Bleomycin alters intratumoral immune response of EBV-associated gastric cancer by ENTPD8 and PCOLCE2

Long Chen (chenlong@cmc.edu.cn)
Chengdu Medical College

Research Article

Keywords: EBVaGC, immunotherapy, bleomycin, ENTPD8 and PCOLCE2

Posted Date: May 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2958586/v1

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

Additional Declarations: No competing interests reported.
Bleomycin alters intratumoral immune response of EBV-associated gastric cancer by ENTPD8 and PCOLCE2

Long Chen1,2 *

1. Department of Immunology, School of Basic Medical Sciences, Chengdu Medical College, Chengdu, Sichuan 610500, China.
2. Non-Coding RNA and Drug Discovery Key Laboratory of Sichuan Province, Chengdu Medical College, Chengdu 610500, China.

* Correspondence: Long Chen. E-mail: chenlong@cmc.edu.cn (LC)

Abstract

Background: EBV-associated gastric cancer (EBV aGC) with high PD-L1 level, is most likely to be the next subgroup benefited from immunotherapy. However, complicated with histological and aetiological heterogeneity, tolerance persists which was usually alleviated by clinical adjuvant chemotherapy (bleomycin). Identifying biomarkers of intratumoral immune response was critical for further understanding the direct mechanism of immunotherapy effectiveness.

Method: Firstly, to identify gene sets involved in both GC tumorigenesis and EBV infection, a transcriptome sequencing data (GSE51575) was collected for different expression gene (DEG) screening and functional enrichment analysis. Through constructing a prognostic model based on 25 repeated DEGs and evaluating immune correlations subsequently, the influence of ENTPD8 and PCOLCE2 in prognosis and immunotherapy was confirmed. In addition, the binding energy between bleomycin and targets was calculated based on hydrogen bond.

Result: A total of 572 down- and 162 up-regulated genes in normal tissue vs. GC tissue while 196 down- and 240 up-regulated genes in EBVnGC vs. EBV aGC were detected with logFC≥2 and p-value<0.05. Among them, ENTPD8 and PCOLCE2 were reduced in EBV aGC which was associated with prognosis significantly and mediated dysregulation of immune response inversely. Besides, the expression trends of ENTPD8 (positive) and PCOLCE2 (negative) were also opposite when binding to bleomycin with the most stable binding energy -4.589 kcal/mol and -4.025 kcal/mol, respectively.

Conclusion: Summarily, the improvement of immunotherapy caused by bleomycin as an adjuvant chemotherapy drug may mainly depend on the fluctuation of intratumoral immune response in EBV aGC mediated by the expression of ENTPD8 and PCOLCE2.

Keywords: EBV aGC, immunotherapy, bleomycin, ENTPD8 and PCOLCE2
Introduction

Epstein-Barr virus (EBV), also known as human herpesvirus 4 (HHV4), is a double-stranded DNA virus infecting over 90% adult population and has been initially discovered in Burkitt lymphoma[1]. EBV driven multiple factors contribute to tumorigenesis, including hypermethylation of tumor suppressor genes, inflammatory changes in the mucosa, host immune evasion and dysregulation of cell cycle pathway[1]. GC is a major concern globally, has been the second leading cause of cancer-related death with 95,000 new patients diagnosed each year. The overall 5-year survival rate is dismal at 20% with median survival less than 12 months [1]. About 9% of GCs have been identified as EBV-positive. The EBV-associated GC (EBVaGC) might result from the monoclonal proliferation of EBV-infected cells which predominantly localizes in the upper part of the stomach and is often accompanied by mucosal atrophy and intestinal metaplasia [2, 3]. Besides, the typical clinical features of EBVaGC include superficial, depressed, ulcerated, saucer-like tumor, thickening of gastric wall, poorly differentiated adenocarcinoma with lymphocyte infiltration and a lace pattern within the mucosa [4, 5]. A clinicopathological study also showed that the frequency of lymph node metastasis (LNM) is significantly reduced in EBVaGCs associated with lower mortality when compared with EBV-negative controls (EBVnGCs) [6]. Besides, the molecular characteristics of EBVnGCs include the focal amplification of 9p24.1 (encompassing JAK2, PD-L1 and PD-L2), recurrent PIK3CA mutations, extreme DNA hypermethylation (CDKN2A, p16\[^{INK4A}\]) and increased immune infiltrates [7-9]. It suggested that EBVaGC might escape the immunological recognition of T lymphocytes via the PD-1/PD-L1 pathway. Also, it may be regarded as the underlying cause of responding to immunotherapy (immune checkpoint blockade, ICB) in EBVaGC patients [10].

In recent decades, chemotherapy has served as the backbone of the treatment for advanced GC [11]. Bleomycin (BLM) forms a key component of curative regimens for lymphoma and germ cell tumors initially [12, 13]. Recently, it has been shown anti-tumor activity in malignancies, including head and neck, cervix, vulva and gastric cancers [14-17]. BLM, a metallo-glycopeptide antibiotic produced by actinobacterium Streptomyces verticillu, requires a transition metal ion, usually Fe (II) which is necessary for DNA cleavage activity[18]. Cytotoxic activity of BLM is through oxidation in deoxyribose of thymidylate and other nucleotides via its capacity to generate radical oxygen species (ROS), which produced single-strand and double-strand breaks in DNA, chromosomal aberration, gaps, fragments and translocation. Then BLM would lead to extended cell cycle arrest,
apoptosis, and mitotic cell death via some cancer signal transduction pathways such as PI3K-Akt, p53 and PPAR pathways [12, 19]. Liu et al. demonstrated that SMAD3 activity was changed with BLM binding, and subsequently triggered the SMAD cascade response suppressing GC progress [20]. In addition, BLM is also able to induce ROS-mediated endoplasmic reticulum stress and autophagy, which result in the surface exposure of chaperones (calreticulin) and liberation of ATP which then induce immunogenic cell death which relies on the infiltration of CD8\(^+\) T lymphocytes recruited by dendritic cells (DCs), secretion of interferon-\(\gamma\) and expansion of Foxp3\(^+\) regulatory T (Treg) cells induced by TGF-\(\beta\) [14, 21].

Although the known data in patients was limited, pre-clinical data has shown EBVaGC could resistant to current chemotherapy options including docetaxel [22] or 5-FU [23]. EBVaGC is a special subtype, which presents with abundant PD-L1/PD-L2 expression. EBV infection induces immune responses, recruits immune cells and modulates immune-related molecular components[24]. Compared with EBVnGC, EBVaGC shows a high proportion of lymphoepithelioma-like tumors with better survival [25]. The deregulated immune response genes in EBVaGC revealed a higher level of lymphocyte infiltration and activation of the immune checkpoint pathway (PD-1/PD-L1) [26]. Among EBVaGC patients, the PD-L1\(^+\) group character with a ‘hot’ immune microenvironment, less aggressive clinicopathological characteristics, better prognosis, and higher sensitivity to anti-PD-1 immunotherapy such as pembrolizumab [11], nivolumab [27] and avelumab [26].

Immunotherapy mainly induces antitumor effects by regulating the tumor microenvironment (TME) usually depends on the dynamic interaction between tumor cells and immunomodulators [28]. TME, a complex ecosystem consisting of various cell types (tumor infiltrating lymphocytes) and other noncellular components from the extracellular matrix with obvious heterogeneity, can play critical roles in tumorigenesis and immunotherapeutic efficacy [1, 29]. However, the current dilemma facing immunotherapy is tolerance caused by varying degrees of immunosuppression. EBVaGC is often accompanied by more CD8\(^+\) cytotoxic T lymphocytes (CTLs) and mature dendritic cells than EBVnGC. Moreover, the immune escape caused by regulatory T cells and immunosuppressive (M2) macrophages have been recognized to contribute to GC prognosis [30, 31]. Infiltrated M2 macrophage could inhibit T cell function like PD-L1 which leads to immune tolerance and poorer prognosis[32].
In TME, the extracellular adenosine triphosphate (ATP) is released by immune cells and drives immune responses through the P2X and P2Y receptors [33]. The most dominant ectonucleotidases that dephosphorylate nucleotides under physiological conditions are members of the Ecto-nucleoside triphosphate diphosphohydrolase (E-NTPDase) family [34, 35]. They are omnipresent enzymes for hydrolyzing nucleoside triphosphates and nucleoside diphosphates, and they have an extra plasmic orientation [36]. This family of ectonucleotidases is composed of 8 members (NTPDase-1 to -8) [34, 35]. NTPDase -8, located at the plasma membrane, is the latest member to be discovered and is also characterized by 2 plasma membrane spanning domains and an active site facing the extracellular milieu [37]. ENTPD8 could initiate many cellular responses such as the clearance of extracellular ATP which is essential for the prevention of innate intestinal inflammation by inhibiting glycolysis metabolic in immune cells [38].

Fibrotic deposition of collagen is the main factor affecting intratumoral immune cell infiltration. Collagen is secreted as a soluble procollagen molecule with an NH2- (N) and a COOH (C)-terminal propeptide attached [39]. Propeptide removal is considered essential for efficient incorporation of soluble collagen into insoluble collagen fibrils, especially enzymatic cleavage of the C-terminal propeptide by bone morphogenic protein (BMP) 1 [40]. The proteolytic activity of BMP1 is potentiated by 2 extracellular glycoproteins, procollagen C-proteinase enhancer1 (PCPE1) and PCPE2 (PCOLCE2) [41]. PCOLCE2, expressed from the sequence tag at 3q21-q24 chromosomal region, was found to encode a protein with a domain structure similar to that of PCPE1 and 43% identity in amino acid sequence [41]. PCPE proteins appear to control multiple biological events, including matrix assembly, collagen cross-linking, as well as cell adhesion and migration [42, 43].

There is a strong urge to identify biomarkers for predicting the response of ICB because the EBVaGC could be therapeutic targets. In our study, based on the transcriptome sequencing data of normal tissues, GC tissues, EBVaGC and EBVnGC clinical samples from GEO (gene expression omnibus) and TCGA (the cancer genome atlas), differentially expressed genes (DEGs) were screened and performed corresponding functional enrichment analysis. The identified key genes (ENTPD8 and PCOLCE2) involved in both EBVaGC formation and immune response regulation would be regarded as the biomarkers to evaluate immunotherapy response. The new marker may also be applied to treat EBV-related malignancy by developing novel immunotherapy. Subsequently, we analyzed the interaction between BLM and the biomarkers using three-dimensional (3D) protein
modeling and molecular docking analysis. This study attempts to confirm the molecular mechanism of BLM to enhance the EBVaGC immunotherapy response. The results would provide a basis for the wider application of ICBs in the EBVaGC treatment.

**Methods**

**Tissues, cell culture, mRNA microarray and RNAseq**

The transcriptome sequencing dataset GSE51575 was obtained from GEO database [44]. In this study, fresh GC tissues from 203 patients who underwent open surgery before chemotherapy were frozen and stored from the Department of Surgery, Samsung Medical Center. Informed written consent was acquired from patients and this protocol was approved by the Institutional Review Boards. Routine in situ hybridization for EBV-encoded small RNA (EBER) was performed to classify EBVaGC or EBVnGC. After microscopic analyses for tumor necrosis, size, margin, and H&E staining, a total of 14 EBVnGCs and 12 EBVaGCs were selected with eligible RNA quality for this study (Table 1). Total RNA from normal (N) and tumor (T) tissues of each sample was isolated using Trizol extractions (Invitrogen). The whole transcriptome sequencing (Hi RNaseq) were performed by Illumina HiSeq 2000 instrument. The Agilent G4851A platform (Agilent Technologies; CA, USA) was used for subsequent RNA array profiling analysis. Gene expression data analysis was subsequently conducted using the R language.

**Table 1, The clinical information of all samples**

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Sample number of each type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue Type</td>
<td>Gastric normal (n=26); Gastric tumor (n=26)</td>
</tr>
<tr>
<td>EBV Type</td>
<td>EBV (-) tumor (n=14); EBV (+) tumor (n=12)</td>
</tr>
<tr>
<td>Histology</td>
<td>CA tumor (n=17); CLR tumor (n=6); LELC tumor (n=3)</td>
</tr>
<tr>
<td>Lauren Type</td>
<td>Intestinal (n=13); Diffuse (n=8); Mixed (n=3); Indeterminate (n=2)</td>
</tr>
<tr>
<td>Age</td>
<td>50–60 (n=12); 60–70 (n=20); 70–80 (n=12); other (n=8)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male (n=42); Female (n=10)</td>
</tr>
<tr>
<td>Tumor Stage</td>
<td>Stage I A-C (n=1); Stage II A-C (n=10); Stage III A-C (n=10)</td>
</tr>
<tr>
<td>Her2+ Type</td>
<td>Her2+: - (n=49); Her2+: 1+ (n=2); Her2+: 1+++ (n=1)</td>
</tr>
</tbody>
</table>
Background deletion, and quantile normalization were performed using the quantile method. The sequenced reads of RNAseq were aligned to human reference genome (hg19) by using the STAR alignment tool [44]. DEGs of normal tissue vs. tumor tissue and EBVnGC vs. EBVaGC were detected by the empirical Bayes method [45, 46]. The p-value was adjusted for multiple comparisons using the Benjamini-Hochberg procedure [47]. Genes with an adjusted p-value of < 0.05 and logFC ≥ 1.0 were considered as differentially expressed. Analysis and visualization of Gene Ontology terms associated with DEGs were performed with DAVID[48] and ClueGO [49]. The enriched biological GO (BP: biological process; CC: cellular component; MF: molecular function) and pathway terms were identified [50]. One significant DEG set was further applied to BiNGO tool of Cytoscape to visualize significant gene ontology hierarchy and functional interaction maps between DEGs [51]. Some other databases used are listed in Table 2.

<table>
<thead>
<tr>
<th>Database ID</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA</td>
<td><a href="https://www.cancer.gov/">https://www.cancer.gov/</a></td>
</tr>
<tr>
<td>The Human Protein Atlas</td>
<td><a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a></td>
</tr>
<tr>
<td>Expasy</td>
<td><a href="http://web.expasy.org/protparam/">http://web.expasy.org/protparam/</a></td>
</tr>
<tr>
<td>PSIPRED</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
</tr>
<tr>
<td>DNMIVD</td>
<td><a href="http://119.3.41.228/dnmivd/index/">http://119.3.41.228/dnmivd/index/</a></td>
</tr>
<tr>
<td>GSCALite</td>
<td><a href="http://bioinfo.life.hust.edu.cn/web/GSCALite/">http://bioinfo.life.hust.edu.cn/web/GSCALite/</a></td>
</tr>
<tr>
<td>Linked Omics</td>
<td><a href="http://www.linkedomics.org/">http://www.linkedomics.org/</a></td>
</tr>
<tr>
<td>STRING</td>
<td><a href="https://string-db.org/">https://string-db.org/</a></td>
</tr>
<tr>
<td>DrugBank</td>
<td><a href="https://go.drugbank.com/">https://go.drugbank.com/</a></td>
</tr>
<tr>
<td>SWISS-MODEL</td>
<td><a href="https://swissmodel.expasy.org/interactive#alignment">https://swissmodel.expasy.org/interactive#alignment</a></td>
</tr>
<tr>
<td>GEPIA</td>
<td><a href="http://gepia.cancer-pku.cn/index.html">http://gepia.cancer-pku.cn/index.html</a></td>
</tr>
<tr>
<td>AutodockVina 1.2.2</td>
<td>(<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>)</td>
</tr>
<tr>
<td>TIP</td>
<td><a href="http://biocc.hrbmu.edu.cn/TIP/index.jsp">http://biocc.hrbmu.edu.cn/TIP/index.jsp</a></td>
</tr>
</tbody>
</table>
**Survival curves**

Overall survival analyses were performed using the R package survival, and the patients were dichotomized based on median expression. Kaplan-Meier estimator of survival was used to construct the survival curves. Log-rank tests (corresponding to a two-sided z test) were used to compare overall survival between patients in different groups, and hazard ratio (HR) (95% confidence interval) was provided for comparison of two groups. The p-values were adjusted for multiple testing based on the false discovery rate (FDR) according to the Benjamini-Hochberg method. Proportional hazard assumptions were tested.

**Analysis of gene expression and function in GC based on TCGA database**

Based on normalized and batch corrected RSEM mRNA expression, differential expression analysis for STAD was provided, which have over ten paired tumor and normal samples. The fold change is calculated by mean (Tumor) / mean (Normal), and the p-value was estimated by t-test and was further adjusted by FDR.

For expression and survival analysis, the clinical data was used while the samples with a competing risk for death were filtered out. Merged mRNA expression and correlated clinical survival data, the GC samples were divided into high and low expression groups based on median mRNA value. Then, the R package survival was used to fit the survival time and survival status within two groups. The Logrank tests were performed for each gene in GC. In analysis between expression and subtype or stage, mRNA expression and clinical stage data were merged by sample barcode. The subtype and stage subgroup must have at least 5 samples. The pathologic stage and clinical stage classify samples into Stage I, II, III, and IV. Trend analysis of stages was performed by Mann-Kendall Trend Test.

The correlation of expression and pathway activity was estimated by the gene different expression between pathway active group and inhibited group, which are defined by median pathway scores. Reverse phase protein array (RPPA) data from (TCPA database) were used to calculate the pathway activity score of GC. The famous cancer-related pathways included are TSC/mTOR, RTK, RAS/MAPK, PI3K/AKT, Hormone ER, Hormone AR, EMT, DNA Damage Response, Cell Cycle, and Apoptosis pathways. Samples were divided into 2 groups (high and low) by median gene expression, the difference of pathway activity score was defined by the student t-test. The p-value was adjusted by FDR which is significant when ≤0.05. Once the score (high
level)>score (low level), gene A may activate this pathway, otherwise have an inhibitory effect.

**Molecular docking of bleomycin with proteins**

To analyze the binding affinities and modes of interaction between the drug candidate and their targets, AutodockVina 1.2.2, a silico protein-ligand docking software was employed [52]. The molecular structures of BLM (DrugBank Accession Number: DB00290) was retrieved from DrugBank database. Since the research about ENTPD8 and PCOLCE2 protein were limited, their 3D structures are not clear without fixed PDB files. The 3D coordinates of ENTPD8 and PCOLCE2 were modelling by SWISS-MODEL based on the amino acid sequences and the PDB files were downloaded for docking. All protein and molecular files were converted into PDBQT format with all water molecules excluded and polar hydrogen atoms were added. The grid box was centered to cover the domain of each protein and to accommodate free molecular movement. The grid box was set to 30 Å × 30 Å × 30 Å, and grid point distance was 0.05nm. Molecular docking studies were performed by Autodock Vina 1.2.2 and the results with top 9 binding energy were listed.

**Statistics**

Statistical analysis were performed using SPSS for Windows V.13.0 (SPSS Inc., Chicago, IL, USA) and R version 3.4.0. The associations of ENTPD8 and PCOLCE2 level by immunohistochemistry staining with clinicopathological characteristics were evaluated using the chi-square test. The prognostic relevance of ENTPD8 and PCOLCE2 level was assessed using univariate and multivariate COX regression analysis. Survival curves were plotted by Kaplan-Meier method and log-rank test. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analysis was calculated by using the R package “clusterProfile” based on gene expression matrix. All statistical hypothesis performed in this paper were two-tailed, and a p-value≤0.05 was considered significant.

**Results**

The dysregulation of immune system in EBVaGC

As described in the Methods part, a transcriptome sequencing dataset (GSE51575) from GEO was obtained for analyzing the mechanism and characteristics of EBVaGC. In this protocol, a total of 52 clinical samples were collected which contained 26 pairs of GC tissues and corresponding tumor-adjacent tissues. Furthermore, all tumors were classified as EBVnGC (n=14) and EBVaGC (n=12) ([Figure S1A](#)). The quality control showed that the sequencing accuracy was similar, 2
absorption peaks were found at both intensity 5 and intensity 10 from the expression density cartogram of all samples (Figure S1B). The PCA dimensionality reduction clustering based on calibration data showed that normal tissues and cancer tissues could be divided into 2 distinct groups. Though EBCnGC and EBVaGC also show slight differences, the spatial distribution is relatively close (Figure S1C). To screen the key molecules and molecular mechanisms involved in EBVaGC carcinogenesis, the variance of transcript profile among normal tissue vs. GC tissue or EBVnGC vs. EBVaGC were compared respectively (Figure S1D and E). The DEGs were identified with logFC≥1 and p-value≤0.05 and the top 20 significant DEGs were labeled with gene symbols.

Abnormal gene transcription often leads to dysregulation of intracellular signaling pathways (activation or inhibition) and is accompanied by significant alteration in biological functions. To this end, we performed gene set enrichment analysis (GSEA) twice covering the whole gene transcript of normal tissue vs. GC tissue or EBVnGC vs. EBVaGC, respectively. Finally, the 15 significantly enriched pathways (p-value<0.05) of normal tissue vs. GC tissue were almost correlative with energy metabolism such as fatty acid metabolism (p-value=0.0061), glycine serine and threonine metabolism (p-value=0.0106) and pentose phosphate pathway (p-value=0.0378) (Figure 1A). Most genes participated in these pathways were down-regulated in GC, which might suggested that energy metabolism was inhibited which is consistent with the metabolic reprogramming theory of cancer. Also, several cancer pathways were enriched like colorectal cancer (Figure S2A). Furthermore, EBVaGC had significant alterations in energy metabolism pathways with down-regulated genes. Wang et al [53] characterized EBVaGC-specific cellular pathways and detected the alterations in macromolecular biosynthetic processes such as carbohydrate, lipid and protein. Deregulation of cholesterol transport and lipoprotein clearance pathways is also evident in EBVaGC [44]. With EBV infection, besides the pathogenic microbial pathways were enriched (Figure S2B), 10 significantly enriched immune pathways (p-value<0.05) of EBVnGC vs. EBVaGC such as primary immunodeficiency (p-value=0.006), the intestinal immune network for IGA production (p-value=0.0178) and antigen processing and presentation (p-value=0.0401) were also detected (Figure 1B and S2C). According to the enrichment score, it would suggest that in EBVaGC, the immune regulatory mechanism was active.
The significantly enriched GSEA items of normal tissue vs. GC tissue and EBVnGC vs. EBVaGC. The lines with various colors represent different pathways. The gray curve represents the sequencing of pathway-related genes according to their expressions. C, The immune score heatmap shows the 23 activity scores of anti-cancer immunity across 7-step cancer-immunity cycle. Each row represents one immune status and the gene number of which annotated in the reference was shown in the row names. D, The bar plot shows the relative proportion of infiltrating immune cells of all samples. Each color represents one type of immune cell.

The different intratumoral immune subtype is often associated with the progression and treatment of GC. Thus, the immune score of all 52 samples was calculated within 7 steps containing release of cancer antigens, cancer antigen presentation, priming and activation, trafficking of immune cells to tumors, infiltration of immune cells into tumors, recognition of cancer cells by T cells and killing of cancer cells (Figure 1C). Compared with EBVnGC, the immune score of all
steps in EBVaGC was higher nearly. In addition, based on the specific biomarkers, the proportion of immune cells was shown in the histogram (Figure 1D). Among them, the CD8⁺ naïve cells were reduced while the CD8⁺ memory cells, CD8⁺ effector cells and T_{reg} cells were recruited in EBVaGC. In one word, the results would further confirm that the EBVaGC subtype might own more active immune response which leads to a better prognosis and immunotherapy sensitivity.

**Enrichment analysis of DEGs in normal tissue vs. GC tissue and EBVnGC vs. EBVaGC compare.**

To identify the critical biomarkers that participated in immune regulation, the DEGs of normal tissue vs. GC tissue and EBVnGC vs. EBVaGC were screened with logFC≥2.0 and p-value≤0.05. In normal tissue vs. GC tissue comparison, a total of 734 DEGs (572 down-regulated and 162 up-regulated) were detected while there were 436 DEGs (196 down-regulated and 240 up-regulated) in EBVnGC vs. EBVaGC (Figure 2A and B). Moreover, there was no significant correlation between the DEG expression trend and clinical indicators such as age and gender. Following by the enrichment analysis of 2 DEG sets (Figure 2C and B, Figure S3 and S4), the top 10 items of BP, MF, CC and KEGG pathway were listed. The DEGs in GC might affect functions of gastric like digestion (BP), gastric acid secretion (KEGG) and collagen-containing extracellular matrix (CC). However, the DEGs between EBVnGC and EBVnGC were enriched in various immune-related functions such as regulation of T cell activation (BP), antigen processing and presentation of exogenous peptide antigen (BP), MHC class II receptor activity (MF) and intestinal immune network for IgA production (KEGG). Interestingly, the DEG enrichment result was similar to the GSEA of whole genes above which further showed the immune importance of DEGs in EBVaGC subtype.
Repeated DEGs both in normal tissue vs. GC tissue and EBVnGC vs. EBVaGC

As well known, the EBVaGC is a specific subtype with abnormal immune regulation. To classify the crucial markers associated with both tumorigenesis and immunoreaction exposed to EBV, a total of 25 repeated DEGs (16 dual down-regulated and 9 dual up-regulated) were recognized and the transcriptome sequencing data of all genes (down-regulated: DUOX2, CA9, COL2A1, SOWAHA, RGN, AP0A1, CKB, TPD52L1, WFDC2, TM4SF4, HMGCS2, SLC15A1, ENTPD8, PCOLCE2, FNDC5 and PPP1R9A; up-regulated: MMP7, PLA2G7, NPSR1, APOC1, CXCL9, CHI3L1, IDO1, PLA2G2A and CXCL10) were shown respectively (Figure 3A-C). Furthermore, the DEGs own higher relevance in normal tissue vs. GC tissue when compared with EBVnGC vs. EBVaGC (Figure 3D and E). It may suggest that the repeated DEGs would like to share some regulatory mechanism and function together in GC formation, while the correlation was reduced due to the EBV infection could cause abnormal transcription as an exogenous variate. To further
understand the functions of repeated DEGs in EBVaGC, the enrichment analysis of down- and up-regulated genes was performed again (Figure 3F and G, Figure S5 and S6). Among KEGG results, pathways associated with metabolism (metabolic pathway, p-value=0.0016) and immunity (cytokine-cytokine receptor interaction, p-value=0.0167) were the most significant. Notably, the enriched functions of repeated DEGs were also similar to the GSEA result above illustrating that these 25 genes as a whole would play critical regulatory roles in EBVaGC formation and immune response.

Figure 3, Enrichment analysis of 25 repeated DEGs. A, The venn diagram shows the dual down-and up-regulated genes of normal tissue vs. GC tissue and EBVnGC vs. EBVaGC. B and C, Based on the sequencing data, histograms
were drawn to show the expression levels of these 25 repeated genes in different sample types. The labels were listed on the right. D and E. The heatmaps shows the correlations between each gene pair. The red and blue circle indicate positive and negative correlation in gene expression, respectively. The size and color of circle indicate degree and significance of correlation, respectively. F and G. A table lists the significantly enriched KEGG items of down-and up-regulated genes with a panel shown associated genes.

**The roles of repeated DEGs in cancer-related pathways**

Though it was demonstrated that the 25 screened repeated genes might mediate pathways associated with metabolism and immunity which have been verified to play key roles in GC [54, 55], it was necessary to establish the connection between DEGs and cancer related pathways. According to the mRNA expression profile and clinical information of GC samples (n=380), the fold change of DEGs in normal tissue vs. GC tissue and different subtypes compared were calculated (Figure 4A and B). Nevertheless, some genes such as DUOXA were detected without significantly different expression may be due to the sample number was too larger with heterogeneity. In addition, the expression trend among Stage I-IV was also visualized with a line chart (Figure 4C). Among them, SOWAHA and ENTPD8 were reduced abidingly while the CKB and SLC15A1 were induced from Stage I to Stage IV. The fluctuation of gene level will directly affect the state of intracellular pathways. To characterize the roles of repeated genes in GC, the correlation between gene expression and the status of 10 classical pathways involved in cancer development was calculated (Figure 4D). For example, the apoptosis pathway was the most associated with DEGs. The PI3K-AKT (WFDC2, SLC15A1, RGN and APOA1) and RASMAPK (TM4FS4, RGN and DUOXA2) pathways were only detected to activate. Interestingly, both PI3K and MAPK have been confirmed to mediate the immune system in GC [56, 57].
Figure 4. Analysis of 25 repeated genes in GC based on TCGA. 

A, The expression fold change of each gene in normal tissue vs. GC tissue. The bubble plot presents the logFC and FDR via color and size. The row list gene symbols. The bubble color from purple to red represents the fold change between normal vs. GC. The size of dots is positively correlated with the FDR significance. 

B, The expression fold change of each gene in compares of different subtypes. The bubble color from white to red represents the significance of FDR; the bubble size is positively correlated with the significance. The black outline border of bubbles indicates FDR $\leq$ 0.05. 

C, Trend plot presents the trend of gene expression from stage I to stage IV. The trend line color from blue to red represents the tendency from fall to rise. 

D, Percentage of cancers in which a gene has an effect (FDR $\leq$ 0.05) on the pathway among selected cancer types, the number in each cell indicates the percentage.
**Prognostic model of 25 repeated DEGs in GC**

As the analysis results above, these 25 DEGs may not only regulate cancer-related pathways, but they can also mediate the intratumoral immune system. Therefore, to estimate the value of this DEG set in the prognosis of GC patients, a mathematical model based on gene level was constructed with the LASSO Cox regression model. In this model, all GC patients could be divided into 2 groups with different risk scores (high and low risk) based on the expression of PCOLCE2, ENTPD8 and APOA1 (lambda.min=0.0444) (Figure 5A-C). Among them, both PCOLCE2 and APOA1 were positively correlative with risk score while ENTPD8 was negative. It indicated that the expression levels of PCOLCE2, ENTPD8 and APOA1 are the most important factors for prognosis in this model. The formula of risk score was:

\[
\text{Riskscore} = (0.0306) \times \text{APOA1} + (0.0565) \times \text{ENTPD8} + (0.116) \times \text{PCOLCE2}
\]

According to the risk score, the patients were averagely divided into high-risk and low-risk groups (n=185). The median survival times were 1.7 and 4.6 years, respectively (Figure 5D). The Log-rank p-value = 5.64e-05, HR (high group) = 2.001 and 95% CI (1.428, 2.804). It indicated that the prognostic model can divide patients into 2 groups with significant differences in survival. To evaluate whether this model can objectively predict prognosis, the AUC values of 1-years (0.605, 95% CI (0.545-0.665)), 3-years (0.696, 95% CI (0.618-0.773)), 5-years (0.615, 95% CI (0.48-0.751)) and 9-years (0.661, 95% CI (0.401-0.922)) were calculated (Figure 5E). Among them, the AUC value of 3-years in this prognostic model was about 0.7 indicating that PCOLCE2, ENTPD8 and APOA1 had a better prognostic evaluation effect and clinical reference value.
Figure 5, The prognostic mode based on 25 repeated DEGs. A, The coefficients of selected features are shown by lambda parameter. The abscissa represents the value of lambda, and the ordinate represents the coefficients of the independent variable. B, The relationship between partial likelihood deviance and log (λ) is plotted using LASSO Cox regression model. C, The Riskscore, survival time and survival status of selected dataset. The top scatterplot represents the Riskscore from low to high. Different colors represent different groups. The scatter plot distribution represents the Riskscore of different samples correspond to the survival time and survival status. The bottom heatmap show the expression of PCOLCE2, ENTPD8 and APOA1. D, Kaplan-Meier survival analysis of different groups was made by log-rank test. HR (High exp) represents the hazard ratio of the low-expression sample relatives to the high-expression sample. E, The ROC curve and AUC of the gene. The higher values of AUC corresponding to higher predictive power.

The expression of ENTPD8 and PCOLCE2
During the construction of the prognostic model, it was found that not all provided DEGs could effectively contribute to the risk score of patients, suggesting that some of them would not significantly change the overall survival of patients. Thus, the survival curves of each gene were drawn based on expression level, while only ENTPD8 (Logrank p-value=0.014) and PCOLCE2 (Logrank p-value=0.00093) could significantly lead to a poorer prognosis of GC with opposite level (Figure S7, Figure 6A). No matter in mRNA or protein level, it was detected that the content of both ENTPD8 and PCOLCE2 was reduced in GC tissue (Figure 6B and C). In addition, according to TCGA dataset of GC, both ENTPD8 and PCOLCE2 were also down-regulated in EBVaGC subtype (Figure 6D). Meanwhile, the expression trends of ENTPD8 and PCOLCE2 were not significant in clinical Stage I, II, III, and IV (Figure 6E). It might suggest that both ENTPD8 and PCOLCE2 could be recognized as specific biomarkers of EBVaGC which also could affect prognosis. Similarly, after calculating the pathway activation score (Figure 6F), it was found that ENTPD8 could active cell cycle pathway (p-value=1.5e-02) while inhibiting EMT (p-value=3.1e-04) and Hormone ER (p-value=3.5e-02) pathways. Also, PCOLCE2 could inhibit Hormone AR (p-value=2.5e-02) and apoptosis (p-value=2.2e-06) pathway.
Figure 6. The expression of ENTPD8 and PCOLCE2 in GC. A, The overall survival curves of ENTPD8 and PCOLCE2 in GC. The red and blue lines represent the samples with a higher or lower gene level, respectively. The legend was listed on the right. B, The boxplots show the mRNA expression of ENTPD8 and PCOLCE2 in normal tissue and GC. C, The immunohistochemical pictures of normal and GC tissues (n=2) which were stained by ENTPD8 and PCOLCE2 antibodies. The target protein appears brown. D and E, The expression of ENTPD8 and PCOLCE2 in different molecular subtypes (CIN, EBV, GS and MSI) and Stage I-IV. F, The pathway activation score of ENTPD8 and PCOLCE2 in cell cycle, EMT, Hormone ER, Hormone AR and apoptosis pathways.

The immune regulation mediated by ENTPD8 and PCOLCE2 in GC

In the present study, the most significant difference between EBVnGC and EBVaGC was the intratumoral immune response which would determine the immunotherapy effectiveness. According to a dataset of GC with anti-PD-1/PD-L1 (PRJEB25780), both ENTPD8 (Wilcoxon p-value=0.51) and PCOLCE2 (Wilcoxon p-value=0.19) were down-regulated in responder group without significance (Figure 7A and B), suggesting ENTPD8 and PCOLCE2 might be regarded as targets and biomarkers of immunotherapy in EBVaGC. To study the mechanism of ENTPD8 and
PCOLCE2 in GC immunity regulation comprehensively, the correlation between the expression level and various immune events was calculated (Figure 7C). Interestingly, the result showed that ENTPD8 and PCOLCE2 own opposite correlations of lymphocyte infiltration and immune factor expression. For example, among all lymphocytes, only the infiltration of macrophage would affect prognosis (Logrank p-value=0.004) (Figure 7D, Figure S8). Furthermore, through calculating the correlation coefficient, it was shown that ENTPD8 (partial Cor=-0.299; p-value=4.60e-09) was negative with macrophage infiltration while the PCOLCE2 (partial Cor=0.34; p-value=1.88e-11) was positive (Figure 7E and F).

![Figure 7. The immunoscore of ENTPD8 and PCOLCE2 expressions. A and B, The expression of ENTPD8 and PCOLCE2 in GC samples who were divided into non-response (blue) and response (red) groups. C, The heatmap shows the correlation between ENTPD8 and PCOLCE2 levels and various immune indexes. The red box represents positive correlation and the blue one represents negative correlation. D, The survivorship curves of GC show the correlation between macrophage infiltration and prognosis. The red and blue curves represent the higher or lower]
level of infiltration. E and F, The correlation between macrophage infiltration and ENTPD8 or PCOLCE2 expression. Each point represents a single GC samples. The blue lines represent the correlation coefficient.

The binding sites of BLM in protein structure of ENTPD8 and PCOLCE2

Recently, it has been identified that the expression levels of ENTPD8 and PCOLCE2 might influence EBVaGC immunity, which also plays role in anti-PD-1/PD-L1 immunotherapy. Screening drugs which could regulate the expression or stability of ENTPD8 and PCOLCE2 is expected to be used as adjuvant agents to enhance the effectiveness of immunotherapy and reduce resistance. According to previous studies, the correlations between drugs and levels of ENTPD8 and PCOLCE2 were calculated (Figure 8A). A total of 21 drugs were identified to regulate levels of ENTPD8 and PCOLCE2 simultaneously. Among them, BLM in 50μM is the most effective with FDR≤0.05. BLM could inhibit DNA metabolism and is used as an antineoplastic with 45% systemic absorption. The molecular formula is C$_{55}$H$_{84}$N$_{17}$O$_{21}$S$_{3}$ and the molecular weight is 1415.552 (Figure 8B). Due to the lack of protein structure studies, ENTPD8 and PCOLCE2 have no definite 3D conformation. To perform molecular-protein docking analysis, 3D structure modeling of ENTPD8 and PCOLCE2 was performed based on amino acid sequences and PDB files describing spatial structure were obtained (Figure 8C and D, Figure S9). To evaluate the affinity of BLM for ENTPD8 and PCOLCE2, molecular docking analysis was performed. The binding poses and interactions were obtained with Autodock Vina v.1.2.2 and binding energy for each interaction was generated (Figure 8E and F, Table 3 and 4, Figure S10). It was shown that BLM bound to its protein targets through visible hydrogen bonds and strong electrostatic interactions. Moreover, the hydrophobic pockets of each target were occupied successfully by the five candidate drugs. For ENTPD8, the target site of BLM with the lowest binding energy of -4.589 kcal/mol was constituted with 4 hydrogen bonds (TRP440, THR125, GLU168 and GLU168), indicating highly stable binding. Similarly, the most stable binding site of BLM in PCOLCE2 was constituted with 2 hydrogen bonds (CYS86 and LEU85) with -4.025 kcal/mol binding energy.

Table 3, The number and location of hydrogen bonds formed between bleomycin and ENTPD8 in molecular docking analysis
<table>
<thead>
<tr>
<th>Binding energy</th>
<th>Number</th>
<th>Location of hydrogen bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.589 kcal/mol</td>
<td>n=4</td>
<td>TRP440, THR125, GLU168, GLU168</td>
</tr>
<tr>
<td>-4.561 kcal/mol</td>
<td>n=3</td>
<td>ILE438, ASP48, HIS53</td>
</tr>
<tr>
<td>-4.524 kcal/mol</td>
<td>n=1</td>
<td>SER52</td>
</tr>
<tr>
<td>-4.465 kcal/mol</td>
<td>n=1</td>
<td>ASN354</td>
</tr>
<tr>
<td>-4.387 kcal/mol</td>
<td>n=2</td>
<td>SER86, LEU130</td>
</tr>
<tr>
<td>-4.326 kcal/mol</td>
<td>n=1</td>
<td>LEU130</td>
</tr>
<tr>
<td>-4.304 kcal/mol</td>
<td>n=3</td>
<td>TYR357, TRP440, ASP205</td>
</tr>
<tr>
<td>-4.275 kcal/mol</td>
<td>n=3</td>
<td>SER55, HIS53, TYR357</td>
</tr>
<tr>
<td>-4.265 kcal/mol</td>
<td>n=2</td>
<td>ASP205, ASP205</td>
</tr>
<tr>
<td>-4.025 kcal/mol</td>
<td>n=2</td>
<td>CYS86, LEU85</td>
</tr>
<tr>
<td>-3.84 kcal/mol</td>
<td>n=2</td>
<td>ASN84, ASN84</td>
</tr>
<tr>
<td>-3.633 kcal/mol</td>
<td>n=2</td>
<td>ASP151, SER164</td>
</tr>
<tr>
<td>-3.624 kcal/mol</td>
<td>n=4</td>
<td>SER117, VAL116, TYR153, GLY155</td>
</tr>
<tr>
<td>-3.612 kcal/mol</td>
<td>n=1</td>
<td>CYS231</td>
</tr>
<tr>
<td>-3.61 kcal/mol</td>
<td>n=2</td>
<td>CYS154, PRO146</td>
</tr>
<tr>
<td>-3.594 kcal/mol</td>
<td>n=2</td>
<td>TYR212</td>
</tr>
<tr>
<td>-3.538 kcal/mol</td>
<td>n=2</td>
<td>LEU85, VAL221</td>
</tr>
<tr>
<td>-3.525 kcal/mol</td>
<td>n=0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4, The number and location of hydrogen bonds formed between bleomycin and PCOLCE2 in molecular docking analysis
Figure 8. Molecular docking analysis of BLM in 3D structures of ENTPD8 and PCOLCE2. A, The bubble plot is used to summarize the correlations between BLM and targets. Blue bubbles represent negative correlations, red bubbles represent positive correlations. Bubble size is positively correlated with the FDR significance. The black outline border indicates FDR ≤ 0.05. B, The 2D chemical formula of BLM. C and D, The 3D structure of ENTPD8 and PCOLCE2 were modeled by SWISS-MODEL. E and F, The top 3 stable binding modes of BLM to ENTPD8 and PCOLCE2 were shown via cartoon representation, respectively. In the local magnification, the dashed lines represent hydrogen bonds.

Discussion

Our study portrayed the transcriptome landscape of tumor-neighboring morphologically
normal tissues and dysplasia samples containing EBVnGCs and EBVaGCs. An immune-related gene set containing 25 repeated DEGs was identified as a regulator of immunotherapy in EBVaGC. EBVaGC is an important subtype supposed to be potentially sensitive to immunotherapy like ICB (PD-1/PD-L1) due to the higher frequency of PD-L1 (30%) and relatively ‘hot’ immune TME than EBVnGC [7, 44]. It was speculated that EBVaGC with higher PD-L1 level is more likely to be immune-inflamed and more sensitive to ICB. However, the impact of PD-L1 expression on prognosis remains controversial in GC [58]. Once the intratumoral T cells are activated, PD-1 is strictly regulated and strongly induced which could engage with ligand (PD-L1) to form an immunosuppressive TME [59]. The PD-L1+ EBVaGC showed deeper invasion and poorer prognosis with a higher degree of lymphocyte infiltration [60, 61]. By collecting and analyzing public clinical data, Wei et al have provided an important perspective on biomarkers for immunotherapy in EBVaGC. The ORR was 0% in PD-L1- EBVaGC, while 63.3% in PD-L1+ EBVaGC. A significantly prolonged PFS was also observed in PD-L1+ EBVaGC [62].

Through in-depth exploration, the TME is complex and diverse which also has been shown to have important impacts on prognosis and the efficacy of immunotherapy [63, 64]. However, based on several clinical reports [26, 65-67], the efficacy of ICB may not be universally satisfactory in all EBVaGC and still faces many challenges, such as the underlying TME biomarkers have not been reported. In this study, the characteristics of TME in EBVaGC patients were analyzed according to the expression of ENTPD8 and PCOLCE2. It will deepen the understanding of the effects of the TME on antitumor and immune responses and will guide more effective immunotherapy strategies. However, the results based on retrospective analyses of public clinical reports were limited to confirming the role of PD-L1 in EBVaGC immunotherapy [68]. In our study, the PD-L1/2 was negatively correlated with ENTPD8 and positively correlated with PCOLCE2 (Figure 7C). However, whether ENTPD8 or PCOLCE2 regulates PD-L1 and the mechanism still need to be verified by prospective clinical trials.

While CD8+ T cell subsets are known to be associated with the mechanism of action of these immunotherapeutic agents, a comprehensive and diverse panel of markers providing comparable prognostic accuracy is desirable for clinical applications. The transfusion of macrophages could successfully activate and recruit CD8+ T cells in the spontaneous GC mouse model [69]. In the present study, the ENTPD8 and PCOLCE2 were also found to correlate with infiltration of
macrophage which was negatively correlated with prognosis (Figure 7D) and might be regarded as supplementary markers. In GC, the macrophage is a crucial component of the innate immune system. The ratio of pro-inflammatory to anti-inflammatory macrophages in GC is closely related to survival because the polarization of pro-inflammatory subtype would lead to the apoptosis of GC cells. [70]. In addition, macrophages could secrete CXCL8 and contribute to the immunosuppressive microenvironment by inducing PD-L1+ subtype[71]. In Chen et al. study, a CD68+STING+ macrophage subtype was identified to result in inferior OS with higher density, suggesting that STING or macrophages are negative prognosticators of GC [72]. Interestingly, the CD68+STING+ macrophage was also associated with a negative response to anti-PD-1/PD-L1 therapy. However, due to the limitations of experimental conditions, we have not confirmed the specific mechanism of ENTPD8 and PCOLCE2 regulating macrophage polarization through molecular experiments.

Clinical immunotherapy in EBVnGC still has limitations. Recently, the common solution was to use adjuvant conventional therapies (radiotherapy and chemotherapy) to improve the efficacy taking into account the ability to induce immunogenic cell death such as pembrolizumab (MK-3475) versus paclitaxel [73]. Similarly, the BLM could also induce programmed cell death depending on the infiltration of CD8+ T lymphocytes [14, 21]. After molecular docking, we identified top 9 binding sites of BLM in ENTPD8 and PCOLCE2. However, several limitations still need to be highlighted. First of all, the analyses for the impact of ENTPD8 and PCOLCE2 status on the immunotherapy efficacy were based on public data, which might exist certain biases. Due to the limited proportion of EBVaGC, the analysis would be a good supplement to the scarce clinical trial data about EBVaGC immunotherapy. Secondly, the number of samples for transcriptome analysis was relatively small, further validation was warranted for guiding in-depth research. Nevertheless, combining public data and tumor samples, our present study proved an important impact of bleomycin on immunotherapy in EBVaGC. Our findings provide significant views for future clinical research on immunotherapy combined with BLM in EBVaGC.

**Consent for publication**

All the authors were consent for publication.

**Availability of data and materials**

The supporting data can be acquired via correspondence author.
Competing interests
The authors declare that there are no competing interests associated with the manuscript.

Funding
This study was supported by Natural Science Foundation of Sichuan Province (2023NSFSC1552).

Author Contribution
L.C. conducted the data screening, statistical analysis, and writing of the manuscript. L.C. provided technical support. L.C. provided comments on the structure of the article. L.C. had read and provided final approval of the version to be submitted. The final manuscript was approved by all the authors above.

Corresponding author
Correspondence to Long Chen.

Acknowledgements
The authors are grateful to all the school health care staffs participating in this study.

Abbreviations
GC: gastric carcinoma; EMT: epithelial-mesenchymal transition; GSEA: gene set enrichment analysis; HR: hazard ratio; FDR: false discovery rate; IHC: immunohistochemistry; EBV: Epstein-Barr virus; HHV4: human herpesvirus 4; EBVaGC: EBV-associated GC; EBVnGCs: EBV-negative controls; ICB: immune checkpoint blockade; BLM: bleomycin; ROS: radical oxygen species; DCs: dendritic cells; Treg: regulatory T cells; CTLs: cytotoxic T lymphocytes; ATP: adenosine triphosphate; E-NTPDase: Ecto-nucleoside triphosphate diphosphohydrolase; BMP: bone morphogenic protein; PCPE1: procollagen C-proteinase enhancer1; TME: tumor microenvironment; GEO: gene expression omnibus; TCGA: the cancer genome atlas; DEGs: differentially expressed genes; 3D: three-dimensional; RPPA: reverse phase protein array; KEGG: kyoto encyclopedia of genes and genomes; GO: gene ontology; BP: biological process; CC: cellular component; MF: molecular function.

Reference
2. Murphy, G., et al., Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric


41. Steiglitz, B.M., D.R. Keene, and D.S. Greenspan, *PCOLCE2 encodes a functional procollagen


Supplementary Files

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

- supplementaryfigures.docx