials and clinical significance.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China (81903155 to Y.L.), and Beijing Municipal Natural Science Foundation (7202023 to Y.L.), and Beijing Hospitals Authority Youth Program (QML20181106 to Y.L.)

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

Author details

Ying Liu[1], Zheming Lu[2*] and Hongying Huang[3*]

1 Laboratory of Genetics, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, People's Republic of China

2 Laboratory of Biochemistry and Molecular Biology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, People's Republic of China

3 Department of Pathology, New York University Langone Medical Center, New York, USA

*Address all correspondence to: luzheming@bjmu.edu.cn; hongying.huang@nyumc.org

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] Parkin DM. The global health burden of infection-associated cancers in the year 2002. International Journal of Cancer. 2006;**118**(12):3030-3044

[2] Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: A cancer hallmarks analysis. Cell Host & Microbe. 2014;**15**(3):266-282

[3] 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**(Pt 10):1276-1285

[4] Swaminathan S. Noncoding RNAs produced by oncogenic human herpesviruses. Journal of Cellular Physiology. 2008;**216**(2):321-326

[5] Chen SJ, Chen GH, Chen YH, Liu CY, Chang KP, Chang YS, et al. Characterization of Epstein-Barr virus miRNAome in nasopharyngeal carcinoma by deep sequencing. PLoS One. 2010;**5**(9):e12745

[6] Farrell PJ. Epstein-Barr virus. The B95-8 strain map. Methods in Molecular Biology. 2001;**174**:3-12

[7] de Jesus O, Smith PR, Spender LC, Elgueta Karstegl C, Niller HH, Huang D, et al. Updated Epstein-Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV. The Journal of General Virology. 2003;**84**(Pt 6):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**(1):164-170

[9] Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, et al. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. Journal of Virology. 2013;**87**(2):1172-1182

[10] Zeng MS, Li DJ, Liu QL, Song LB, Li MZ, Zhang RH, et al. Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. Journal of Virology. 2005;**79**(24):15323-15330

[11] Liu P, Fang X, Feng Z, Guo YM, Peng RJ, Liu T, et al. Direct sequencing and characterization of a clinical isolate of Epstein-Barr virus from nasopharyngeal carcinoma tissue by using next-generation sequencing technology. Journal of Virology. 2011;**85**(21):11291-11299

[12] Kwok H, Tong AH, Lin CH, Lok S, Farrell PJ, Kwong DL, et al. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. PLoS One. 2012;**7**(5):e36939

[13] Kwok H, Chiang AK. From conventional to next generation sequencing of Epstein-Barr virus genomes. Viruses. 2016;**8**(3):60

[14] Kwok H, Wu CW, Palser AL, Kellam P, Sham PC, Kwong DL, et al. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. Journal of Virology. 2014;**88**(18):10662-10672

[15] Santpere G, Darre F, Blanco S, Alcami A, Villoslada P, Mar Alba M, et al. Genome-wide analysis of wild-type Epstein-Barr virus genomes derived from healthy individuals of the 1,000 genomes project. Genome Biology and Evolution. 2014;**6**(4):846-860

[16] Palser AL, Grayson NE, White RE, Corton C, Correia S, Ba Abdullah MM, et al. Genome diversity of Epstein-Barr virus from multiple tumor types and

normal infection. Journal of Virology. 2015;**89**(10):5222-5237

[17] Liu Y, Yang W, Pan Y, Ji J, Lu Z, Ke Y. Genome-wide analysis of Epstein-Barr virus (EBV) isolated from EBV-associated gastric carcinoma (EBVaGC). Oncotarget. 2016;**7**(4):4903-4914

[18] Wang S, Xiong H, Yan S, Wu N, Lu Z. Identification and characterization of Epstein-Barr virus genomes in lung carcinoma biopsy samples by next-generation sequencing technology. Scientific Reports. 2016;**6**:26156

[19] Tu C, Zeng Z, Qi P, Li X, Yu Z, Guo C, et al. Genome-wide analysis of 18 Epstein-Barr viruses isolated from primary nasopharyngeal carcinoma biopsy specimens. Journal of Virology. 2017;**91**(17):e00301-17

[20] Yao Y, Xu M, Liang L, Zhang H, Xu R, Feng Q, et al. Genome-wide analysis of Epstein-Barr virus identifies variants and genes associated with gastric carcinoma and population structure. Tumour Biology. 2017;**39**(10):1010428317714195

[21] Chen JN, Zhou L, Qiu XM, Yang RH, Liang J, Pan YH, et al. Determination and genome-wide analysis of Epstein-Barr virus (EBV) sequences in EBV-associated gastric carcinoma from Guangdong, an endemic area of nasopharyngeal carcinoma. Journal of Medical Microbiology. 2018;**67**(11):1614-1627

[22] Hong S, Liu D, Luo S, Fang W, Zhan J, Fu S, et al. The genomic landscape of Epstein-Barr virus-associated pulmonary lymphoepithelioma-like carcinoma. Nature Communications. 2019;**10**(1):3108

[23] 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. The Oncologist. 2019;**24**(9):e905–e913

[24] Peng RJ, Han BW, Cai QQ, Zuo XY, Xia T, Chen JR, et al. Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. Leukemia. 2019;**33**(6):1451-1462

[25] Xu M, Yao Y, Chen H, Zhang S, Cao SM, Zhang Z, et al. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. Nature Genetics. 2019;**51**(7):1131-1136

[26] Hui KF, Chan TF, Yang W, Shen JJ, Lam KP, Kwok H, et al. High risk Epstein-Barr virus variants characterized by distinct polymorphisms in the EBER locus are strongly associated with nasopharyngeal carcinoma. International Journal of Cancer. 2019;**144**(12):3031-3042

[27] Xu M, Zhang WL, Zhu Q, Zhang S, Yao YY, Xiang T, et al. Genome-wide profiling of Epstein-Barr virus integration by targeted sequencing in Epstein-Barr virus associated malignancies. Theranostics. 2019;**9**(4):1115-1124

[28] Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet. 2019;**394**(10192):64-80

[29] de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: A worldwide incidence analysis. The Lancet Global Health. 2020;**8**(2):e180–e190

[30] Mahdavifar N, Ghoncheh M, Mohammadian-Hafshejani A, Khosravi B, Salehiniya H. Epidemiology and inequality in the incidence and mortality of nasopharynx cancer in Asia. Osong Public Health and Research Perspectives. 2016;**7**(6):360-372

[31] Young LS, Yap LF, Murray PG. Epstein-Barr virus: More than 50 years old and still providing surprises. Nature Reviews Cancer. 2016;**16**(12):789-802

[32] Shibata D, Weiss LM. Epstein-Barr virus-associated gastric adenocarcinoma. The American Journal of Pathology. 1992;**140**(4):769-774

[33] Tokunaga M, Land CE, Uemura Y, Tokudome T, Tanaka S, Sato E. Epstein-Barr virus in gastric carcinoma. The American Journal of Pathology. 1993;**143**(5):1250-1254

[34] Takada K. Epstein-Barr virus and gastric carcinoma. Molecular Pathology. 2000;**53**(5):255-261

[35] 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**(3):824-833

[36] Fukayama M, Ushiku T. Epstein-Barr virus-associated gastric carcinoma. Pathology, Research and Practice. 2011;**207**(9):529-537

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

[38] Chang KL, Chen YY, Chen WG, Hayashi K, Bacchi C, Bacchi M, et al. EBNA-1 gene sequences in Brazilian and American patients with Hodgkin's disease. Blood. 1999;**94**(1):244-250

[39] 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**(7):877-882

[40] Xiong J, Zhao W. What we should know about natural killer/T-cell lymphomas. Hematological Oncology. 2019;**37**(Suppl 1):75-81

[41] Halabi MA, Jaccard A, Moulinas R, Bahri R, Al Mouhammad H, Mammari N, et al. Clonal deleted latent membrane protein 1 variants of Epstein-Barr virus are predominant in European extranodal NK/T lymphomas and disappear during successful treatment. International Journal of Cancer. 2016;**139**(4):793-802

[42] Nagamine M, Takahara M, Kishibe K, Nagato T, Ishii H, Bandoh N, et al. Sequence variations of Epstein-Barr virus LMP1 gene in nasal NK/T-cell lymphoma. Virus Genes. 2007;**34**(1):47-54

[43] Li Z, Xia Y, Feng LN, Chen JR, Li HM, Cui J, et al. Genetic risk of extranodal natural killer T-cell lymphoma: A genome-wide association study. The Lancet Oncology. 2016;**17**(9):1240-1247

[44] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 2011;**28**(10):2731-2739

[45] Chen JN, Ding YG, Feng ZY, Li HG, He D, Du H, et al. Association of distinctive Epstein-Barr virus variants with gastric carcinoma in Guangzhou, southern China. Journal of Medical Virology. 2010;**82**(4):658-667

[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.

The Journal of Infectious Diseases. 2008;**197**(11):1594-1597

[47] Icheva V, Kayser S, Wolff D, Tuve S, Kyzirakos C, Bethge W, et al. Adoptive transfer of epstein-barr virus (EBV) nuclear antigen 1-specific t cells as treatment for EBV reactivation and lymphoproliferative disorders after allogeneic stem-cell transplantation. Journal of Clinical Oncology. 2013;**31**(1):39-48

[48] Fox CP, Haigh TA, Taylor GS, Long HM, Lee SP, Shannon-Lowe C, et al. A novel latent membrane 2 transcript expressed in Epstein-Barr virus-positive NK- and T-cell lymphoproliferative disease encodes a target for cellular immunotherapy. Blood. 2010;**116**(19):3695-3704

[49] Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. Journal of Clinical Oncology. 2014;**32**(8):798-808

[50] Cho SG, Kim N, Sohn HJ, Lee SK, Oh ST, Lee HJ, et al. Long-term outcome of Extranodal NK/T cell lymphoma patients treated with postremission therapy using EBV LMP1 and LMP2a-specific CTLs. Molecular Therapy. 2015;**23**(8):1401-1409

[51] Tang KW, Larsson E. Tumour virology in the era of high-throughput genomics. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2017;**372**(1732):20160265

[52] Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: A synthetic analysis. The Lancet Global Health. 2016;**4**(9):e609-e616

[53] Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Research. 2012;**22**(4):593-601

[54] Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. Nature Genetics. 2015;**47**(2):158-163

[55] Liu Y, Lu Z, Xu R, Ke Y. Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology. Oncotarget. 2016;**7**(5):5852-5864

[56] Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. Nature Genetics. 2020;**52**(3):320-330

[57] Koneva LA, Zhang Y, Virani S, Hall PB, McHugh JB, Chepeha DB, et al. HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. Molecular Cancer Research. 2018;**16**(1):90-102

[58] Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nature Genetics. 2012;**44**(7):765-769

[59] Zhao LH, Liu X, Yan HX, Li WY, Zeng X, Yang Y, et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. Nature Communications. 2016;**7**:12992