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 large language models (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 diagrams. Furthermore, we described the critical roles of EBV antigens by constructing functional subsets of host proteins targeted by EBV antigens and illustrating these interactions through protein-protein interaction (PPI) networks. Finally, our model delineates the core attributes of host proteins targeted by LMP1, providing unique insights into potential therapeutic strategies against EBV-associated diseases.

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 infectious mononucleosis (IM), nasopharyngeal carcinoma (NPC), Burkitt lymphoma (BL), Hodgkin lymphoma (HL), 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 current landscape of studies on EBV antigens and their interactions with cellular signaling pathways.

Traditional literature mapping techniques, including systematic reviews and meta-analyses, often fall short of capturing the full breadth of literature in a biomedical field. 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. Collectively, this study offers a unique overview of current EBV-host interactions and provides valuable insights for the development of anti-EBV therapeutic strategies in the future.

Results

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

To create an appropriate dataset for information extraction, our initial dataset comprised 36,105 EBV-related research articles available in the PubMed database as of the start date 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). Furthermore, to ensure the precision of our research and to independently study the functionality of each factor, we deliberately excluded all studies that investigated three or more EBV antigens. The final dataset includes 2,289 documented entries of EBV antigens and their associated pathways, derived from 1,839 publications (Figure 1b). To validate the accuracy of the information extracted using ChatGPT, we manually compared the GPT-4 generated summaries with the corresponding original abstracts. The dataset selected for validation comprised 102 entries related to EBNA3C and LMP2. Our results indicate that our prompts achieved an average accuracy of 96.08% (Figure 1c, Table S1), 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). Notably, the functions of LMP1 and LMP2A have consistently remained a hot research focus, with BZLF1 and EBNA1 also receiving significant attention. 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). In addition, within the top 20 most scrutinized pathways, the B-cell-specific BCR (B-cell receptor) signaling pathway and the antigen presentation pathways are prominently featured, which are integral to the immune response. Other research focuses include protein modification or DNA homologous recombination related to EBV infection. 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 (Figure S1). 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. 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 (Figure 2b). 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[9,10]. 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 S2). 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 and pathway analysis 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. Hepatitis B infection implies similarities between host response induced by Hepatitis B Virus (HBV) and EBV infections. Interestingly, HBV and EBV are both oncogenic DNA viruses capable of directly interacting with the host genome[11,12]. This finding suggests the potential for examining the common features of HBV and EBV infections to gain complementary insights. The results also indicate a possible link between immune compromise induced by Chronic Myeloid Leukemia and EBV infection, opening up intriguing areas for future exploration. 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). 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[13]. 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.

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 ten 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[14].

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[15–17] and MYC[18,19] 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[20] and verified the role of LMP1 in viral latency maintenance via mimicry of the CD40 signaling pathways[21]. 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[17,22,23].

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[24–27]. 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[28]. 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[29]. Recent studies have also uncovered the role of circular RNAs in promoting the progression of EBV-associated gastric carcinoma through the transactivation of METTL3[30]. Along this pathway, caspases have been found to inhibit the m6A RNA modification process, thereby promoting EBV replication[31]. Additionally, EBV infection is known to induce TCL1 gene expression in Burkitt's lymphoma, which also serves as an activator for the Akt pathway[32].

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[33]. Moreover, MYC abundance represses EBV reactivation and controls the lytic switch in Burkitt lymphoma[34]. 3D genome mapping of EBV-positive lymphoblastoid cell lines (LCLs) revealed distinctive EBV super-enhancer (SE) targets like MCL1, IRF4, and EBF[35]. Bioinformatic analysis identified that H3K27ac

SEs overlap with STAT5 and NFAT targets, indicating their roles in promoting EBV latency[10]. 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[36]. These EBV-specific TFs, such as SPI1, EBF1, and PAX5, collectively contribute to EBV-induced oncogenesis and latency[10].

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[37,38]. For example, LMP1 mimics CD40 to activate downstream NF-κB pathways to induce B-cell proliferation[21]. After infection, EBV hijacks host signaling transduction factors like STATs[39,40] and IRFs[41] 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[42]. 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[43]. 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[44].

Enrichment networks reveal the core attributes of LMP1

To validate and further refine the characterization of EBV antigen-cellular pathway landscape, we employed our enrichment pipeline to investigate the most studied EBV antigen, LMP1 (Figure 5a, Table S2). LMP1 is a multi-functional membrane protein enriched in the lipid rafts of EBV-infected lymphocytes, which is critical for EBV oncogenesis and B cell transformation. LMP1 notably enhances cell proliferation and growth, activates numerous cellular signaling pathways, and facilitates immune evasion. The enrichment of our dataset for LMP1 underscores its profound impact, with related pathways representing nearly fifty percent of the critical components mapped across all EBV antigens, particularly in areas related to cell growth, cell cycle, and immune response (Figure S2a). A more refined analysis focuses on the top 30 most interconnected pathways related to LMP1 (Figure 5b), categorizing its functions into four main areas: Toll-like receptor signaling, FGR signaling, cell cycle control via Retinoblastoma Protein (RB1), and cell apoptosis through caspase activation. Notably, the FGR domain is split into two clusters that distinctly address FGFR regulation in normal and disease states. Broadening our perspective, we examined the top 100 pathways (Figure S2b). The cell growth and cell cycle clusters converge through interactions mediated by PTK6 and signaling by non-receptor tyrosine kinases. Meanwhile, Toll-like receptor pathways merge with the

cell growth cluster due to pathogen-induced inflammatory responses. These factors bridging LMP1's core functional clusters could serve as potential drug candidates to disrupt its activity.

Next, we questioned whether the pathways identified from our literature review reflect biological processes at transcriptomic level. RNA-seq analysis indicates that LMP1 overexpression induces differentially expressed genes (DEGs) that predominantly constitute the LMP1-targeted pathway network (Figure 5c). The DEGs based on literature show close interactions, demonstrating the efficacy of our analyses (Figure 5d). To further explore the core protein factors of these host cellular pathways, we employed the PPI analysis on the obtained LMP1 target proteins. The top three enriched clusters encompass the apoptosis, immune response, and cell cycle-related host factors (Figure 5e). The apoptotic cluster centers around TP53, with BCL2 and PML showing the most differential expression. The second cluster consists of LMP1-induced host immune response, including the interferon regulatory factors (IRFs), TNF receptor associated factors (TRAFs), and NF-κB associated factors. The last cluster includes various signaling transduction pathways contributing to LMP1-induced B cell transformation marked by CD40 and STATs activation. These PPI clusters encapsulate the current understanding of LMP1-targeted host proteins. By strategically targeting the key factors within these networks, we could potentially cripple the oncogenic pathways driven by LMP1, paving the way for novel therapeutic interventions.

Discussion

This study, for the first time, presents a mapping between EBV antigens, host factors, and cellular pathways through screening EBV-related articles using LLMs. Beyond direct language processing and transformation through natural language processing algorithms, this study is based on experimentally validated findings, which adds to its credibility.

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. As LLM technology advances, generating summaries from complete articles could become more cost-effective in terms of money and time. Future analyses employing more flexible models are likely to provide a more accurate interpretation of EBV-induced cellular changes. Lastly, 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. 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.

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, 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.

Methods

Data sources

The data in this study were obtained from the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) using EDirect[45]. 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 prompting for publication acquisition and data processing

To ensure accuracy, we fine-tuned ChatGPT by providing it with the EBV antigen list obtained from NCBI before submitting prompts. An example prompt 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. Keep your reply as consistent as possible. All EBV proteins were listed before. 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 a combination of manual correction and synonym substitution for invalid GPT-4 generated outputs using custom R scripts for further data cleaning. Specifically, we standardized 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”.

GPT-4 generated data evaluation

To evaluate the accuracy of the GPT-generated results, we selected a total of 102 entries recorder for EBNA3C and LMP2 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, 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 (Figure S1).

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[46], Reactome (https://reactome.org/), and STRINGdb.

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).

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-kappaB pathway in carcinoma cells by inhibiting IKK phosphorylation. *Mol Cancer* **9**, 1 (2010).

7. González-Márquez, R., Schmidt, L., Schmidt, B. M., Berens, P. & Kobak, D. The landscape of biomedical research. *Patterns* 100968 (2024)

doi:10.1016/j.patter.2024.100968.

8. Wu, X. *et al.* reguloGPT: Harnessing GPT for Knowledge Graph Construction of Molecular Regulatory Pathways. Preprint at https://doi.org/10.1101/2024.01.27.577521 (2024).

9. Bal, E. *et al.* Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. *Nature* **607**, 808–815 (2022).

10. Zhou, H. *et al.* Epstein-Barr Virus Oncoprotein Super-enhancers Control B Cell Growth. *Cell Host & Microbe* **17**, 205–216 (2015).

11. Qian, Z. *et al.* HBV integrations reshaping genomic structures promote hepatocellular carcinoma. *Gut* gutjnl-2023-330414 (2024) doi:10.1136/gutjnl-2023-330414.

12. Lestou, V. S. *et al.* Non-random integration of epstein-barr virus in lymphoblastoid cell lines. *Genes Chromosom. Cancer* **8**, 38–48 (1993).

13. Mills, K. H. G. IL-17 and IL-17-producing cells in protection versus pathology. *Nat Rev Immunol* **23**, 38–54 (2023).

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

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

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

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

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

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

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

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

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

23. Kenney, J. L., Guinness, M. E., Curiel, T. & Lacy, J. Antisense to the epstein-barr virus (EBV)-encoded latent membrane protein 1 (LMP-1) suppresses LMP-1 and bcl-2 expression and promotes apoptosis in EBV-immortalized B cells. *Blood* **92**, 1721–1727 (1998).

24. Ribeiro, J. *et al.* P53 deregulation in Epstein-Barr virus-associated gastric cancer. *Cancer Lett* **404**, 37–43 (2017).

25. Sarkozy, C. *et al.* Mutational landscape of gray zone lymphoma. *Blood* **137**,

1765–1776 (2021).

26. Lee, J. E. *et al*. Clinicopathologic and genomic characteristics of mucinous gastric adenocarcinoma. *Gastric Cancer* **25**, 697–711 (2022).

27. Lee, H. S., Chang, M. S., Yang, H.-K., Lee, B. L. & Kim, W. H. Epstein-barr virus-positive gastric carcinoma has a distinct protein expression profile in comparison with epstein-barr virus-negative carcinoma. *Clin Cancer Res* **10**, 1698–1705 (2004).

28. Vereide, D. T. *et al.* Epstein-Barr virus maintains lymphomas via its miRNAs. *Oncogene* **33**, 1258–1264 (2014).

29. Mansouri, S., Pan, Q., Blencowe, B. J., Claycomb, J. M. & Frappier, L. Epstein-Barr virus EBNA1 protein regulates viral latency through effects on let-7 microRNA and dicer. *J Virol* **88**, 11166–11177 (2014).

30. Zhang, J.-Y. *et al.* ebv-circRPMS1 promotes the progression of EBV-associated gastric carcinoma via Sam68-dependent activation of METTL3. *Cancer Lett* **535**, 215646 (2022).

31. Zhang, K. *et al*. Caspases Switch off the m ^{6} A RNA Modification Pathway to Foster the Replication of a Ubiquitous Human Tumor Virus. *mBio* **12**, e01706-21 (2021).

32. Bell, A. & Rickinson, A. B. Epstein-Barr virus, the TCL-1 oncogene and Burkitt's lymphoma. *Trends Microbiol* **11**, 495–497 (2003).

33. Dhanasekaran, R. *et al*. The MYC oncogene - the grand orchestrator of cancer growth and immune evasion. *Nat Rev Clin Oncol* **19**, 23–36 (2022).

34. Guo, R. *et al.* MYC Controls the Epstein-Barr Virus Lytic Switch. *Mol Cell* **78**, 653-669.e8 (2020).

35. Jiang, S. *et al.* The Epstein-Barr Virus Regulome in Lymphoblastoid Cells. *Cell Host Microbe* **22**, 561-573.e4 (2017).

36. Arvey, A. *et al.* An Atlas of the Epstein-Barr Virus Transcriptome and Epigenome Reveals Host-Virus Regulatory Interactions. *Cell Host & Microbe* **12**, 233–245 (2012).

37. Park, G. B. *et al.* CD80 (B7.1) and CD86 (B7.2) induce EBV-transformed B cell apoptosis through the Fas/FasL pathway. *Int J Oncol* **43**, 1531–1540 (2013).

38. Kis, L. L., Takahara, M., Nagy, N., Klein, G. & Klein, E. IL-10 can induce the expression of EBV-encoded latent membrane protein-1 (LMP-1) in the absence of EBNA-2 in B lymphocytes and in Burkitt lymphoma- and NK lymphoma-derived cell lines. *Blood* **107**, 2928–2935 (2006).

39. Weber-Nordt, R. M. *et al.* Constitutive activation of STAT proteins in primary lymphoid and myeloid leukemia cells and in Epstein-Barr virus (EBV)-related lymphoma cell lines. *Blood* **88**, 809–816 (1996).

40. Hsiao, J.-R. *et al.* Constitutive activation of STAT3 and STAT5 is present in the majority of nasopharyngeal carcinoma and correlates with better prognosis. *Br J Cancer* **89**, 344–349 (2003).

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

42. Cohen, J. I. Herpesviruses in the Activated Phosphatidylinositol-3-Kinase-δ

Syndrome. *Front Immunol* **9**, 237 (2018).

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

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

45. Kans, J. Entrez Direct: E-utilities on the Unix Command Line. in *Entrez Programming Utilities Help [Internet]* (National Center for Biotechnology Information (US), 2024).

46. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

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 interactions based on 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. Enrichment networks reveal the core attributes of LMP1. (a) The bar plot shows the numbers of total articles of the top 7 most studied EBV antigens. (b) The enrichment network highlights the top 30 most highly connected LMP1-associated pathways. (c) The enrichment network shows the active transcription of LMP1-targeted pathways. Pathways containing LMP1-induced DEGs are highlighted. (d) The correlation heatmap shows the relationships between genes up-regulated by LMP1 overexpression. (e) The top 3 PPI networks demonstrate the functional clusters of LMP1-targeted host proteins. Differentially expressed genes induced by LMP1 are highlighted in orange.

Supplementary materials

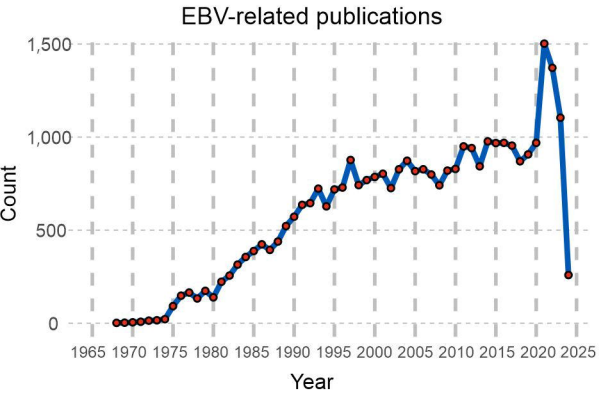
Figure S1. The Sankey diagram illustrates the manual annotation from GPT-4 generated pathways to the corresponding formal KEGG pathway entries.

Figure S2. Enrichment networks reveal the core attributes of LMP1. (a) The enrichment network of all LMP1 targeted host proteins (P.adjust<0.05). (b) The enrichment network shows the top 100 most highly connected LMP1-associated pathways.

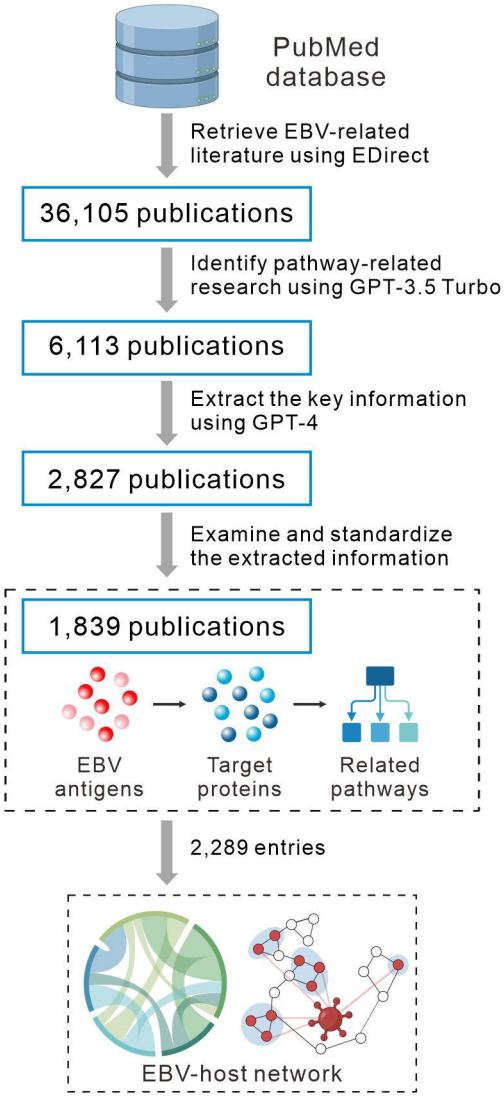
Table S1. Validation for GPT-generated results.

Table S2. Ranking of most studied EBV antigens (with >10 related articles).

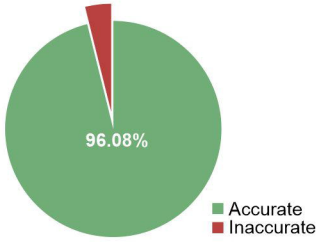
a


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

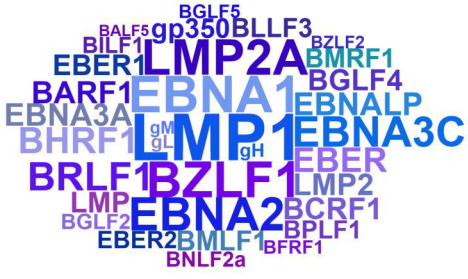
b


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

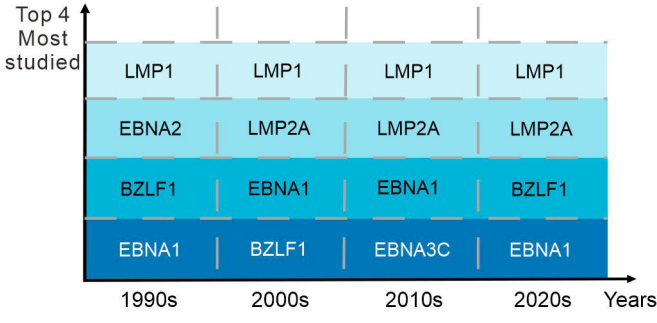
c


GPT-4 prompting accuracy
96.08%
Accurate
Inaccurate

d


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

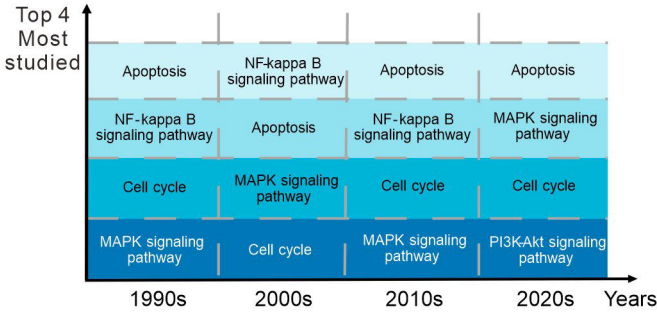
e

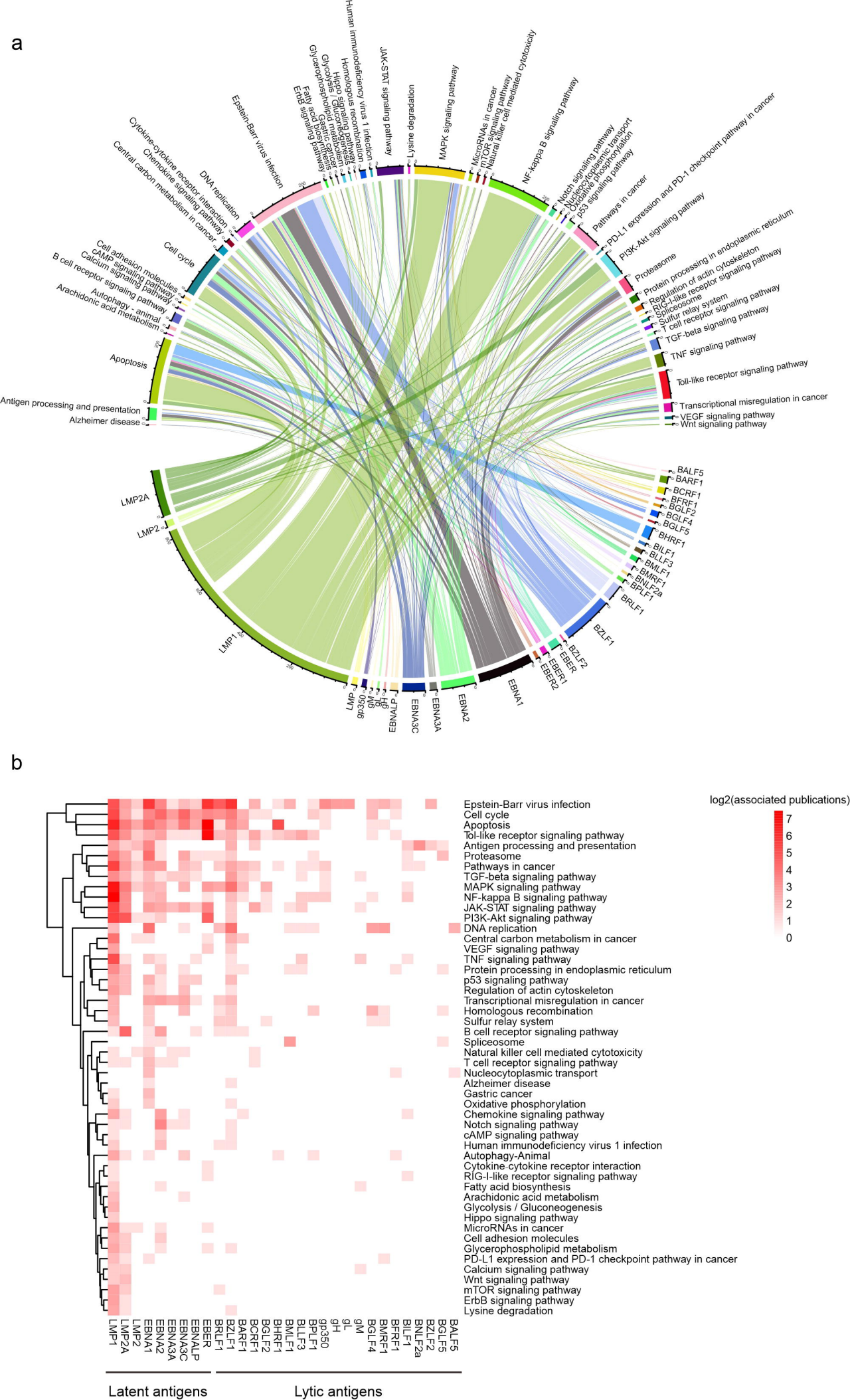

Top 4 Most studied
LMP1 LMP1 LMP1 LMP1
EBNA2 LMP2A LMP2A LMP2A
BZLF1 EBNA1 EBNA1 BZLF1
EBNA1 BZLF1 EBNA3C EBNA1
1990s 2000s 2010s 2020s Years

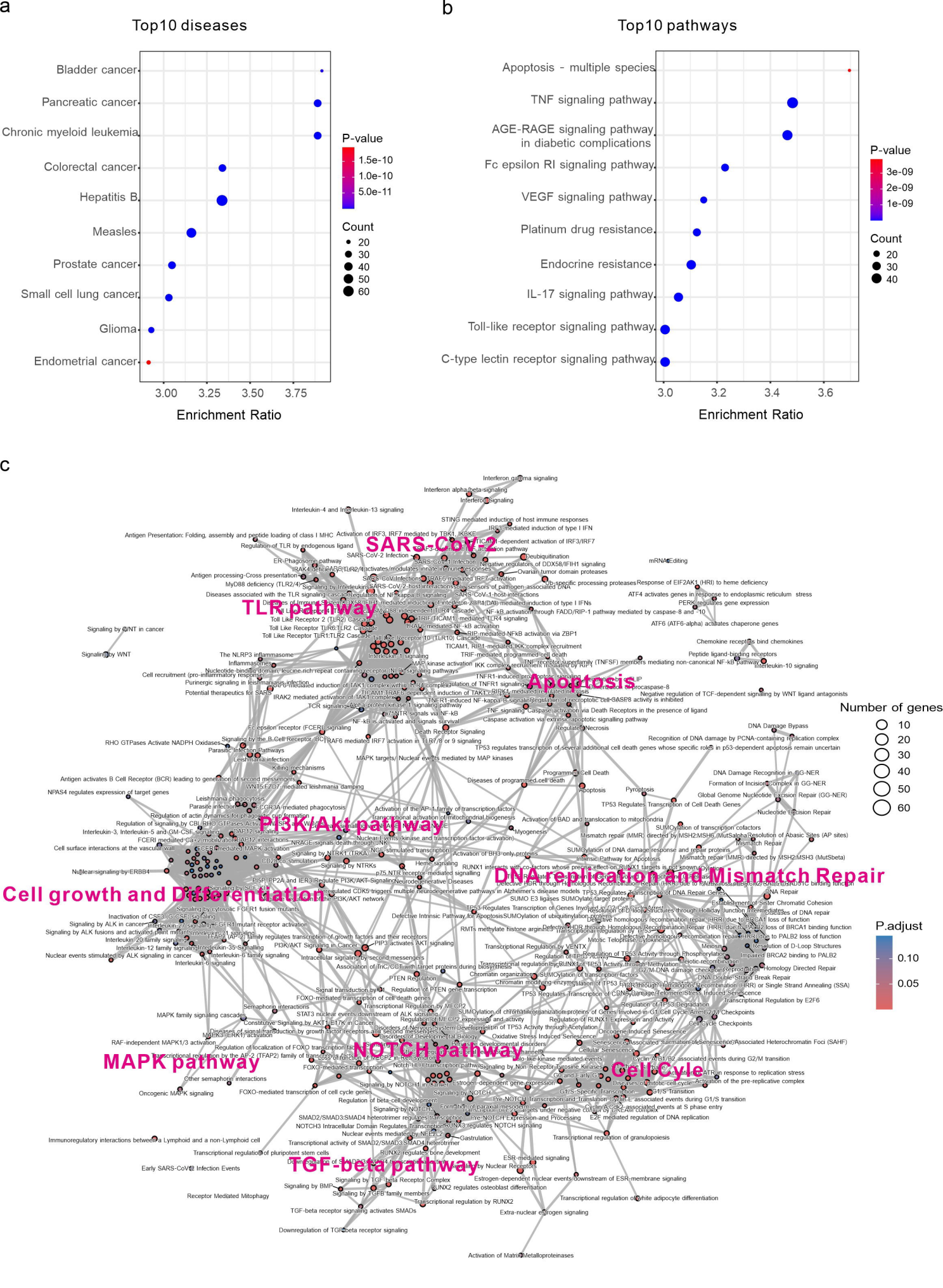
f

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

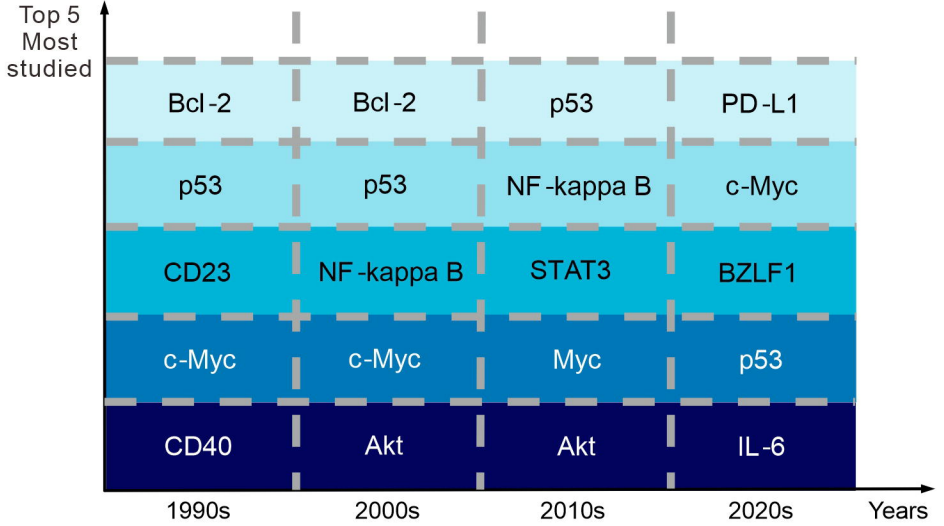
g


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

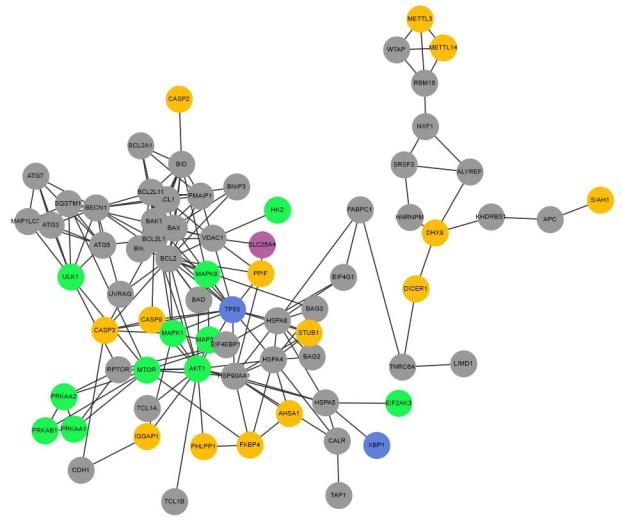

a
Epstein-Barr virus infection
DNA replication
Cytokine-cytokine receptor interaction
Central carbon metabolism in cancer
Chemokine signaling pathway
Cell cycle
Cell adhesion molecules
cAMP signaling pathway
Calcium signaling pathway
B cell receptor signaling pathway
Autophagy - animal
Arachidonic acid metabolism
Apoptosis
Antigen processing and presentation
Alzheimer disease
Human immunodeficiency virus 1 infection
Homologous recombination
Hippo signaling pathway
Glycolysis / Gluconeogenesis
Glycerophospholipid metabolism
Gastric cancer
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
9p350
gH
gL
gM
BGLF4
BMRF1
BFRF1
BLLF1
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
CellCycle
TGF-beta pathway
Number of genes
10
20
30
40
50
60
P.adjust
0.10
0.05

a

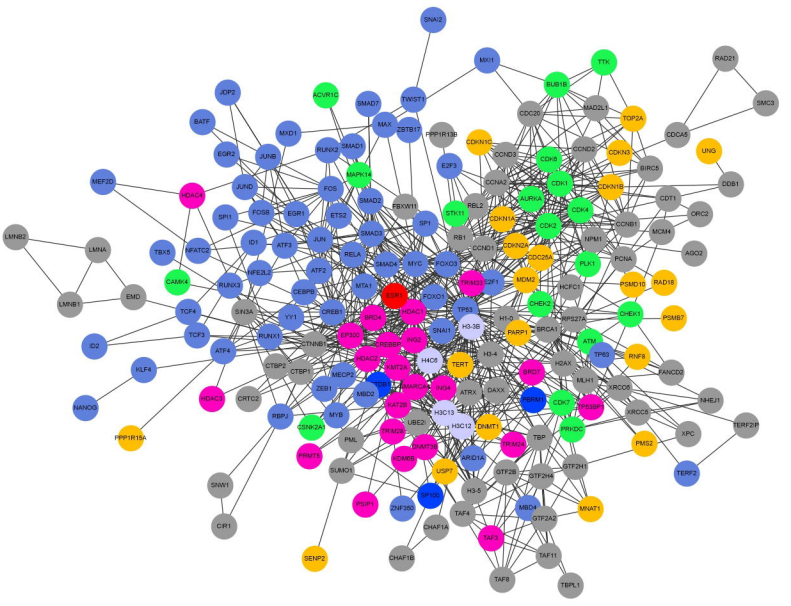


b



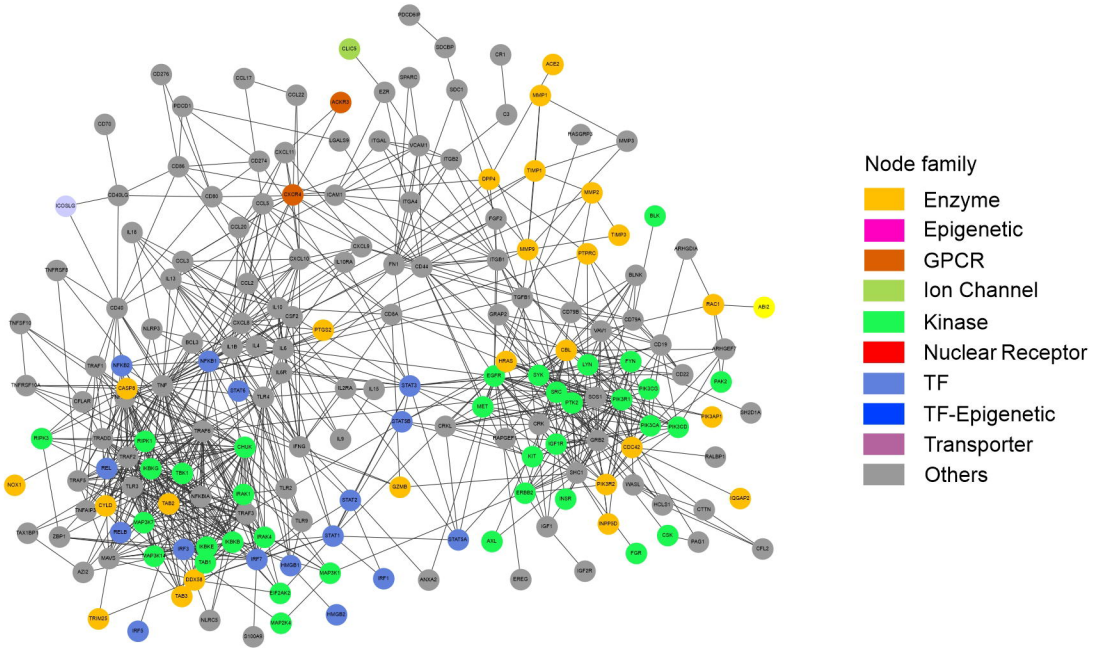
TP53-associated apoptotic signaling pathways

c

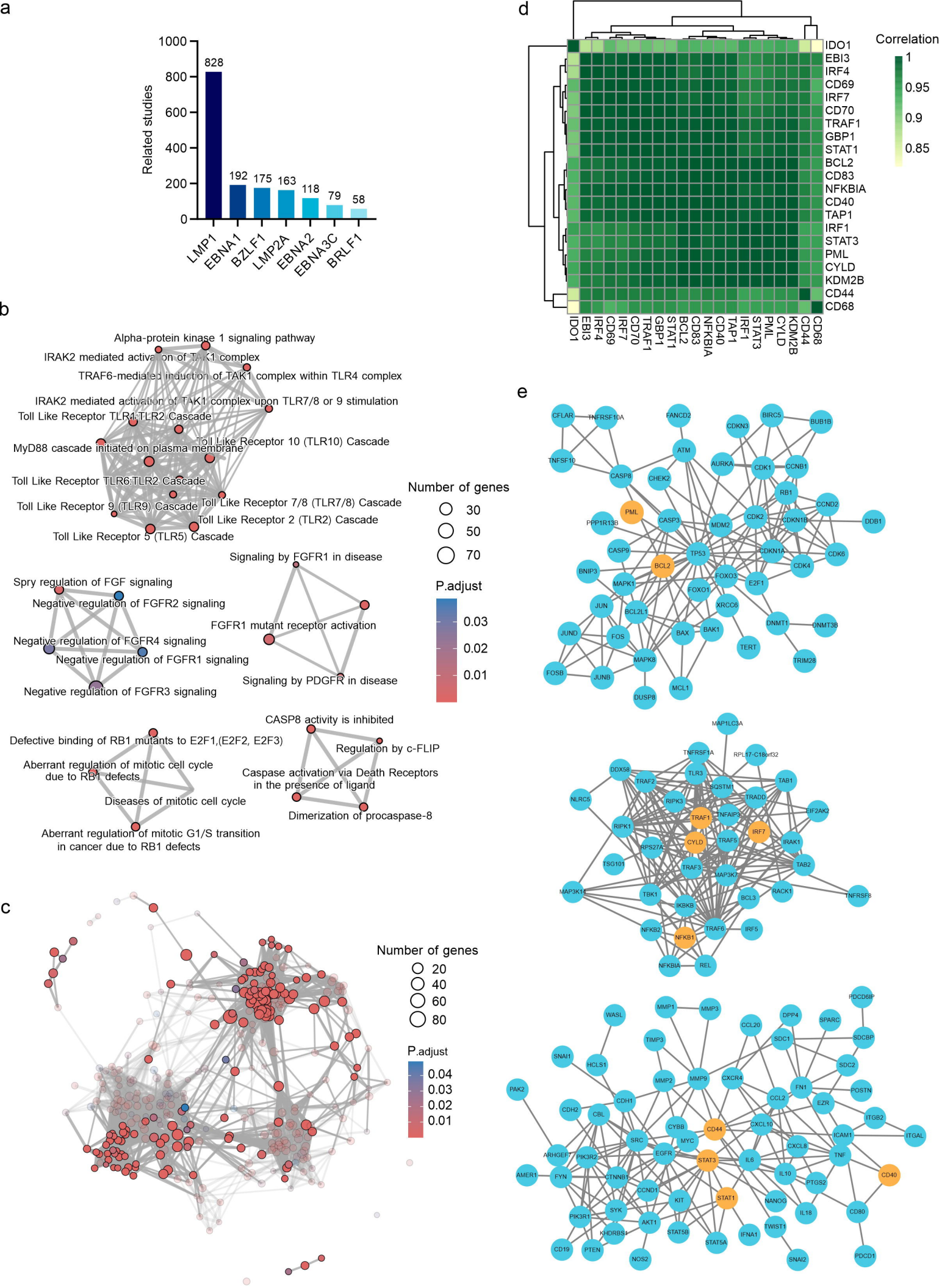


EBV-associated TFs and TF-related factors

d



Immune response and related factors


a
Related studies
1000
800
600
400
200
0
828
192
175
163
118
79
58
LMP1
EBNA1
BZLF1
LMP2A
EBNA2
EBNA3C
BRLF1
b
Alpha-protein kinase 1 signaling pathway
IRAK2 mediated activation of TAK1 complex
TRAF6-mediated induction of TAK1 complex within TLR4 complex
IRAK2 mediated activation of TAK1 complex upon TLR7/8 or 9 stimulation
Toll Like Receptor TLR1:TLR2 Cascade
MyD88 cascade initiated on plasma membrane
Toll Like Receptor 10 (TLR10) Cascade
Toll Like Receptor TLR6:TLR2 Cascade
Toll Like Receptor 9 (TLR9) Cascade
Toll Like Receptor 7/8 (TLR7/8) Cascade
Toll Like Receptor 2 (TLR2) Cascade
Toll Like Receptor 5 (TLR5) Cascade
Signaling by FGFR1 in disease
Spry regulation of FGF signaling
Negative regulation of FGFR2 signaling
FGFR1 mutant receptor activation
Negative regulation of FGFR4 signaling
Negative regulation of FGFR1 signaling
Signaling by PDGFR in disease
Negative regulation of FGFR3 signaling
CASP8 activity is inhibited
Defective binding of RB1 mutants to E2F1,(E2F2, E2F3)
Regulation by c-FLIP
Aberrant regulation of mitotic cell cycle due to RB1 defects
Caspase activation via Death Receptors in the presence of ligand
Diseases of mitotic cell cycle
Dimerization of procaspase-8
Aberrant regulation of mitotic G1/S transition in cancer due to RB1 defects
Number of genes
30
50
70
P.adjust
0.03
0.02
0.01
c
Number of genes
20
40
60
80
P.adjust
0.04
0.03
0.02
0.01
d
Correlation
1
0.95
0.9
0.85
IDO1
EBI3
IRF4
CD69
IRF7
CD70
TRAF1
GBP1
STAT1
BCL2
CD83
NFKBIA
CD40
TAP1
IRF1
STAT3
PML
CYLD
KDM2B
CD44
CD68
e
CFLAR
TNFRSF10A
FANCD2
CDKN3
BIRC5
BUB1B
TNFSF10
ATM
AURKA
CDK1
CCNB1
CASP8
CHEK2
RB1
CCND2
PML
CASP3
MDM2
CDK2
CDKN1B
DDB1
PPP1R13B
CASP9
TP53
CDKN1A
CDK6
BNIP3
BCL2
CDK4
MAPK1
FOXO3
E2F1
FOXO1
JUN
BCL2L1
XRCC6
DNMT1
DNMT3B
JUND
FOS
BAX
BAK1
TERT
TRIM28
FOSB
JUNB
MAPK8
MCL1
DUSP8
MAP1LC3A
TNFRSF1A
RPL17-C18orf32
DDX58
TLR3
TRAF2
SQSTM1
TAB1
NLRC5
RIPK3
TRADD
TRAF1
TNFAIP3
EIF2AK2
RIPK1
IRF7
CYLD
TRAF5
IRAK1
RPS27A
TSG101
TRAF3
TAB2
MAP3K7
MAP3K14
TBK1
RACK1
BCL3
TNFRSF8
IKBKB
NFKB2
TRAF6
IRF5
NFKB1
NFKBIA
REL
PDCD6IP
WASL
MMP1
MMP3
DPP4
SPARC
CCL20
SDC1
SDCBP
TIMP3
SNAI1
HCLS1
SDC2
MMP2
MMP9
CXCR4
FN1
POSTN
PAK2
CCL2
EZR
CDH1
CDH2
CBL
ITGB2
CYBB
CD44
CXCL10
ICAM1
ITGAL
SRC
MYC
CXCL8
ARHGEF7
PIK3R2
EGFR
STAT3
IL6
TNF
AMER1
FYN
CTNNB1
IL10
CD40
PTGS2
CCND1
KIT
STAT1
NANOG
SYK
IL18
CD80
PIK3R1
AKT1
TWIST1
KHDRBS1
STAT5B
IFNA1
CD19
PTEN
STAT5A
NOS2
SNAI2
PDCD1