Decoding critical targets and signaling pathways in EBV-mediated diseases using large language models

Jingwen Yu[1], Yaohao Wang[1], Haidong Wang[1], Zhi Wei[2], Yonggang Pei[1, *]

[1] School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen, Guangdong, China

[2] Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA

Running title: Utilizing large language models to analyze EBV-mediated diseases

Keywords: EBV, Antigens, Pathways, LLMs

Corresponding author:

Yonggang Pei Ph.D.

Assistant professor

School of Public Health and Emergency Management

Southern University of Science and Technology

Email: peiyg@sustech.edu.cn

Tel: 86-0755-88015615

Abstract

Epstein-Barr virus (EBV), a member of the gamma-herpesvirus, is the first identified human oncovirus and is associated with various malignancies. Understanding the intricate interactions between EBV antigens and cellular pathways is crucial to unravel the molecular mechanisms in EBV-mediated diseases. However, fully elucidating EBV-host interactions and the associated pathogenesis remains a significant challenge. In this study, we utilized LLMs to screen 36,105 EBV-relevant scientific publications, presenting a descriptive overview of the interactions between EBV antigens and host cellular pathways through comprehensive interaction networks. We described the critical roles of EBV antigens by constructing functional subsets of host proteins targeted by EBV antigens and illustrated the interactions using protein-protein interaction (PPI) networks. Furthermore, we developed an antigen-pathway network that highlights the connections with EBV-associated diseases, including DLBCL, BL, NPC, gastric cancer (GC), and post-transplant lymphoproliferative disorders (PTLD). Utilizing our dataset and public dataset, we validated the efficacy of using BLLF3-targeted TLR2-associated factors as marker genes for DLBCL. Next, we confirmed the co-expression of calcium pathway factors induced by LMP1 over-expression in BL. Finally, based on shared results suggesting that LMP1 actively regulates the glycolysis pathway, we identified and validated the correlation and co-expression of LMP1-induced PARP1, HIF1A, HOXB3, and key transcription-related factors, revealing the complete picture of LMP1's influence on the glycolysis pathway. Our study presents a comprehensive functional encyclopedia of the interactions

between EBV antigens and host signaling pathways across various EBV-associated disease contexts, providing valuable insights for the development of therapeutic strategies.

Introduction

Epstein-Barr virus (EBV) was discovered in 1964 and has infected over 90% of the world's population[1]. The oncovirus contributes to various human diseases, such as Burkitt lymphoma (BL), nasopharyngeal carcinoma (NPC), Hodgkin lymphoma (HL), diffuse large B cell lymphoma (DLBCL), multiple sclerosis (MS), and so on. In the 60 years since its discovery, the intricate roles of EBV in regulating interconnected host cellular pathways have provided valuable insights into the mechanisms underlying EBV-related disease progression. EBV infection in human cells triggers a series of changes in cellular biology including cell cycle transition, cell apoptosis, DNA mismatch repair, and multiple cellular signaling pathways. Such modulations are mediated by various EBV antigens, promoting viral latency and oncogenesis. For instance, EBV-encoded Latent Membrane Protein 1 (LMP1) mimics CD40 signaling via its two C-terminal activating regions, namely CTAR1 and CTAR2, which are shown to activate both the canonical and noncanonical NF-κB pathways[2–5]. EBV nuclear antigen 1 (EBNA1) inhibits the canonical NF-κB pathway by inhibiting the phosphorylation of IKKα/β[6]. Although the NF-κB pathway serves as a key factor in EBV latent infection and oncogenesis, the intricate and intertwined network of EBV antigens and NF-κB factors presents a significant challenge for the development of

anti-EBV therapeutic strategies. Therefore, it is imperative to comprehensively examine the landscape of current studies on EBV antigens and their interactions with cellular signaling pathways to further explore the EBV-mediated pathogenesis.

Traditional literature mapping techniques, including systematic reviews and meta-analyses, often fall short of capturing the full breadth of EBV literature. However, the rapid development in machine learning offers promising tools for analyzing potential key factors in EBV-mediated pathogenesis from vast amounts of literature. A recent study characterizes the biomedical literature into a 2D map[7]. The embedding contains 21 million papers, presenting the unlimited possibilities of LLMs-assisted literature study. Another work introduced a knowledge graph of the m6A regulatory pathway utilizing GPT-4 to elucidate m6A's regulatory mechanisms in cancer phenotypes across various cancers[8]. This burgeoning technology has catalyzed a proliferation of studies exploring its capabilities and applications in the fields of biological and medical research, driving innovation and prompting ongoing discourse on its potential impacts and future developments.

Our study aims to utilize LLMs for the first time to analyze decades of research on how EBV antigens regulate cellular factors and signaling pathways. We present a comprehensive framework of EBV antigens, their interacting host targets, and associated pathways, constructing an extensive network that delineates virus-host interaction. We then mapped the diverse background of EBV-related diseases for antigen-pathway relationships and revealed the critical interactions between EBV antigens and cellular pathways in EBV-mediated diseases. Collectively, this study

offers a unique overview of current EBV-host interactions and provides valuable resources for the development of anti-EBV therapeutic strategies in the future.

Methods

Data sources

The data in this study were obtained from the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) using EDirect. The initial data collection was performed on February 28, 2024. The final dataset comprises 36,105 EBV-related articles, including only original articles and reviews published in English. The RNA-seq dataset employed to investigate differentially expressed genes induced by LMP1 overexpression was obtained from the Gene Expression Omnibus (GEO) database under accession number GSE247495. Differential gene expression analysis was conducted using the R package DESeq2 (1.40.2). Genes with an adjusted p-value (padj) less than 0.05 and a log2 fold change greater than log2(2) were considered significantly differentially expressed.

ChatGPT-4 prompt engineering

To ensure accuracy, we provided ChatGPT with the lists of EBV antigens and EBV-related diseases through system messages. The prompt for antigen-target-pathway summary is as follows: “If the abstract concludes that EBV-encoded proteins affect another protein or pathway, show me only the EBV protein, affected protein and pathway. If not, return 'Error' and nothing else. Use a temperature of 0. All EBV

proteins were listed before. The abstract: XXX”. The prompt for EBV-related disease mapping: “Return which specific EBV-related disease does the abstract attached in the end investigates. If unknown, return 'Error' and nothing else. The abstract: XXX”.

To minimize costs, we initially utilized the free version GPT-3.5 Turbo via the OpenAI API. Items not resulting in "Error" were then processed using GPT-4. Next, we employed synonym substitution to standardize the naming of various aliases. For instance, “EBV nuclear antigen 1”, “EBNA1” and “EBNA-1” are all annotated as “EBNA1”; “EB2”, “Mta”, “SM” and “BS-MLF1” are all annotated as “BMLF1”. (Table S1)

GPT-4 generated data evaluation

To evaluate the accuracy of the GPT-generated results, we randomly selected a total of 115 entries for manual inspection. The inspection focused on the EBV antigens mentioned in the corresponding articles, the host factors associated with these antigens, and the annotated KEGG pathways. The criteria for a "correct result" are as follows: (1) the article clearly describes the interaction between the EBV antigen and its corresponding host factors; (2) the article identifies the biological changes influenced by this interaction and their relevance to the annotated KEGG results.

Annotation from GPT summarized pathways to KEGG database

The results generated by LLMs are highly variable. To establish a unique

standard for readability, we annotated the LLM-generated pathway outputs to the KEGG pathway database due to its simplicity, accessibility, and widespread use within the academic community. This mapping was performed based on our customized R script and empirical manual substitution. A Sankey diagram shows a complete mapping of the substitution (Table S21).

Gene enrichment analysis

LLM-acquired target host protein annotation is performed using the R package STRINGdb (2.12.1). Annotated IDs are submitted to UniProt for symbol conversion (https://www.uniprot.org/). The generated list of genes of interest (GOIs) was subjected to gene enrichment analysis using three primary tools: the KEGG database[9], Reactome (https://reactome.org/), and STRINGdb.

Omics analysis

PCA reduction and cohort analyses are performed using UCSC Xena[10] (https://xena.ucsc.edu/) and GEPIA2 (http://gepia2.cancer-pku.cn/)[11]. The dataset comprised TCGA Diffuse Large B-Cell Lymphoma, GTEx Spleen and GTEx Blood. LMP1 overexpression RNA-seq data is normalized and analyzed using DESeq2.

Visualization

The figures were created with BioRender (https://biorender.com), the R package wordcloud2 (0.2.1), circlize (0.4.16), or pheatmap (1.0.12) with customized scripts.

PPI network is visualized in Cytoscape (3.9.1) and colored based on protein family types. Other visualizations are made by a combination of R packages including ggplot2 (3.5.0), ggalluvial (0.12.5), and enrichplot (1.20.3).

Results

LLMs-assisted data mining reveals the key insights in EBV research

To create a literature pool for information extraction, our initial dataset comprised 36,105 EBV-related research articles available in the PubMed database as of the start of this project (February 28, 2024). These publications span a temporal range of 57 years, from 1968 to 2024 (Figure 1a). Overall, there has been a steady increase in the number of research articles related to EBV over these years, with an average of approximately 1000 publications annually.

Next, to better characterize the current EBV studies, we summarized the available articles into a table of PubMed ID, EBV antigens, target proteins, and the related signaling pathways using ChatGPT-4. The returned dataset contained information from 2,827 publications related to EBV antigen-pathway interactions and was subsequently subjected to rigorous data reduction and filtering (Figure 1b). The final dataset includes 2,289 documented entries of EBV antigens and their associated pathways, derived from 1,839 publications (Figure 1b, Table S3). To validate the accuracy of the information extracted using ChatGPT, we manually compared a random subset of 115 entries (~5%) of the GPT-4 generated summaries with the corresponding original abstracts. The results indicate that our prompts achieved an

average accuracy of 93.04% (Figure 1c, Table S4), demonstrating that our GPT-4-based in-house pipeline can effectively analyze additional literature.

Analysis of EBV antigens and their impacts on cellular pathways reveals that LMP1 is the most frequently studied antigen, followed by EBNA1, BZLF1, LMP2A, and EBNA2, which are primarily latent or lytic proteins (Figure 1d). Throughout the 2010s, the significance of EBNA3C has increased, making it the fourth most studied antigen (Figure 1e). The most studied pathways focus on the mechanisms underlying EBV infection (Figure 1f). Pathways involving EBV-mediated cell apoptosis are the next most thoroughly studied areas. Abundant studies concentrate on the oncogenic mechanism of EBV infection, mainly on how EBV antigens regulate cell cycle and replication. The most extensively studied mechanisms include the NF-κB signaling pathway, MAPK signaling pathway, TLR signaling pathway, PI3K/Akt signaling pathway, TNF signaling pathway, and TGF-β signaling pathway (Figure 1f). On the temporal scale, research on cell apoptosis, NF-κB signaling pathway, cell cycle, and MAPK signaling pathway have always been the four unyielding topics (Figure 1g), suggesting their critical roles in EBV-mediated diseases.

Pathway mapping reveals distinct signatures of EBV antigens

To clarify the critical pathways that EBV antigens regulate, we manually annotated the GPT-generated pathways to the KEGG pathway entries (Table S2). The chord diagram denotes the frequency of top 33 most studied EBV antigens and correlated pathways, indicating an intricate and interconnected network of cellular

pathways induced by EBV infection (Figure 2a). To clearly illustrate the concentrated interactions between EBV antigens and their related pathways, we present a chord diagram as a heatmap (Figure 2b). In this figure, the pathways are clustered using the maximum linkage clustering method, and EBV antigens are arranged empirically based on their roles in EBV latent or lytic infection. We observed that EBV latent membrane proteins, LMP1 and LMP2A, exhibit similar interaction patterns across different host cellular pathways, although research on LMP1 has been more thorough as previously discussed (Figure 1d). The resemblance of these two EBV antigens lies in the shared role in their ability to reprogram host immune response through the MAPK, PI3K/Akt, or TNF pathways, and their oncogenic functions in regulating the p53 or PD1/PD-L1 checkpoint pathway. Despite the significant scarcity of LMP2A-related publications compared to LMP1, the studies collectively indicate that LMP2A contributes more to the BCR signaling pathway and the Wnt pathway than LMP1 (Figure 2b).

EBV nuclear antigens (EBNAs) contribute to the establishment and maintenance of viral latency, especially EBNA1, which is present in all EBV latent programs except for latency 0. EBNA1 directly binds to the host genome and is essential for the replication and maintenance of EBV episome[1]. It also serves as a transcription factor regulating multiple gene expression. Our analysis shows that EBNA1 differs from other EBNAs in terms of regulating DNA replication, and oxidative phosphorylation. EBNA2 plays a crucial role in the initial stage of EBV-mediated B-cell transformation. Interestingly, EBNA2 has been confirmed to bind to super-enhancer sites that are

critical for the maintenance of EBV latency[12,13]. Our findings highlight the distinctive role of EBNA2 in cAMP signaling pathways, fatty acid biosynthesis, and cell adhesion lipid molecules, indicating its potential to facilitate B-cell transformation and proliferation through lipid metabolism. Similar to EBNA1, EBNA3A and EBNA3C are shown to interfere with the host genome and promote immune evasion, cell cycle control, and inhibition of apoptosis as transcription factors. EBNALP binds to EBNA2 and functions as a coactivator to promote the transformation of B-cells during EBV infection[1]. For immediate-early lytic genes that are responsible for EBV reactivation, both BZLF1 and BRLF1 are well-studied. BZLF1 activates lytic gene expression, initiating the lytic cycle, while BRLF1 synergizes with BZLF1 to promote lytic gene transcription, leading to viral replication and release. Studies on BZLF1 seem more diverse on multiple cellular oncogenic pathways.

Other EBV antigens have received relatively less attention compared to the previously mentioned latent and lytic antigens (Table S5). EBV glycoproteins play roles in infection and entry into host cells. Glycoprotein L (gL) cooperates with gH and gB to form the infection complex, mediating EBV-cell fusion and entry.

Gene enrichment uncover defining characteristics of EBV-induced cellular modulations

To further investigate the associated diseases or pathways from the level of target proteins, we annotated the EBV research literature on a higher resolution after enrichment of GPT-generated target proteins using the KEGG database and the

Reactome database. Disease-wise, KEGG enrichment revealed that EBV infection may be associated with traits typical of a wide range of diseases, including bladder cancer, pancreatic cancer, colorectal cancer, small cell lung cancer, and others (Figure 3a). The results suggest a possible correlation between these malignant diseases and EBV infection. However, it's important to note that the KEGG database does not include EBV-related diseases like Burkitt lymphoma. Consequently, the diseases highlighted are primarily autoimmune-related and various forms of cancer (Figure 3a).

The pathways enriched solely from target proteins and those retrieved from literature exhibit notable similarities, with extensive studies on pathways such as apoptosis, VEGF signaling pathway, TNF signaling pathway, Toll-like receptor signaling pathway, and IL-17 signaling pathway (Figure 3b). This indicates the coherence between GPT-generated pathways and the original articles. The last three pathways highlight the importance of studying how innate immunity is affected by viral antigens. The IL-17 signaling pathway triggers downstream NF-κB signaling pathway and is known to be a key pathogenic factor of autoimmune diseases[14]. However, the relationship between the IL-17 pathway and EBV infection remains unexplored. The high expression of IL-17 in multiple sclerosis patients could present a novel area of EBV research and provide druggable candidates for certain autoimmune diseases[15].

To provide a more comprehensive overview of the functions of EBV-associated target proteins within their host, we used the Reactome database for annotation and identified 375 biological processes ($p<0.05$) related to EBV infection (Figure 3c).

This network shows that EBV-targeted host proteins predominantly form clusters in approximately 10 functional pathway groups. The largest clusters include innate immunity pathways, represented by the TNF signaling pathway, Toll-like receptor signaling cascades, and NF-κB signaling pathways. These are followed by cell growth and differentiation pathways, exemplified by the PI3K/AKT signaling pathways and related factors such as interleukins. Subsequently, DNA replication and mismatch repair, along with cell cycle-related pathways, are prominently featured, particularly those regulated by TP53. Moreover, the NOTCH signaling pathway and its regulation of the TGF-β signaling pathways mediated by the SMAD proteins are also critical for cell proliferation, differentiation, transitions, and apoptosis in our model of cellular signaling crosstalk[16].

Intriguingly, we have identified some active mediators bridging the interconnected cellular signaling crosstalk (Figure 3c). These active members include deubiquitination and sumoylation that regulate protein modification and proteostasis; TP53-mediated cellular senescence on DNA replication and cell cycle; PTEN and downstream AKT signaling activation mediated by PIP3; BCR signaling for multiple pathways spanning the immune response to cell proliferation; Fc epsilon receptor (FCERI) signaling for the same function. Targets related to SARS-CoV-2 infection are enriched alongside the immune signaling pathways, suggesting a shared immune response elicited by EBV and SARS-CoV-2. This indicates the potential for developing broad-spectrum antivirals for both viruses by targeting shared oncogenic factors. Additionally, more discrete, smaller clusters were identified, including the

Wnt signaling pathway, mRNA editing, as well as the unfolded protein response (UPR) essential for maintaining cellular homeostasis and survival under conditions of protein misfolding and ER stress.

Analysis of host protein-protein interactions identifies key mediators in EBV-mediated diseases

To further identify critical cellular factors involved in EBV-associated diseases, we focused on EBV-targeted host proteins independent of pathway enrichment, which may provide insights into key target proteins. Bibliographic analysis of publications over time reveals trends in EBV-related host proteins (Figure 4a). p53 and c-Myc are the two most studied proteins across all periods. Genomic profiling has identified frequent TP53[17–19] and MYC[20,21] mutations or rearrangements in EBV-positive malignancies. CD23 and CD40 are two critical factors for EBV infection on the B-cell membrane. Earlier studies identified that EBNA2 specifically induces CD23 expression vital for B-cell transformation[22] and verified the role of LMP1 in viral latency maintenance via mimicry of the CD40 signaling pathways[23]. As previously discussed in Figure 2, multiple EBV antigens including BHRF1 and LMP1 are associated with Bcl-2 related apoptosis in lymphoma and immortalized B-cells[19,24,25].

To better characterize the pathways and networks of EBV-related host factors described above, we annotated all the target proteins using the STRING database and performed protein-protein interaction (PPI) analysis. The top 3 enriched clusters are shown (Figure 4). The first PPI cluster centers around TP53 genes, apoptotic

pathways, mTOR signaling, and the MAPK pathway cascades along with gene silencing mediated by DICER1 and m6A methylation (Figure 4b). This cluster reflects the role of EBV in suppressing the host apoptotic pathways and related signaling transduction pathways to promote viral latency and oncogenesis. Clinical trials and genomic profiling studies have identified the downregulation and mutation of TP53 in multiple EBV-associated cancers[26–29]. EBV-encoded microRNAs (EBERs) are shown to help maintain Burkitt's lymphomas by targeting caspase 3 to inhibit apoptosis and aid the transformation of naïve B-cells[30]. Such inhibition usually works in harmony with other factors like Bcl-2. Besides triggering classic apoptotic pathways, EBV also bypasses the host antiviral systems by decreasing host Dicer protein via EBNA1-mediated microRNAs degradation[31]. Recent studies have also uncovered the role of circular RNAs in promoting the progression of EBV-associated gastric carcinoma through the transactivation of METTL3[32]. Along this pathway, caspases have been found to inhibit the m6A RNA modification process, thereby promoting EBV replication[33]. Additionally, EBV infection is known to induce TCL1 gene expression in Burkitt's lymphoma, which also serves as an activator for the Akt pathway[34].

The second cluster highlights EBV's roles in transcriptional regulation and epigenetic modification, particularly through transcriptional factors (TFs) (Figure 4c). MYC is a key TF associated with multiple cancers including EBV-associated lymphomas[35]. Moreover, MYC abundance represses EBV reactivation and controls the lytic switch in Burkitt lymphoma[36]. 3D genome mapping of EBV-positive lymphoblastoid cell lines (LCLs) revealed distinctive EBV super-enhancer (SE)

targets like MCL1, IRF4, and EBF[37]. Bioinformatic analysis identified that H3K27ac SEs overlap with STAT5 and NFAT targets, indicating their roles in promoting EBV latency[13]. On a larger scale, broad screening of TFs integrated over 700 high-throughput sequencing datasets and provides an atlas of EBV-related transcriptomic and epigenetic host-virus regulatory interactions[38]. These EBV-specific TFs, such as SPI1, EBF1, and PAX5, collectively contribute to EBV-induced oncogenesis and latency[13].

The last cluster primarily comprises factors related to the innate immune response, including cellular receptors, cytokines, and chemokines that regulate immune cell activation, inflammatory response, and the tumor microenvironment (TME) (Figure 4d). EBV orchestrates immune evasion by cellular receptors like CD40, CD80, IL4, IL6, IL10 and their ligands[39,40]. For example, LMP1 mimics CD40 to activate downstream NF-κB pathways to induce B-cell proliferation[23]. After infection, EBV hijacks host signaling transduction factors like STATs[41,42] and IRFs[43] to promote cell survival and proliferation. LMP1 and LMP2A then activate the JAK/STAT pathway, to facilitate cell growth factors and anti-apoptotic signaling. Meanwhile, EBV and other herpesviruses induce the PI3K/Akt pathway to aid viral infection, latency, and reactivation[44]. This cluster is also rich in enzymes, multiple cellular growth factors (CGFs), and extracellular matrix (ECM) proteins, which represent the invasiveness of EBV-infected cells. Early research discovered that viral LMP1 and BZLF1 proteins up-regulate the expression and activity of matrix metalloproteinase 1 (MMP1), and thereby confer the invasive properties of the cells[45]. Besides common growth factors,

we also identified Angiotensin-Converting Enzyme 2 (ACE2) in this cluster. Study indicates that the lytic replication of EBV induces ACE2 expression in human epithelial cells, which enhances the entry of SARS-CoV-2[46].

Exploring the critical relationships in EBV-mediated diseases

Due to the complex background of EBV pathogenesis, we next seek to elucidate the influence of EBV antigens on host pathways in the context of multiple EBV-associated malignancies (Figure 5a-b). In this section we demonstrate how our dataset and generated PPI networks can be utilized to assist EBV-host interaction studies using basic omics analyses. We plot the connections identified for EBV antigens and host pathways onto circular connection hubs for DLBCL, BL, NPC, GC, and PTLD (Figure 5c, Figure 6a and Figure S1).

Approximately 9% of DLBCLs are EBV positive with type II lantency[1]. In DLBCL-centered hubs, we noticed the special linkage of EBV encoded dUTPase BLLF3 and the TNF signaling pathways (Figure 5c). This indicates that BLLF3 induce DLBCL through the TNF pathways. dUTPase is crucial for maintaining the balance of intracellular nucleotides, particularly during DNA replication[47]. Our dataset pointed to study exploring BLLF3 and its interaction with host factor TLR2[48]. Our PPI networks show several known key factors closely interacting with TLR2 in the background of EBV infection, such as IRAK1, HMGB1, TLR9, TNF, NFKB1, IL1B and IL10. Next, we conducted pan-cancer analysis to examine the utility of these proteins in distinguishing between DLBCL tumor and normal samples (Figure

5d). The results indicated clear difference between the tumor and samples from normal tissues, suggesting their potential as marker genes for DLBCL diagnosis. As DLBCL does not have corresponding gene expression data for normal tissue in TCGA, we used whole blood and spleen, both of which are immune cells, as normal samples for comparison. Taken apart, we noticed the high expression of HMGB1, TNF, NFKB1 and IL10 in DLBCL tumor samples which function in TNF pathway (Figure 5e). At last, we examined the correlation of TNF expression and DLBCL patients' survival probability (Figure 5f). The results show the survival possibility of patients with high TNF expression dropped to 75% in approximately four years while samples with low expression survived 100%. These findings above show the potential in studying BLLF3 mediated TNF pathway activation in the carcinogenesis of DLBCL.

LMP1 is the most studied EBV antigens in our literature pool (Figure 5b). In BL, literature indicates that LMP1 activates the calcium signaling pathway by overexpressing the host factor CAMK4[49] (Figure 6a, Table S3). Our PPI network shows that HDAC4 and CREB1 directly interact with CAMK4, and the secondary factors that interact with these two, such as MEF2D, RUNX3, RUNX2, NFATC2, JUND, FOS, RELA, TP53, TAF4, MECP2, CTNNB1, ATF4, and CRTC2 (Figure 4c). We observed that although a dox-induced overexpression of LMP1 in the Akata cells did not result in a significant increase in CAMK4, its interacting proteins listed above exhibited a co-expression pattern and consistent upregulation (Figure 6b, c). This suggests that the calcium signaling pathway may play a role in the oncogenesis of BL by a LMP1-dependent manner.

Finally, we locate studies working on similar pathways associated with the same antigens across cell lines from different EBV-related malignancies to compensate the utility. For instance, studies on BL and NPC both demonstrated that LMP1 is associated with glycolysis (Figure 6a, Figure S1). EBV typically establishes type I latency in BL. Overexpressing LMP1 in BL cell lines suggests that LMP1 activates PARP1 to induce glycolysis and promote oncogenesis in a HIF-1α-dependent manner[45]. Study in NPC cell lines demonstrates that LMP1 represses Hox genes via stalling the key mediators of DNA transcription like RNA pol II to modulate glycolysis[51]. These studies in cell lines with different EBV latencies demonstrate the role of LMP1 in regulating glycolysis. We next explore the transcriptomic correlation for PARP1, HIF1A, HOXB3, and key transcription-related factors (Figure 6d-e). Result show high correlation, which suggests a more complete mechanism that LMP1 activates PARP1, which subsequently binds to HIF-1α and RNA pol II and activates HoxB3, finally modulate glycolysis.

Discussion

This study, for the first time, presents a mapping of EBV antigens, host factors, and cellular pathways from available EBV-related articles using LLMs. We have summarized the major pathways associated to each EBV antigens, and provided a descriptive overview of core PPI clusters. At last, we provided examples of ustilizing our dataset for marker gene analyses.

However, certain limitations in our research need to be addressed. While our

initial data pool contained 36,105 EBV-related publications, our strict prompts and data filtering reduced this to 2,289 entries, encompassing 1,839 unique publications (~5.1%) for downstream analyses. If loosening the criteria, we will be able to obtain more entries for deeper analyses, although the accuracy may decrease. Secondly, our literature study was based solely on conclusions drawn from abstracts of the available publications which has its intrinsic limitation in the study of network biology. Although we used the frequency of topics as a proxy for their importance in the EBV antigen-pathway mapping, the sheer number of publications may not fully represent actual biological significance. Occasionally, a higher citation rate may influence their proportions, potentially translating non-experimental results. Future efforts may focus on using supervised machine learning algorithms to explore the transcriptomic networks of EBV-induced cellular changes for enhanced utility and usability. In the last part, we showcase through basic omics analyses that our EBV literature analysis bears the potential to inspire researchers to engage with former published work of scientists worldwide for novel effort in illustrating the mechanism of EBV-host interaction. Finally, EBV latency types are critical in understanding EBV gene expression and virus-host interaction. However, such information is largely unattainable in abstracts. Future research based on more powerful tools can extract the main text for related information.

This study represents a pioneering effort in the application of LLMs in EBV-related research, highlighting the potential of LLMs in advancing our understanding of viral interactions and cellular pathways. Unlike broad-scale context generation, our

study clarifies the relationships between EBV antigens and various cellular pathways based on the manual annotation of GPT-generated summaries, presenting networks of EBV-related host proteins for future research. Beyond EBV, the big data era and advancements in various LLMs underscore the urgent need for advanced, digitized platforms in all areas of viral research. Such tools are pivotal for accelerating research progress, advancing basic virology, and fostering the development of novel antivirals.

References

1. Damania, B., Kenney, S. C. & Raab-Traub, N. Epstein-Barr virus: Biology and clinical disease. *Cell* S0092867422011217 (2022) doi:10.1016/j.cell.2022.08.026.

2. Hatzivassiliou, E., Miller, W. E., Raab-Traub, N., Kieff, E. & Mosialos, G. A fusion of the EBV latent membrane protein-1 (LMP1) transmembrane domains to the CD40 cytoplasmic domain is similar to LMP1 in constitutive activation of epidermal growth factor receptor expression, nuclear factor-kappa B, and stress-activated protein kinase. *J Immunol* **160**, 1116–1121 (1998).

3. Kaykas, A. & Sugden, B. The amino-terminus and membrane-spanning domains of LMP-1 inhibit cell proliferation. *Oncogene* **19**, 1400–1410 (2000).

4. Gewurz, B. E. *et al.* Genome-wide siRNA screen for mediators of NF-κB activation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2467–2472 (2012).

5. Luftig, M. *et al.* Epstein–Barr virus latent membrane protein 1 activation of NF-κB through IRAK1 and TRAF6. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15595–15600 (2003).

6. Valentine, R. *et al.* Epstein-Barr virus-encoded EBNA1 inhibits the canonical NF-

463 kappaB pathway in carcinoma cells by inhibiting IKK phosphorylation. *Mol Cancer* **9**, 1
464 (2010).
465 7. González-Márquez, R., Schmidt, L., Schmidt, B. M., Berens, P. & Kobak, D. The
466 landscape of biomedical research. *Patterns* 100968 (2024)
467 doi:10.1016/j.patter.2024.100968.
468 8. Wu, X. *et al.* reguloGPT: Harnessing GPT for Knowledge Graph Construction of
469 Molecular Regulatory Pathways. Preprint at https://doi.org/10.1101/2024.01.27.577521
470 (2024).
471 9. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*
472 *Acids Res* **28**, 27–30 (2000).
473 10. Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena
474 platform. (2020).
475 11. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server
476 for large-scale expression profiling and interactive analysis.
477 12. Bal, E. *et al.* Super-enhancer hypermutation alters oncogene expression in B cell
478 lymphoma. *Nature* **607**, 808–815 (2022).
479 13. Zhou, H. *et al.* Epstein-Barr Virus Oncoprotein Super-enhancers Control B Cell
480 Growth. *Cell Host & Microbe* **17**, 205–216 (2015).
481 14. Mills, K. H. G. IL-17 and IL-17-producing cells in protection versus pathology. *Nat*
482 *Rev Immunol* **23**, 38–54 (2023).
483 15. Luchtman, D. W., Ellwardt, E., Larochelle, C. & Zipp, F. IL-17 and related cytokines
484 involved in the pathology and immunotherapy of multiple sclerosis: Current and future

485 developments. *Cytokine Growth Factor Rev* **25**, 403–413 (2014).

486 16. K, L. Signaling Cross Talk between TGF-β/Smad and Other Signaling Pathways.
487 *Cold Spring Harbor perspectives in biology* **9**, (2017).

488 17. Menter, T. *et al.* Mutational landscape of B-cell post-transplant lymphoproliferative
489 disorders. *Br J Haematol* **178**, 48–56 (2017).

490 18. Bruce, J. P. *et al.* Whole-genome profiling of nasopharyngeal carcinoma reveals
491 viral-host co-operation in inflammatory NF-κB activation and immune escape. *Nat*
492 *Commun* **12**, 4193 (2021).

493 19. Frontzek, F. *et al.* Molecular profiling of EBV associated diffuse large B-cell
494 lymphoma. *Leukemia* **37**, 670–679 (2023).

495 20. Panea, R. I. *et al.* The whole-genome landscape of Burkitt lymphoma subtypes.
496 *Blood* **134**, 1598–1607 (2019).

497 21. Garcia-Reyero, J. *et al.* Genetic lesions in MYC and STAT3 drive oncogenic
498 transcription factor overexpression in plasmablastic lymphoma. *Haematologica* **106**,
499 1120–1128 (2021).

500 22. Wang, F. *et al.* Epstein-Barr virus nuclear antigen 2 specifically induces expression
501 of the B-cell activation antigen CD23. *Proc Natl Acad Sci U S A* **84**, 3452–3456 (1987).

502 23. Uchida, J. *et al.* Mimicry of CD40 signals by Epstein-Barr virus LMP1 in B
503 lymphocyte responses. *Science* **286**, 300–303 (1999).

504 24. Oudejans, J. J. *et al.* BHRF1, the Epstein-Barr virus (EBV) homologue of the BCL-2
505 protooncogene, is transcribed in EBV-associated B-cell lymphomas and in reactive
506 lymphocytes. *Blood* **86**, 1893–1902 (1995).

507 25. Kenney, J. L., Guinness, M. E., Curiel, T. & Lacy, J. Antisense to the epstein-barr
508 virus (EBV)-encoded latent membrane protein 1 (LMP-1) suppresses LMP-1 and bcl-2
509 expression and promotes apoptosis in EBV-immortalized B cells. *Blood* **92**, 1721–1727
510 (1998).
511 26. Ribeiro, J. *et al.* P53 deregulation in Epstein-Barr virus-associated gastric cancer.
512 *Cancer Lett* **404**, 37–43 (2017).
513 27. Sarkozy, C. *et al.* Mutational landscape of gray zone lymphoma. *Blood* **137**, 1765–
514 1776 (2021).
515 28. Lee, J. E. *et al.* Clinicopathologic and genomic characteristics of mucinous gastric
516 adenocarcinoma. *Gastric Cancer* **25**, 697–711 (2022).
517 29. Lee, H. S., Chang, M. S., Yang, H.-K., Lee, B. L. & Kim, W. H. Epstein-barr virus-
518 positive gastric carcinoma has a distinct protein expression profile in comparison with
519 epstein-barr virus-negative carcinoma. *Clin Cancer Res* **10**, 1698–1705 (2004).
520 30. Vereide, D. T. *et al.* Epstein-Barr virus maintains lymphomas via its miRNAs.
521 *Oncogene* **33**, 1258–1264 (2014).
522 31. Mansouri, S., Pan, Q., Blencowe, B. J., Claycomb, J. M. & Frappier, L. Epstein-Barr
523 virus EBNA1 protein regulates viral latency through effects on let-7 microRNA and dicer.
524 *J Virol* **88**, 11166–11177 (2014).
525 32. Zhang, J.-Y. *et al.* ebv-circRPMS1 promotes the progression of EBV-associated
526 gastric carcinoma via Sam68-dependent activation of METTL3. *Cancer Lett* **535**, 215646
527 (2022).
528 33. Zhang, K. *et al.* Caspases Switch off the m ^{6} A RNA Modification Pathway to Foster

529 the Replication of a Ubiquitous Human Tumor Virus. *mBio* **12**, e01706-21 (2021).

530 34. Bell, A. & Rickinson, A. B. Epstein-Barr virus, the TCL-1 oncogene and Burkitt's

531 lymphoma. *Trends Microbiol* **11**, 495–497 (2003).

532 35. Dhanasekaran, R. *et al.* The MYC oncogene - the grand orchestrator of cancer

533 growth and immune evasion. *Nat Rev Clin Oncol* **19**, 23–36 (2022).

534 36. Guo, R. *et al.* MYC Controls the Epstein-Barr Virus Lytic Switch. *Mol Cell* **78**, 653-

535 669.e8 (2020).

536 37. Jiang, S. *et al.* The Epstein-Barr Virus Regulome in Lymphoblastoid Cells. *Cell Host*

537 *Microbe* **22**, 561-573.e4 (2017).

538 38. Arvey, A. *et al.* An Atlas of the Epstein-Barr Virus Transcriptome and Epigenome

539 Reveals Host-Virus Regulatory Interactions. *Cell Host & Microbe* **12**, 233–245 (2012).

540 39. Park, G. B. *et al.* CD80 (B7.1) and CD86 (B7.2) induce EBV-transformed B cell

541 apoptosis through the Fas/FasL pathway. *Int J Oncol* **43**, 1531–1540 (2013).

542 40. Kis, L. L., Takahara, M., Nagy, N., Klein, G. & Klein, E. IL-10 can induce the

543 expression of EBV-encoded latent membrane protein-1 (LMP-1) in the absence of EBNA-

544 2 in B lymphocytes and in Burkitt lymphoma- and NK lymphoma-derived cell lines. *Blood*

545 **107**, 2928–2935 (2006).

546 41. Weber-Nordt, R. M. *et al.* Constitutive activation of STAT proteins in primary

547 lymphoid and myeloid leukemia cells and in Epstein-Barr virus (EBV)-related lymphoma

548 cell lines. *Blood* **88**, 809–816 (1996).

549 42. Hsiao, J.-R. *et al.* Constitutive activation of STAT3 and STAT5 is present in the

550 majority of nasopharyngeal carcinoma and correlates with better prognosis. *Br J Cancer*

551 **89**, 344–349 (2003).

552 43. Wang, L. *et al.* LMP1 signaling pathway activates IRF4 in latent EBV infection and a
553 positive circuit between PI3K and Src is required. *Oncogene* **36**, 2265–2274 (2017).

554 44. Cohen, J. I. Herpesviruses in the Activated Phosphatidylinositol-3-Kinase-δ
555 Syndrome. *Front Immunol* **9**, 237 (2018).

556 45. Lu, J., Chua, H.-H., Chen, S.-Y., Chen, J.-Y. & Tsai, C.-H. Regulation of matrix
557 metalloproteinase-1 by Epstein-Barr virus proteins. *Cancer Res* **63**, 256–262 (2003).

558 46. Verma, D., Church, T. M. & Swaminathan, S. Epstein-Barr Virus Lytic Replication
559 Induces ACE2 Expression and Enhances SARS-CoV-2 Pseudotyped Virus Entry in
560 Epithelial Cells. *J Virol* **95**, e0019221 (2021).

561 47. Fleischmann, J., Kremmer, E., Greenspan, J. S., Grässer, F. A. & Niedobitek, G.
562 Expression of viral and human dUTPase in Epstein-Barr virus-associated diseases.
563 *Journal of Medical Virology* **68**, 568–573 (2002).

564 48. Williams, M. & Ariza, M. E. EBV Positive Diffuse Large B Cell Lymphoma and
565 Chronic Lymphocytic Leukemia Patients Exhibit Increased Anti-dUTPase Antibodies.
566 *Cancers (Basel)* **10**, 129 (2018).

567 49. Mosialos, G. *et al.* A Ca2+/calmodulin-dependent protein kinase, CaM kinase-Gr,
568 expressed after transformation of primary human B lymphocytes by Epstein-Barr virus
569 (EBV) is induced by the EBV oncogene LMP1. *J Virol* **68**, 1697–1705 (1994).

570 50. Hulse, M. *et al.* Poly(ADP-ribose) polymerase 1 is necessary for coactivating
571 hypoxia-inducible factor-1-dependent gene expression by Epstein-Barr virus latent
572 membrane protein 1. *PLoS Pathog* **14**, e1007394 (2018).

51. Jiang, Y. *et al.* Repression of Hox genes by LMP1 in nasopharyngeal carcinoma and modulation of glycolytic pathway genes by HoxC8. *Oncogene* **34**, 6079–6091 (2015).

Acknowledgments: We thank all members of our laboratory for helpful advice and assistance.

Funding: This work was supported by the Science, Technology and Innovation Commission of Shenzhen Municipality (grant no. JCYJ20230807093208017 to Y.P.), the Guangdong Natural Science Foundation (grant no. 2024A1515013126 to Y.P.).

Author contributions: Y.P. conceived the application of LLM in the EBV-related literature analysis. J.Y. designed and carried out the data mining and visualization. Y.W., H.W., and Z.W. provided expertise in technical support and data analysis. J.Y. wrote the manuscript and Y.P. revised the manuscript.

Competing interests: The authors declare no potential conflicts of interest.

Figure legends

Figure 1. LLMs-assisted data mining reveals the key insights in EBV research. (**a**) The trend of EBV-related publications in different years. (**b**) The pipeline in this study is illustrated as a flowgraph. (**c**) Evaluation of GPT-generated output is conducted by examining 102 entries recorded for EBNA3C and LMP2. (**d**) The relative frequency

of EBV antigens (log2 transformed) in our analysis is illustrated in the word cloud. (**e**) Statistical analysis highlights the research focus on various EBV antigens over time, showcasing only the four most studied antigens for each decade. (**f**) The relative frequency of EBV-related pathways was shown in the word cloud. (**g**) Statistical analysis highlights the research focus on various EBV-related pathways over time, showcasing only the four most pathways for each decade.

Figure 2. Pathway mapping reveals the distinct signatures of EBV antigens. (**a**) The chord diagram shows the linkage between the top 33 EBV antigens and their corresponding pathways. (**b**) Heatmap view of the linkage between EBV latent or lytic antigens and the associated pathways.

Figure 3. Gene enrichment and pathway analysis uncover defining characteristics of EBV-induced cellular modulations. (a-b) KEGG enrichment shows the top 10 EBV-related diseases or pathogens (a) and signaling pathways (b). (c) Reactome enrichment maps the landscape of EBV-targeted pathways and their correlations based on the shared host proteins.

Figure 4. Analysis of host protein-protein interactions identifies key mediators in EBV-mediated diseases. (a) Statistical analysis reveals the focus of study every decade on EBV-targeted host proteins. TF, transcription factor; GPCR, G Protein-Coupled Receptor. (b-d) Top 3 PPI networks describe the basic functional clusters of

EBV-targeted host proteins.

Figure 5. Revealing the key relationships in EBV-mediated DLBCLs. (a) The bar plot shows the numbers of total articles of the top 12 most studied EBV-related diseases. (b) The bar plot shows the numbers of total articles of the top 7 most studied EBV antigens. (c) The circular hub graph elucidates the DLBCL-related antigen-pathway relations. (d) PCA analysis highlights distinct DLBCL tumor samples using TLR2-related marker gene expressions. (e) The TLR2-related marker gene expressions between normal tissue and tumor tissue. (f) Survival plot shows high TNF expression leads to less survival probability (red indicates samples with normalized TNF counts > 7.907 and blue indicates less).

Figure 6. Exploring the functions of EBV antigens in BL. (a) The circular hub graph elucidates the BL-related antigen-pathway relations. (b-c) The CAMK4-related factors are co-expressed by LMP1 overexpression. (d-e) The PARP1-related factors are co-expressed by LMP1 overexpression.

Supplementary materials

Figure S1: Determining the key relationships in EBV-mediated NPC, PTLD, and GC. (a-c) The circular hub graph elucidates the NPC-related, PTLD-related and GC-related antigen-pathway relations. NPC, nasopharyngeal carcinoma; PTLD, Posttransplant lymphoproliferative disorders; GC, gastric cancer.;

Table S1: The formal names and aliases used in GPT-generated term substitution;

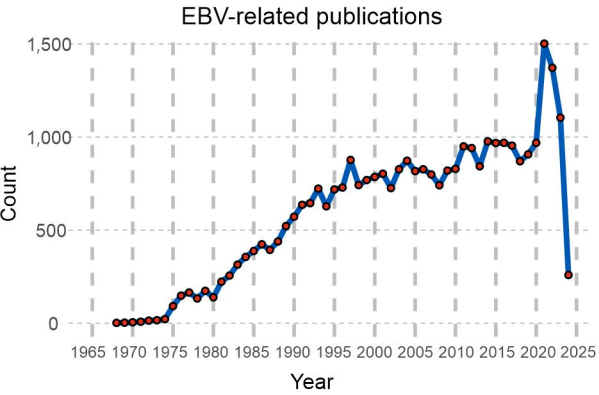
Table S2: The correspondence between GPT-4 generated pathways and manually annotated pathway to KEGG entries;

Table S3: The final dataset includes 2,289 entries of EBV antigens and their associated pathways;

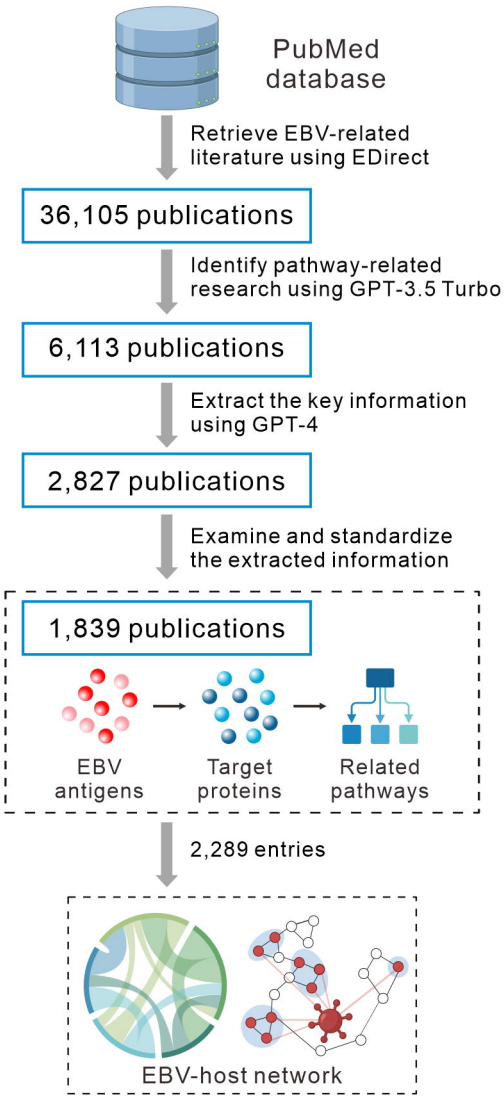
Table S4: Validation for GPT-generated results;

Table S5: Ranking of most studied EBV antigens (with >10 related articles);

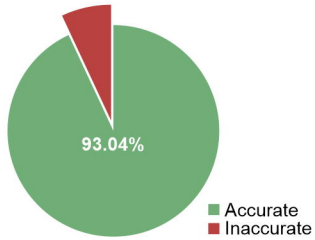
a


EBV-related publications
1,500
1,000
500
0
Count
1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025
Year

b


PubMed database
Retrieve EBV-related literature using EDirect
36,105 publications
Identify pathway-related research using GPT-3.5 Turbo
6,113 publications
Extract the key information using GPT-4
2,827 publications
Examine and standardize the extracted information
1,839 publications
EBV antigens
Target proteins
Related pathways
2,289 entries
EBV-host network

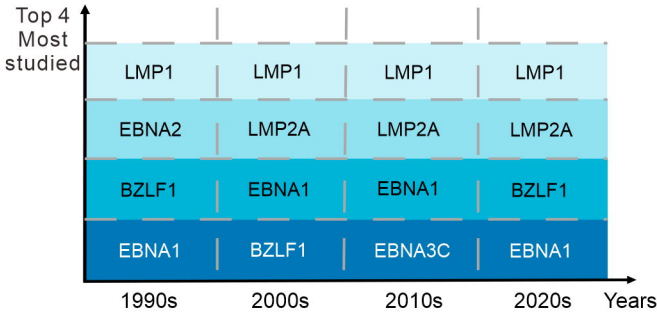
c


GPT-4 prompting accuracy
93.04%
Accurate
Inaccurate

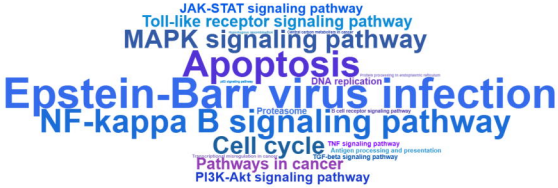
d

BGLF5
BALF5
gp350
BLLF3
BILF1
BZLF2
EBER1
LMP2A
BMRF1
BARF1
EBNA1
BGLF4
EBNA3A
EBNALP
BHRF1
gM
gL
LMP1
EBNA3C
gH
BRLF1
BZLF1
EBER
LMP2
LMP
BGLF2
EBNA2
BCRF1
BPLF1
EBER2
BMLF1
BFRF1
BNLF2a

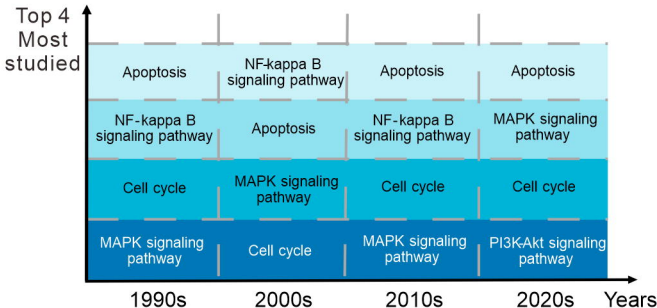
e

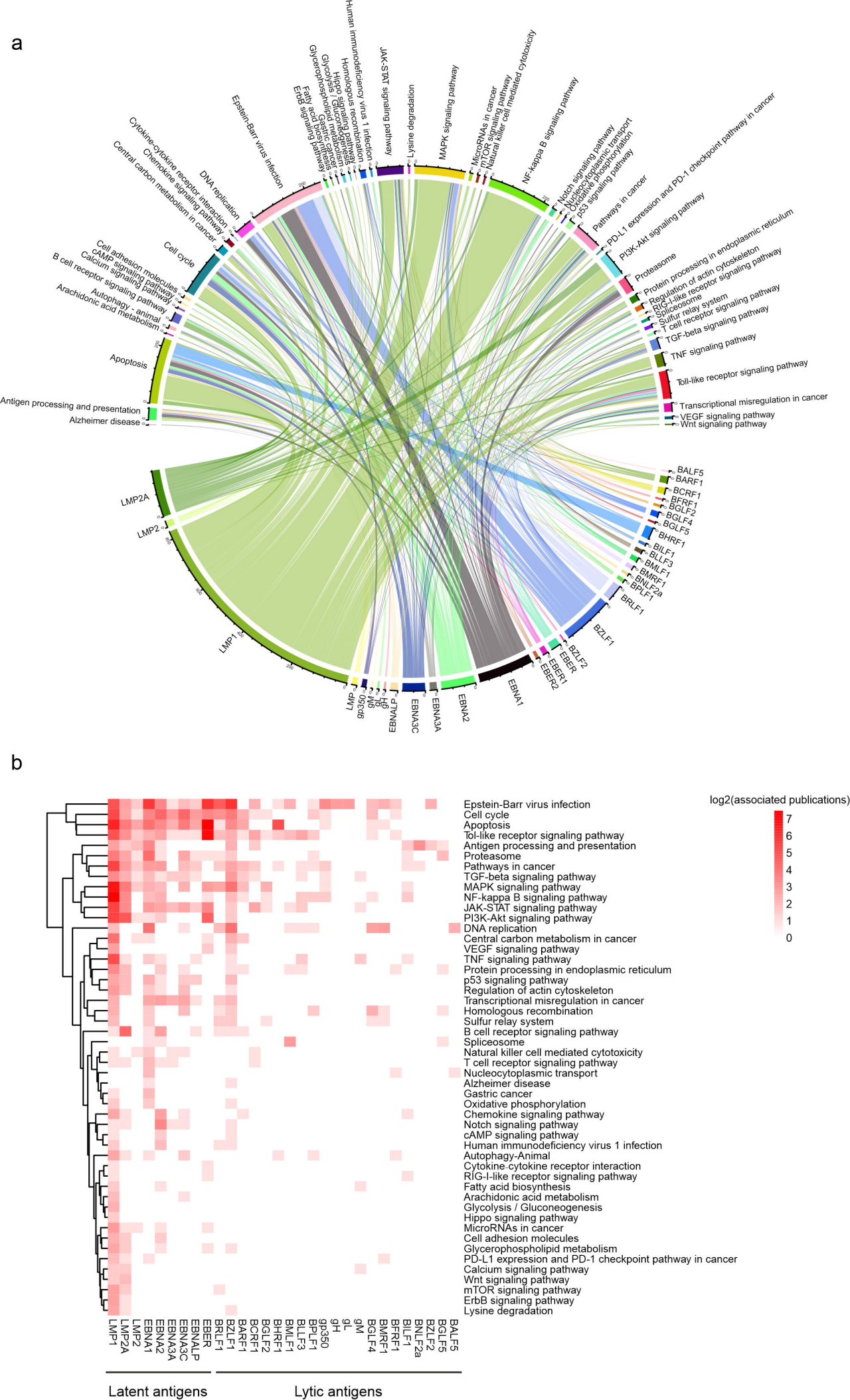

Top 4 Most studied
LMP1 LMP1 LMP1 LMP1
EBNA2 LMP2A LMP2A LMP2A
BZLF1 EBNA1 EBNA1 BZLF1
EBNA1 BZLF1 EBNA3C EBNA1
1990s 2000s 2010s 2020s Years

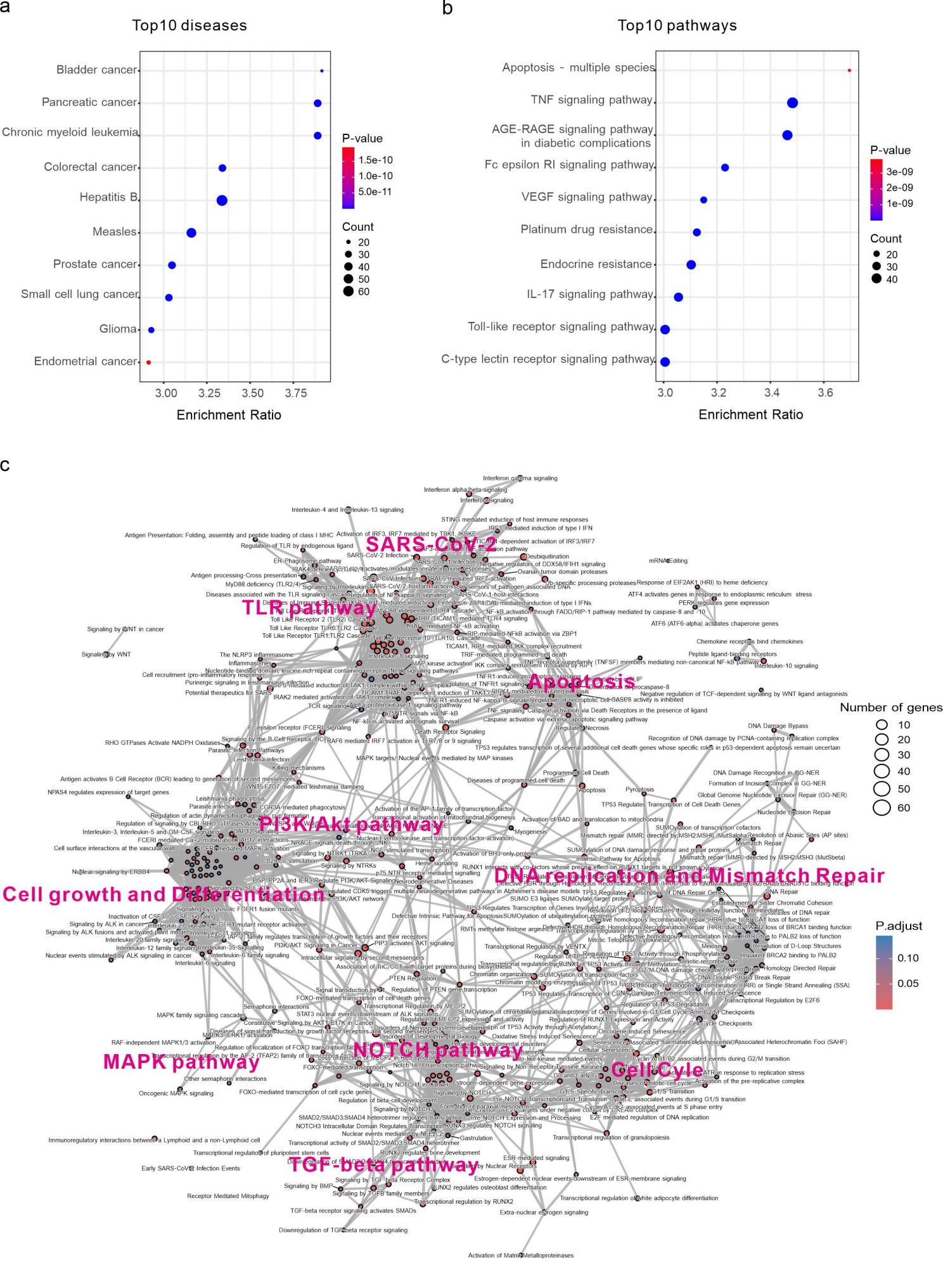
f

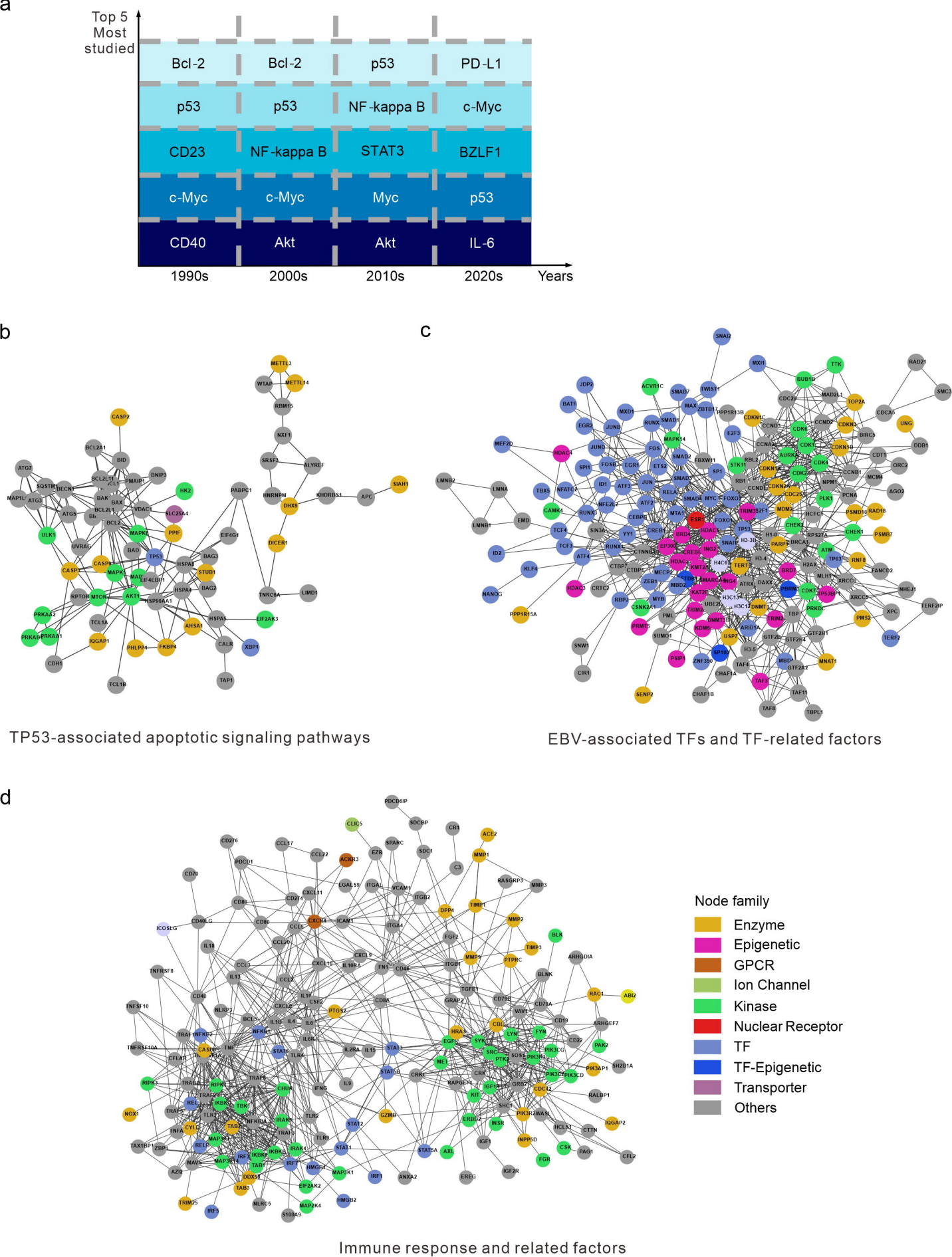

JAK-STAT signaling pathway
Toll-like receptor signaling pathway
MAPK signaling pathway
Apoptosis
DNA replication
Epstein-Barr virus infection
Proteasome
NF-kappa B signaling pathway
Cell cycle
TNF signaling pathway
Pathways in cancer
PI3K-Akt signaling pathway

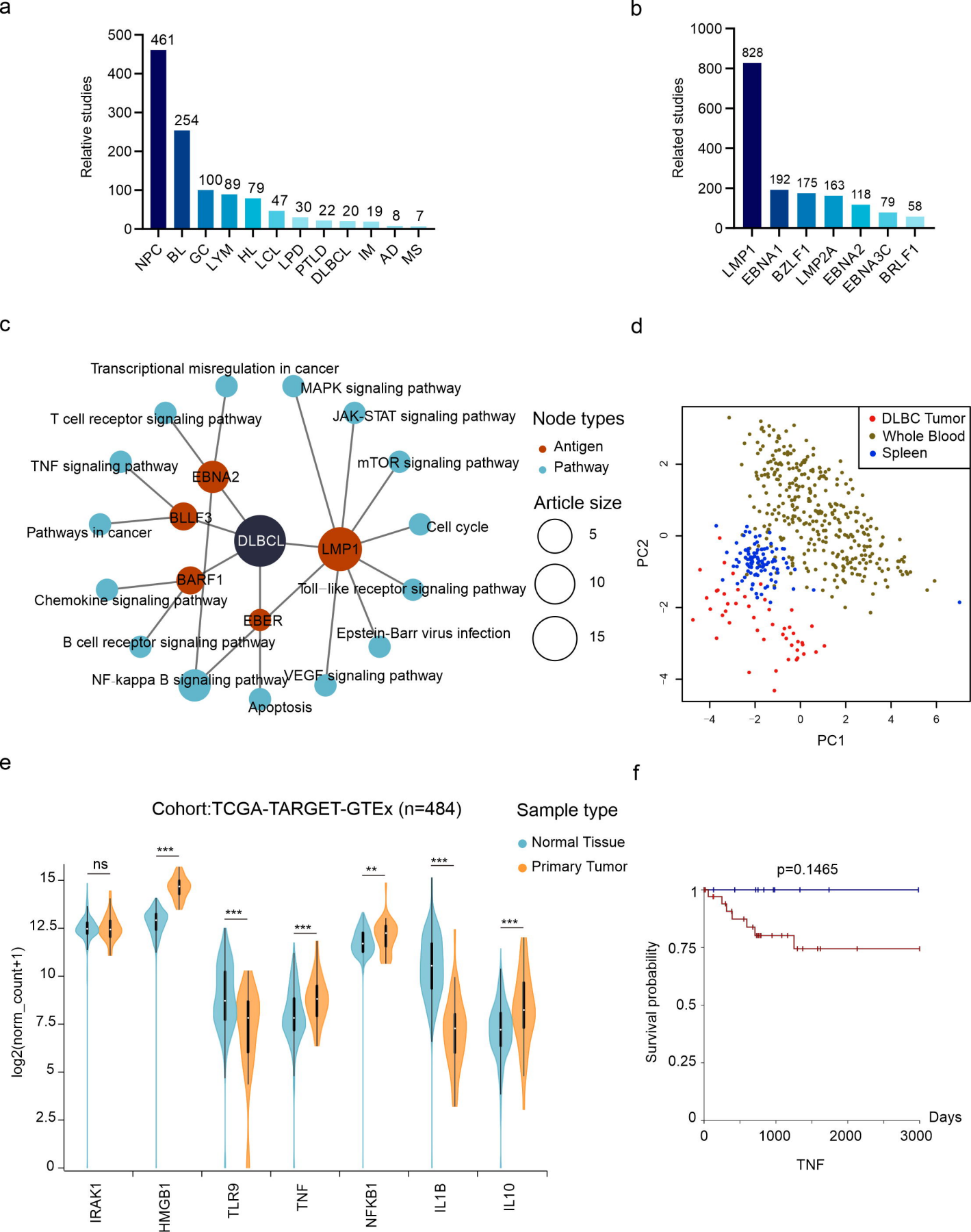
g

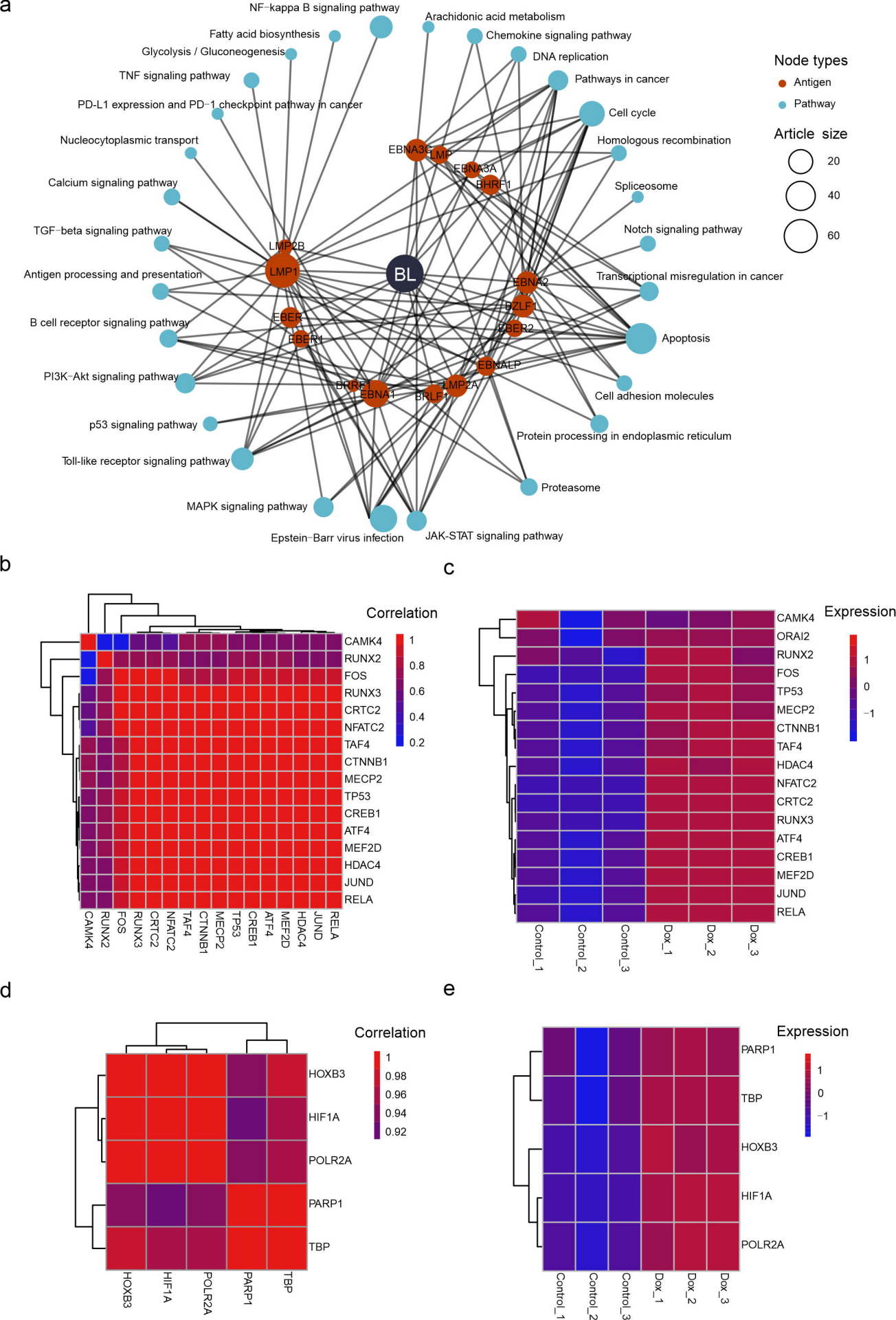

Top 4 Most studied
Apoptosis NF-kappa B signaling pathway Apoptosis Apoptosis
NF-kappa B signaling pathway Apoptosis NF-kappa B signaling pathway MAPK signaling pathway
Cell cycle MAPK signaling pathway Cell cycle Cell cycle
MAPK signaling pathway Cell cycle MAPK signaling pathway PI3K-Akt signaling pathway
1990s 2000s 2010s 2020s Years


a
Epstein-Barr virus infection
DNA replication
Cytokine-cytokine receptor interaction
Chemokine signaling pathway
Central carbon metabolism in cancer
Cell cycle
Cell adhesion molecules
cAMP signaling pathway
Calcium signaling pathway
B cell receptor signaling pathway
Autophagy - animal
Arachidonic acid metabolism
Apoptosis
Antigen processing and presentation
Alzheimer disease
Human immunodeficiency virus 1 infection
Homologous recombination
Hippo signaling pathway
Glycolysis / Gluconeogenesis
Gastric cancer
Glycerophospholipid metabolism
Fatty acid biosynthesis
ErbB signaling pathway
JAK-STAT signaling pathway
Lysine degradation
MAPK signaling pathway
MicroRNAs in cancer
mTOR signaling pathway
Natural killer cell mediated cytotoxicity
NF-kappa B signaling pathway
Notch signaling pathway
Nucleocytoplasmic transport
Oxidative phosphorylation
p53 signaling pathway
Pathways in cancer
PD-L1 expression and PD-1 checkpoint pathway in cancer
PI3K-Akt signaling pathway
Proteasome
Protein processing in endoplasmic reticulum
Regulation of actin cytoskeleton
RIG-I-like receptor signaling pathway
Spliceosome
Sulfur relay system
T cell receptor signaling pathway
TGF-beta signaling pathway
TNF signaling pathway
Toll-like receptor signaling pathway
Transcriptional misregulation in cancer
VEGF signaling pathway
Wnt signaling pathway
LMP2A
LMP2
LMP1
LMP
gp350
gp
gH
gL
EBNALP
EBNA3C
EBNA3A
EBNA2
EBNA1
EBER2
EBER1
EBER
BZLF2
BZLF1
BRLF1
BPLF1
BNLF2a
BMRF1
BMLF1
BLLF3
BLLF1
BHRF1
BGLF5
BGLF4
BGLF2
BFRF1
BCRF1
BARF1
BALF5
b
log2(associated publications)
7
6
5
4
3
2
1
0
Epstein-Barr virus infection
Cell cycle
Apoptosis
Toll-like receptor signaling pathway
Antigen processing and presentation
Proteasome
Pathways in cancer
TGF-beta signaling pathway
MAPK signaling pathway
NF-kappa B signaling pathway
JAK-STAT signaling pathway
PI3K-Akt signaling pathway
DNA replication
Central carbon metabolism in cancer
VEGF signaling pathway
TNF signaling pathway
Protein processing in endoplasmic reticulum
p53 signaling pathway
Regulation of actin cytoskeleton
Transcriptional misregulation in cancer
Homologous recombination
Sulfur relay system
B cell receptor signaling pathway
Spliceosome
Natural killer cell mediated cytotoxicity
T cell receptor signaling pathway
Nucleocytoplasmic transport
Alzheimer disease
Gastric cancer
Oxidative phosphorylation
Chemokine signaling pathway
Notch signaling pathway
cAMP signaling pathway
Human immunodeficiency virus 1 infection
Autophagy-Animal
Cytokine-cytokine receptor interaction
RIG-I-like receptor signaling pathway
Fatty acid biosynthesis
Arachidonic acid metabolism
Glycolysis / Gluconeogenesis
Hippo signaling pathway
MicroRNAs in cancer
Cell adhesion molecules
Glycerophospholipid metabolism
PD-L1 expression and PD-1 checkpoint pathway in cancer
Calcium signaling pathway
Wnt signaling pathway
mTOR signaling pathway
ErbB signaling pathway
Lysine degradation
LMP1
LMP2A
LMP2
EBNA1
EBNA2
EBNA3A
EBNA3C
EBNALP
EBER
BRLF1
BZLF1
BARF1
BCRF1
BGLF2
BHRF1
BMLF1
BLLF3
BPLF1
gp350
gH
gL
gM
BGLF4
BMRF1
BFRF1
BLF1
BNLF2a
BZLF2
BGLF5
BALF5
Latent antigens
Lytic antigens


a
Top10 diseases
Bladder cancer
Pancreatic cancer
Chronic myeloid leukemia
Colorectal cancer
Hepatitis B
Measles
Prostate cancer
Small cell lung cancer
Glioma
Endometrial cancer
3.00
3.25
3.50
3.75
Enrichment Ratio
P-value
1.5e-10
1.0e-10
5.0e-11
Count
20
30
40
50
60
b
Top10 pathways
Apoptosis - multiple species
TNF signaling pathway
AGE-RAGE signaling pathway in diabetic complications
Fc epsilon RI signaling pathway
VEGF signaling pathway
Platinum drug resistance
Endocrine resistance
IL-17 signaling pathway
Toll-like receptor signaling pathway
C-type lectin receptor signaling pathway
3.0
3.2
3.4
3.6
Enrichment Ratio
P-value
3e-09
2e-09
1e-09
Count
20
30
40
c
SARS-CoV-2
TLR pathway
Apoptosis
PI3K/Akt pathway
Cell growth and Differentiation
DNA replication and Mismatch Repair
MAPK pathway
NOTCH pathway
Cell Cycle
TGF-beta pathway
Number of genes
10
20
30
40
50
60
P.adjust
0.10
0.05


a
Top 5 Most studied
Bcl-2
Bcl-2
p53
PD-L1
p53
p53
NF-kappa B
c-Myc
CD23
NF-kappa B
STAT3
BZLF1
c-Myc
c-Myc
Myc
p53
CD40
Akt
Akt
IL-6
1990s
2000s
2010s
2020s
Years
b
TP53-associated apoptotic signaling pathways
c
EBV-associated TFs and TF-related factors
d
Node family
Enzyme
Epigenetic
GPCR
Ion Channel
Kinase
Nuclear Receptor
TF
TF-Epigenetic
Transporter
Others
Immune response and related factors


a
Relative studies
461
254
100
89
79
47
30
22
20
19
8
7
NPC
BL
GC
LYM
HL
LCL
LPD
PTLD
DLBCL
IM
AD
MS
b
Related studies
828
192
175
163
118
79
58
LMP1
EBNA1
BZLF1
LMP2A
EBNA2
EBNA3C
BRLF1
c
Transcriptional misregulation in cancer
MAPK signaling pathway
T cell receptor signaling pathway
JAK-STAT signaling pathway
TNF signaling pathway
mTOR signaling pathway
EBNA2
BLLF3
Pathways in cancer
DLBCL
LMP1
Cell cycle
BARF1
Chemokine signaling pathway
Toll-like receptor signaling pathway
EBER
B cell receptor signaling pathway
Epstein-Barr virus infection
NF-kappa B signaling pathway
VEGF signaling pathway
Apoptosis
Node types
Antigen
Pathway
Article size
5
10
15
d
DLBC Tumor
Whole Blood
Spleen
PC2
PC1
e
Cohort:TCGA-TARGET-GTEx (n=484)
Sample type
Normal Tissue
Primary Tumor
log2(norm_count+1)
ns
***
**
IRAK1
HMGB1
TLR9
TNF
NFKB1
IL1B
IL10
f
p=0.1465
Survival probability
Days
TNF


a
NF-kappa B signaling pathway
Arachidonic acid metabolism
Fatty acid biosynthesis
Chemokine signaling pathway
Glycolysis / Gluconeogenesis
DNA replication
TNF signaling pathway
Pathways in cancer
PD-L1 expression and PD-1 checkpoint pathway in cancer
Cell cycle
Nucleocytoplasmic transport
Homologous recombination
EBNA3C
LMP
EBNA3A
BHRF1
Calcium signaling pathway
Spliceosome
TGF-beta signaling pathway
Notch signaling pathway
LMP2B
LMP1
BL
Antigen processing and presentation
Transcriptional misregulation in cancer
EBNA2
BZLF1
B cell receptor signaling pathway
EBER
EBER1
EBER2
Apoptosis
EBNALP
PI3K-Akt signaling pathway
BRRF1
EBNA1
BRLF1
LMP2A
Cell adhesion molecules
p53 signaling pathway
Protein processing in endoplasmic reticulum
Toll-like receptor signaling pathway
Proteasome
MAPK signaling pathway
Epstein-Barr virus infection
JAK-STAT signaling pathway
Node types
Antigen
Pathway
Article size
20
40
60
b
Correlation
1
0.8
0.6
0.4
0.2
CAMK4
RUNX2
FOS
RUNX3
CRTC2
NFATC2
TAF4
CTNNB1
MECP2
TP53
CREB1
ATF4
MEF2D
HDAC4
JUND
RELA
c
Expression
1
0
−1
CAMK4
ORAI2
RUNX2
FOS
TP53
MECP2
CTNNB1
TAF4
HDAC4
NFATC2
CRTC2
RUNX3
ATF4
CREB1
MEF2D
JUND
RELA
Control_1
Control_2
Control_3
Dox_1
Dox_2
Dox_3
d
Correlation
1
0.98
0.96
0.94
0.92
HOXB3
HIF1A
POLR2A
PARP1
TBP
e
Expression
1
0
−1
PARP1
TBP
HOXB3
HIF1A
POLR2A
Control_1
Control_2
Control_3
Dox_1
Dox_2
Dox_3