A Novel Defined Risk Signature Based on Lactate-Related Genes can Predicts Prognosis of Bladder Cancer

Ding Hu

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Yanyang Jin

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Zizhi Li

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Weichao Huang

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Qingfei Cao

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Guangquan Tong

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Peng Hong

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Tong Shen

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Weikang Wu

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Ming Tong (✉ tongming@jzmu.edu.cn)

Jinzhou Medical University, The First Hospital of Jinzhou Medical University

Chundong Ji

Affiliated Hospital of Panzhihua University

Research Article

Keywords: TCGA, GEO, lactate, prognostic signature, bladder cancer, tumor microenvironment

Posted Date: April 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1387518/v2>

Abstract

Background

Lactate is an important carbon source for cell metabolism as well as a signaling molecule in normal, chronic inflammatory and cancer tissues. However, the relationship between lactate and bladder cancer (BLCA) has not been studied yet.

Methods

Expression profiles of the patients with bladder cancer from The Cancer Genome Atlas were divided into different clusters using non-negative matrix factorization clustering to obtain differentially expressed genes (DEGs). Univariate Cox analysis, Lasso-Cox analysis, Kaplan-Meier survival analysis, and time-dependent receiver operating characteristics curves were used to develop and validate a prognostic signature for the survival rate of patients. "CIBERSORT" and "ESTIMATE" algorithms were used to quantify the immunocyte content and evaluate tumor purity for different clusters and subgroups. Online tool "Tumor Immune Dysfunction and Exclusion" was used to predict the immune escape ability of patients in different scoring subgroups after immunotherapy.

Results

The different clusters according to the expression pattern of lactate-associated genes (LAGs) had significant differences in the survival rate and the tumor microenvironment. Besides, a risk signature constructed by DEGs based on different clusters and verified by the GSE32894 dataset could not only significantly distinguish between the prognosis and clinical features of patients with different scores but was also found to be related to the tumor microenvironment and immunotherapy. Finally, a high accuracy nomogram model was established.

Conclusions

Overall, this research has revealed the relationship between lactate and patient prognosis, tumor microenvironment, and immunotherapy, which may provide novel prognostic biomarkers for patients with BLCA.

Background

Bladder cancer (BLCA) is the tenth most common malignant tumor in the world. The United States Cancer Statistics reported 83,730 new cases and 17,200 bladder cancer-associated deaths in 2021 (1). Non-muscular invasive bladder cancer accounts for approximately 75% of the bladder cancer cases, while muscle-invasive bladder cancer accounts for 10-25% of the cases (2). BLCA is a highly heterogeneous

malignant tumor characterized by inherent biological invasiveness, radiological and chemotherapy resistance leading to high recurrence and progression (3). Although the incidence of BLCA has declined in recent years, the mortality rate has remained high (4). Recently, many gene signatures or biomarkers have been found to have the potential value to predict the prognosis of BLCA, such as autophagy (5), ferroptosis (6), and pyroptosis (7), but they are still in the stage of molecular research and have not yet been applied to clinical practice. Therefore, uncovering prognostic biomarkers for the prognosis of BLCA would be of important significance.

Lactate has long been considered a by-product of glycolysis. However, over the past three decades, scientists have begun to explore the new and unexpected biological functions of this molecule (8, 9). Lactate is now considered an important carbon source for cell metabolism, a signaling molecule in normal, chronic inflammatory and cancer tissues, and is closely related to the tumor microenvironment (TME) (10). TME is a favorable environment for tumor growth rather than anti-tumor immune monitoring, which is caused in part by accelerated metabolism of tumor cells and cancer-related fibroblasts (11). After the tissue has exhausted local resources, adjacent immune cells are forced to treat high concentrations of metabolites, especially lactate, in the absence of nutrition, which drives immunosuppression and facilitates tumor growth. For instance, LDHA-associated lactate accumulation in melanoma has been shown to inhibit tumor surveillance by NK cells and T cells (12). This inhibition is thought to be associated with lactate-mediated down-regulation of nuclear factors in activated T cells. Lactate produced by extracellular secretion in cancer cells plays a key role in improving cancer progression by creating an active environment that promotes tumor onset and development (13). Extracellular lactate levels can stimulate many cell types, including cancer cells (14), macrophages (15), dendritic cells (16), T cells (17), and NK cells (18), triggering intracellular signaling, fine-tuning cellular behavior, and strongly affecting their function in the TME. Although some researches have shown that lactate plays a significant role in tumor cell proliferation, invasion and metastasis, the value of lactate-associated genes (LAGs) in BLCA has not been explored to this date. To that end, we explored the expression levels of LAGs in BLCA patients, their clinical values, and related immune functions using bioinformatics to build a good prognostic predictive model for BLCA patients.

Methods And Materials

Data Collection

The workflow of this research is shown in **Figure 1**. The BLCA expression profile in the Cancer Genome Atlas (TCGA,<https://portal.gdc.cancer.gov>) database included 405 tumor samples and 19 normal samples. Corresponding data in the Gene Expression Omnibus (GEO,<https://www.ncbi.nlm.nih.gov/geo/>) database (GSE32894) included 219 tumor samples. Furthermore, copy number variation (CNV) data and mutation data were also obtained from the TCGA database. The use of the R package "SVA" (19) eliminated the batch effect of non-biotech deviations. 24 LAGs were obtained from GOBP_LACTATE_METABOLIC_PROCESS, GOBP_LACTATE_TRANSMEMBRANE_TRANSPORT, GOMF_LACTATE_DEHYDROGENASE_ACTIVITY, and

GOMF_LACTATE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY gene sets in the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/index.jsp>) (20).

Mutation and Difference Analyses of LAGs

The waterfall plot drawn by the R-package "maftools" (21) was used to represent the mutation landscape of the 24 LAGs, the frequency of CNV of each gene was displayed with a histogram, and the CNV change positions of the 24 LAGs on chromosomes were drawn by using R-package "RCircos". The R-package "limma" (22) was used to analyze the difference between cancer and para-cancer tissues of LAGs, and the differential LAGs were subjected to subsequent analysis, and the heat map was drawn with the R package "pheatmap". A protein-protein interaction (PPI) network for the LAGs was constructed with Search Tool for the Retrieval of Interacting Genes "STRING" (<https://string-db.org/>).

Subgroup Identification of BLCA Based on LAGs

After excluding cases lacking survival information and those with a survival period of fewer than 30 days, we retrieved the expression profiles and clinical information of BLCA patients in 393 cases from the TCGA database for further analysis. The BLCA samples were grouped using the R-package "NMF" and the non-negative matrix factorization (NMF) clustering algorithm (23). Based on the obtained expression patterns of differential LAGs, BLCA patients were divided into different clusters to identify their potential features. Clustering number k was set to 2–10, and the optimal clustering number was finally determined according to cophenetic indicator (the criterion for determining the optimal rank value is the pre-point of the maximum change in the cophenetic value with K).

Establishment of a Risk Signature by Lasso-Cox Method

The R package "limma" (22) was used to obtain the differently expressed genes (DEGs) ($|\log_2FC| > 1$, false discovery rate (FDR) < 0.05) among different clusters. And prognostic related DEGs were obtained by univariate Cox analysis. The dimensions of these DEGs were reduced by the Lasso-Cox method, and a risk signature (LAG-score) was constructed. TCGA dataset was used to construct a risk signature, and the GSE32894 dataset was used to verify the feasibility of that signature. The calculation formula is as follows:

$$LAG \text{ score} = \sum_{k=1}^n coef(k) * exp(k)$$

Among them, n , $coef(k)$ and $exp(k)$ represent the number of genes, corresponding coefficients and corresponding expressions in signature, respectively. Patients in the TCGA and GSE32894 cohorts were grouped into low- and high-score subgroups by the median LAG-score in the TCGA cohort. Kaplan-Meier (K-M) survival curves were used to evaluate the signature's ability to distinguish patients between

different subgroups, and time-dependent receiver operating characteristic (ROC) curves were used to determine the accuracy of signature.

Independent Prognostic Analysis

Univariate and multivariate Cox regression analyses using the R-package "survival" provided independent prognostic analysis of signature and other clinical features of BLCA(24). $P < 0.05$ was considered to have independent prognostic significance.

Gene Set Enrichment Analysis (GSEA)

The "c2.cp.kegg.v7.4.symbols" is a dataset containing the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway(25, 26). It was downloaded from the Molecular Signatures Database "MSigDB" (<https://www.gsea-msigdb.org/gsea/index.jsp>)(20) to identify differences in biological behavior and pathways in patients from different score subgroups in the TCGA cohort using the GSEA algorithm(27).

Tumor Microenvironment (TME) Analysis

Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) is a bioinformatics algorithm to calculate cell composition from gene expression profiles of complex tissues(28). Estimation of Stromal and Immune cells in Malignant Tumours using Expression data (ESTIMATE) is a tool that uses gene expression trends to infer the fraction of stromal and immune cells in tumor samples(29). We used them to quantify the immunocyte content and evaluate tumor purity for different clusters and subgroups. Tumor Immune Dysfunction and Exclusion (TIDE, <http://tide.dfci.harvard.edu>), as a computational framework that simulates two major mechanisms of tumor immune exclusion, can predict the efficacy of immunotherapy(30). We used it to assess the effectiveness of immunotherapy in patients with different LAG-score subgroups in the TCGA cohort.

Nomogram scoring system

We used R package "rms" to construct a nomogram model based on the LAG-score and other independent clinical features to improve the clinical value of the LAG-score(31). A score was matched for each variable in the nomogram scoring system, and the scores of all the variables of each sample were added, amounting to a total score from which the 1-, 3-, and 5-year probability of death of the corresponding patient could be calculated(32). Calibration curve is a method to evaluate the difference between the predicted value and the true value of the model(32, 33), the standard curve is a straight line passing through the origin of coordinate axes with a slope of 1. If the predicted calibration curve is closer to the standard curve, the prediction ability of nomograms is better. Decision curve analysis (DCA) is a method used to evaluate the diagnostic accuracy of a diagnostic model(34), the farther the curve away from the two reference lines, the model has more application value.

Statistical Analysis

All visual and statistical analyses were performed in this study using R version 4.1.0 using the corresponding functional packages. The Chi-square test was used to explore the differences of classified data among different datasets or groups. The Mann-Whitney U test was used for comparison between two groups. The K-M survival analysis was performed using the log rank test. The Spearman method was used to detect correlation. For all the results, $p < 0.05$ was considered statistically significant.

Results

Genetic Variation Landscape of LAGs in BLCA

We summarized the frequency of somatic mutations and CNV in 24 LAGs in BLCA. As shown in **Figure 2A**, at the genetic level, 222 of 412 (53.88%) samples showed LAGs mutations. Among them, TP53 showed the highest frequency of mutations. While LDHA and TIGAR did not find any mutation in any BLCA sample. Of these types of mutations, missense mutations were the most frequent. We also discovered that most of them had copy number deletion at the CNVs level (**Figure 2B**). At the same time, we identified the changes of regulators with CNV characteristics on chromosomes (**Figure 2C**).

Defining of the Expression of LAGs in BLCA

First, a PPI network for 24 LAGs was constructed with online tool "STRING" (**Figure S1A**) and the hub genes, SLC16A3, C12orf5, TP53, LDHA, and LDHB were identified using the "MCODE" plug-in in the "cytoscape" software (35) (**Figure S1B**). The correlation network was used to show the relationship between the 24 LAGs (**Figure 2D**) (the correlation coefficient > 0.2 , the red line indicates a positive correlation, and the blue line indicates a negative correlation). Next, the LAGs with differential expression between cancer and paracancerous tissues were identified by differential analysis. The multi-boxplot showed that 13 of the 24 LAGs had significantly dysregulated expression between cancer and paracancerous tissue, of which the expression levels of PER2, ACTN3, SLC16A7, LDHAL6B and SLC25A12 were down-regulated in cancer tissues, while the expression levels of PARK7, TP53, PNKD, PFKFB2, MRS2, LDHA, SLC16A8 and SLC16A3 were up-regulated in cancer tissues (**Figure 2E**).

The Subtypes Based on LAGs are closely Related to TME

A total of 13 LAGs with dysregulated expression between the cancer and paracancerous tissues were subjected to NMF analysis. The purpose of NMF analysis is to identify potential characteristics in gene expression profiles by decomposing the original matrix into two non-negative matrices (36). Cophenetic indicator was used to determine the optimal k value. Afterwards, the optimal total cluster number was set to $k = 2$ (the two subclasses were designated as cluster 1 ($n=181$) and cluster 2 ($n=212$)). When $k = 2$, the consensus matrix heat map maintained a clear and sharp boundary, indicating that the samples had stable and robust clusters (**Figure 3A-B**). The result of K-M survival curve indicated that the survival rate of cluster 2 was significantly higher than that of cluster 1 ($P=0.04$) (**Figure 3C**). Quantitative analysis of immune cells and estimation of the TME scores were performed on samples from different clusters using the "CIBERSORT"(28) and "ESTIMATE"(29) algorithms, respectively in **Figure 3D-E**. We can find that most

immune cells such as T regulatory cells, macrophages, and NK cells were significantly different between cluster 1 and cluster 2. In addition, the ESTIMATE score, immune score and stromal score of cluster 2 were lower than cluster 1, suggesting that cluster 2 had a higher tumor purity. The above results indicate that the subgroup based on LAGs was highly related to the TME. Moreover, we compared the immunophenotyping results of our study with previous studies to find a correlation. Sankey diagram showed that there were more C2 immune subgroups in cluster 1, while more C1, C3, and C4 immune subgroups were present in cluster 2 (**Figure 3F**).

A Risk Signature Based on Lactate-Related Clusters

To better apply these subtypes to the clinical management of BLCA and to identify a specific score for each patient, we next investigated the differences between these two clusters and identified a specific gene signature. Through analyzing the difference in gene expression among different clusters, we obtained a total of 1205 DEGs ($|\log_2 \text{FC}| > 1$, $\text{FDR} < 0.05$), of which the expression of 578 genes was down-regulated, while 627 genes were up-regulated in cluster 2 (**Supplementary Table 1**). 162 prognostic-related DEGs were identified by univariate Cox analysis ($p < 0.05$) (**Figure S2 and Supplementary Table 2**). Further dimension reduction of these regulators was performed by Lasso-Cox analysis, resulting in the generation of 15 genes and the construction of the risk signature named “LAG-score” (**Figure 4A and Table 1**). Among them, the coefficients of CTSE, PTPRR, GNLY, AIM2, ITGA3, TTTY14, SP6, C19orf18, ADRB2 and GPC2 were negative, indicating that these genes could be used as protection genes of patients, while the coefficients of ANXA1, SERPINE2, CDK6, SLC16A3 and PPBP were positive, indicating that these genes could be used as risk genes of patients. The LAG-score of each patient was obtained by multiplying the expression amount of each gene with its coefficient. The more negative the score the better the prognosis of patients while the more positive the score the worst the prognosis. All TCGA cohort’s patients were divided into high-scoring and low-scoring groups based on the median LAG-score (**Supplementary Table 3**). We reveal that as the score increased, the number of patients who died also gradually increased (**Figure 4B**). The result of K-M survival curve showed that the mortality rate of patients in the high LAG-score subgroup was significantly higher than that in the low LAG-score subgroup ($P < 0.001$) (**Figure 4C**). We performed this validation with the GSE32894 cohort from the GEO database (using the median LAG-score of the TCGA cohort to divide the GSE32894 cohort into high and low LAG-score subgroups), and the results showed that there were significant differences in the prognosis of patients between the two groups ($P = 0.005$) (**Figure 4D-E and Supplementary Table 4**). The results of ROC analysis in the TCGA cohort showed AUC values of 0.783, 0.823, and 0.811 at 1, 3, and 5 years, respectively (**Figure 5A**), while those of the GSE32894 cohort showed AUC values of 0.772, 0.709, and 0.705 at 1, 3, and 5 years, respectively (**Figure 5B**), thereby demonstrating the high accuracy of our model. In addition, the results of t-SNE analysis also demonstrated that there were significant differences in the distribution of patients based on LAG-score (**Figure 5C-D**).

Researchers have previously studied the prognosis model of BLCA under different modified conditions. For instance, Sun et al. found a seven-gene signature related to ferroptosis (6), Chen et al. found a three-gene signature related to m6A (37), Zhang et al. found an eight-gene signature related to hypoxia (38),

and Dong et al. found an eight-gene signature related to immunity (39) in BLCA patients. C-index, also known as consistency index, it and ROC curve are common methods to evaluate the model's accuracy (40, 41). Comparing the AUC values and C-indexes, we found that the AUC value and C-index of our model were 0.783 and 0.747, respectively, which were higher than the signatures of Chen (AUC = 0.636, C-index = 0.615), Dong (AUC = 0.655, C-index = 0.64), Sun (AUC = 0.654, C-index = 0.625), Zhang (AUC = 0.650, C-index = 0.6), indicated that our signature's accuracy was better than other prognostic signatures (Figure 5E-F).

Table 1. Results of multivariate Cox proportional hazard regression analysis of candidate genes in overall survival rate.

Gene	Coefficient
CTSE	-0.134
PTPRR	-0.187
ANXA1	0.160
SERPINE2	0.144
GNLY	-0.571
CDK6	0.172
SLC16A3	0.203
AIM2	-0.232
ITGA3	-0.136
TTY14	-0.530
SP6	-0.152
PPBP	0.255
C19orf18	-0.334
ADRB2	-0.548
GPC2	-0.494

Coefficient: Weights for each gene obtained by multivariate Cox analysis.

Clinical Prognostic Value of LAG-Score

To determine whether the prognostic value of LAG-score is independent of other clinical features such as sex, age, T stage and clinical stage, we performed univariate and multivariate Cox analyses on the TCGA cohort and the GSE32894 cohort, respectively, and found that LAG-score had independent prognostic significance in both cohorts, with HR values of 1.312 and 1.421, respectively (Figure 6A-D). The clinical

correlation heat map shows that LAG-score can distinguish patients with different age, gender, pathological grade, T stage, N stage and clinical stage significantly (**Figure 6E**). The above results indicated that the LAG-score may have a certain clinical value.

Different LAG-Score Subgroups Participated in Different Biological Pathways

To explore the biological pathways in the high- and low-scoring subgroups, we performed GSEA enrichment analysis, which showed that the high LAG-score subgroup was mainly enriched in carcinogenic-related pathways, such as extracellular matrix (ECM) receptor interaction, pathways in cancer, and focal adhesion. In comparison, the low LAG-score subgroup was mainly enriched in immune-related pathways, such as graft versus host disease and allograft rejection (**Figure 7A-B**).

Mutation Analysis Based on LAG-Score

To further explore lactate-mediated tumor heterogeneity, we performed mutation analysis for each of the high and low scoring subgroups and indicated that the high LAG-score subgroup had a slightly lower frequency of somatic mutations than the low LAG-score subgroup (92.31% vs. 95.43%) (**Figure 8A-B**). Moreover, in the high LAG-score subgroup, mutations in all genes except TP53 were less frequent than those in the low LAG-score subgroup (51% vs. 42%), which might be related to the LAG-score. Based on this phenomenon, we further explored the prognosis of both mutant and wild-type TP53 in the high and low score subgroups. Interestingly, the results of the K-M analysis showed that patients with TP53 mutation/high LAG-score subgroup had the worst prognosis, while patients with TP53 mutation/low LAG-score subgroup had the best prognosis ($P < 0.001$) (**Figure 8C**). It is now known that TP53 is involved in many biological processes such as apoptosis (42), immunity (43), and pyroptosis (44). We speculated that this might be related to the fact that the mutant TP53 mediates antitumor reactions such as immunity in the low LAG-score subgroup and mediates the cancer-causing pathway in the high LAG-score subgroup. In addition, we investigated the relationship between LAG-score and TMB, and the results showed that the TMB of samples with low LAG-score subgroup was higher than that of patients with high LAG-score subgroup (**Figure 8D**), and with the increase of LAG-score, TMB decreased gradually ($R = -0.18$, $P < 0.001$) (**Figure 8E**).

Immune Analysis Based on LAG-Score

Some researches have shown that lactate is closely related to the TME and immunotherapy (10). To investigate the relationship between LAG-score and TME, we assessed the immune cells content and TME scores using the "CIBERSORT" and "ESTIMATE" algorithms, respectively, and performed a spearman correlation analysis with the LAG-score. The results showed that more than half of the indicators negatively correlated with the LAG-score (**Figure 9A**). We compared the expression levels of immune checkpoints in the high and low LAG-score subgroups, and the results showed that PD-1, PD-L1, PD-L2, and CTLA4 were more elevated in the high LAG-score subgroup, which is consistent with the results of previous studies (**Figure 9B-E**). A high TIDE score means that patients receiving immunotherapy are more likely to achieve immune escape. Through comparison, we found that the TIDE score of the high LAG-

score subgroup was higher (**Figure 9F**), and the TIDE score increased with the increase of LAG-score, which meant that the samples in the high LAG-score subgroup were more likely to undergo immune escape while receiving immune treatment (**Figure 9G**). Moreover, the Sankey diagram showed that the majority of patients with high LAG-score are distributed in cluster 1 and the known C1 immune subgroup, while the majority of patients with low LAG-score are distributed in cluster 2 and the known C2 immune subgroup (**Figure 9H**).

Establishment of Nomogram and detecting its accuracy

To improve the clinical value of the LAG-score, we constructed a nomogram model that included age, T, stage, and LAG-score using the TCGA set. Calibration charts for 1-, 3-, and 5-year were used to test its predicted performance (**Figure 10A-B**). Consistently, the decision curve also showed that our nomogram had higher accuracy in predicting the prognosis of patients than other clinical features such as clinical stages (**Figure 10C**).

Discussion

BLCA is the tenth most common cancer in the world. It has high heterogeneity in histological and biological characteristics and a high rate of incidence and recurrence (1). In recent years, although the level of medical diagnosis and treatment has developed rapidly, the number of patients dying from BLCA has not decreased significantly (45, 46). Thus, it is necessary to formulate individualized treatment strategies to promote the prognosis of cancer patients. Lactate is the final product of glycolysis and an important carbon source for cell metabolism, serving as a signaling molecule in normal tissues, chronic inflammatory tissues, and cancer tissues (10). In the TME, accelerated metabolism of tumor cells and cancer-associated fibroblasts leads to depletion of local resources of the tissue, forcing neighboring immune cells to handle high concentrations of cellular metabolites, especially lactate, in the absence of nutrition, which drives immunosuppression and promotes tumor growth. Lactate dehydrogenase (LDH) exists in two different subtypes, LDHA and LDHB. LDHA is an enzyme that catalyzes the last step of the Voorburg effect by converting pyruvate and NADH to lactate, while LDHB converts lactate to pyruvate, which promotes oxidative metabolism. At present, although studies on lactate and bladder cancer are rarely reported, in other cancers such as prostate cancer (47), pancreatic cancer (48), and colorectal cancer (49), the down-regulation or loss of the expression level of lactate dehydrogenase LDHB represents a key early event in the development of cancer and is associated with high proliferation, increased tumor cell invasion ability, and poor prognosis results for patients. LDHA is the main cause of TME acidity and contributes to the epithelial-mesenchymal transition (EMT) and metastasis (50, 51). LDHA is up-regulated in the development of many tumors and is significantly associated with poor prognosis, and inhibition of LDHA is a promising method of targeting cancer (52, 53). In the mouse lung cancer model, targeting LDHA significantly inhibited tumorigenesis and tumor progression (54). In the breast cancer model, knocking down LDHA can affect mitochondrial physiology, inhibit glycolysis and significantly delay tumor development (55). These reports show that lactate is a promising research object in the process of tumor development and treatment. At present, more and more studies have

designed different signatures to predict the prognosis of tumor patients according to different biological modifications, such as ferroptosis (6), apoptosis (56), autophagy (57), and glycolysis(58). Although the current researches have shown that lactate is closely related to tumors and the TME, there is no report about designing a lactate-related signature to predict patients' prognosis and its relationship to TME. In this research, we constructed a LAG-score to predict the relationship between lactate and patients' prognosis with BLCA and explore its effects on TME and efficacy of immunotherapy. We first divided BLCA patients into two clusters according to the expression pattern of LAGs using the NMF method. According to our above findings, we can draw the following conclusions: First, LAGs can significantly differentiate between the prognosis of patients. The results of the K-M survival analysis suggested that the survival rate of cluster 2 was significantly higher than that of cluster 1. Second, different LAGs subtypes can reflect differences in TME, mainly differences in T cells, B cells, macrophages and NK cells. Many studies have shown that the efflux of proton-coupled lactate in tumor cells or stromal cells plays a significant role in maintaining the acidic phenotype of tumors, and it promotes tumor progression by regulating TME, including angiogenesis, cell invasion, metastasis development, survival signalling and escape from immune surveillance (59). Extracellular acidosis inhibits T cell-mediated immunity, and neutralizing tumor acidity improves antitumor response and immunotherapy (60). When the extracellular PH is between 6.0 and 6.5, it can induce the dysfunction of human tumor-specific CD8 + T cells, resulting in decreased cytolytic activity and cytokine production. At the same time, proton pump inhibitor therapy has been shown to be effective in restoring T cell function (61). In fact, acidic conditions reduced the expression of IL-6, CCL2, and NOS in M1 macrophages but increased the expression of M2 macrophage markers in the TME. Extracellular acidification also inhibits the antitumor activity of NK cells by inhibiting mTOR (62). Third, LAGs-based classification can be used to distinguish the known immunophenotype. Immune C2 was mainly concentrated in cluster 1 with a worse prognosis, while immune C1, C3, and C4 are distributed in cluster 2 with a better prognosis. This result indicates that our method has certain value for immunophenotyping. At present, more and more studies have proved the feasibility of prognostic signatures. Therefore, as this study progressed, we established LAG-score signature. According to the research results, we can reach the following consensus: First, the LAG-score can be used to identify heterogeneity in patients. Second, this study reveals that lactate is closely related to the clinical outcome of patients. Notably, the LAG-score signature includes both tumor-promoting genes and tumor-inhibiting genes, and assigns different weights to them. In addition, signature includes but is not limited to LAGs, such as SLC16A3. SLC16A3 is monocarboxylate transporter 4 (MCT4) of the SLC16 solute carrier family, plays a significant role in cancer metabolism (63). It is strongly overexpressed in anaerobic tissues and highly glycolytic, has a low affinity for lactate, responsible for lactate export (64). In tumor tissue, the acidification of the extracellular microenvironment resulting from co-transported protons is beneficial for angiogenesis, tumor cell proliferation, invasion, and metastasis (65). Since L-lactate exported by SLC16A3 is used as a fuel for proliferating cancer cells (66), inhibition of SLC16A3 is a hopeful strategy for treating certain tumor types. Third, LAG-score can not only significantly distinguish the prognosis and different clinical features of patients, but also closely relate to the presence of somatic mutation. The low LAG-score subgroup has longer predicted survival time and favorable clinical characteristics. In the mutation analysis, TMB was higher in low LAG-score subgroup and was negatively correlated with the

LAG-score. TP53 which is a participant of various biological modifications exhibited different mutation frequencies in the high and low LAG-score subgroups affecting their prognosis differently. The possible mechanism could be that it mediates anti-tumor reactions such as immunity in the low LAG-score subgroup, while it promotes tumor development in the high LAG-score subgroup. Fourth, different LAG-score subgroups participated in different biological pathways. The high LAG-score subgroup is mainly involved in carcinogenesis pathway, especially ECM-receptor interaction. ECM is a component of the TME, and ECM receptor interaction actively participates in tumor progression (67). Studies have shown a high correlation between ECM and glycolysis (68). Lactate is a product of glycolysis, however, cancer cells still preferentially metabolize glucose to lactate under aerobic conditions (69). The accumulation of lactate acidifies the surrounding environment of cells, further promoting tumor growth. Based on the relevant literature and the results of GSEA in this study, it can be concluded that there may be a direct or indirect relationship between ECM and lactate metabolism. To inhibit the production of lactate in TME and thus inhibit the growth of tumors, ECM may be the starting point. Fifth, LAG-score can distinguish the degree of immune cells' infiltration in the TME and is related to the effectiveness of immunotherapy. Based on the infiltration results of TME cells and analysis by TIDE, it was shown that more immune cells were activated in patients with low LAG-score, showing better response to immunotherapy. Although immune checkpoint analysis revealed a higher expression of PD-1, PD-L1, PD-L2, and CTLA4 in patients with high LAG-score, predictive data from TIDE suggest that patients with high LAG-score were more susceptible to immune evasion during immunotherapy. This may be due to increased glucose uptake and lactate secretion promoting immune escape through immune checkpoint ligand and metabolic competition (70). Moreover, relevant studies have shown that acidified TME can promote the up-regulation of some immune checkpoints and limit the efficacy of immunotherapy, however, neutralizing the acidic environment can improve the efficacy of immunotherapy for patients (61, 71). Finally, our model focuses on the prognosis of BLCA, TME, and effectiveness of immunotherapy compared with other models. We found that our model was more valuable in facilitating individualized treatment showing better AUC value and C-index compared to previous BLCA signatures that were studied under the modified conditions of ferroptosis (6), m6A (37), hypoxia (38), and immunity (39).

Our study started with LAGs by categorizing patients with BLCA into subtypes, identifying DEGs, establishing a risk signature, and linking lactate to prognosis. Although we have researched from multiple perspectives, multiple methods and multiple sets of data, there are still some limitations. First, although our risk signature is established based on lactate related clusters, we cannot prove whether the factors in signature play corresponding roles in the lactate metabolism-related pathways, so future in vitro and in vivo experiments are necessary. Second, although our study could predict the effectiveness of immunotherapy in patients, its real value needs to be further confirmed in clinical trials.

Conclusion

In conclusion, the results of this research show that lactate is closely related to BLCA. This research provides a new risk signature for predicting the prognosis of BLCA patients and an important basis for further study on the relationship between LAGs and immunotherapy effectiveness in BLCA.

Abbreviations

BLCA
bladder cancer
TME
tumor microenvironment
TCGA
The Cancer Genome Atlas
NMF
non-negative matrix factorization
DEGs
differentially expressed genes
ROC
receiver operating characteristics
LAGs
lactate-associated genes
CNV
copy number variation
K-M
Kaplan-Meier
GSEA
Gene Set Enrichment Analysis
ECM
extracellular matrix
TMB
tumor mutation burden
DCA
decision curve analysis
LDH
lactate dehydrogenase
EMT
epithelial-mesenchymal transition

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Publicly available datasets were analyzed in this study. This data can be found here: <https://portal.gdc.cancer.gov/>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

DH, MT and CJ designed the research framework and revised the manuscript. DH and YJ contributed to the data analysis and wrote the manuscript. ZL, WH, QC, GT, WW, PH and TS provided comments during the writing. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge TCGA (<https://portal.gdc.cancer.gov/>) database and GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database for providing their platforms and contributors for uploading their meaningful datasets.

References

1. R. Siegel, K. Miller, H. Fuchs, A. Jemal. (2021). Cancer Statistics, 2021. *CA: a cancer journal for clinicians*. **71**. 7-33. doi:10.3322/caac.21654.
2. O. Sanli *et al.* (2017). Bladder cancer. *Nature reviews. Disease primers*. **3**. 17022. doi:10.1038/nrdp.2017.22.
3. F. Massari *et al.* (2016). Metabolic phenotype of bladder cancer. *Cancer treatment reviews*. **45**. 46-57. doi:10.1016/j.ctrv.2016.03.005.
4. M. Al-Husseini *et al.* (2019). Trends in the incidence and mortality of transitional cell carcinoma of the bladder for the last four decades in the USA: a SEER-based analysis. *BMC cancer*. **19**. 46. doi:10.1186/s12885-019-5267-3.
5. X. Yan *et al.* (2021). Construction and Validation of an Autophagy-Related Prognostic Signature and a Nomogram for Bladder Cancer. *Frontiers in oncology*. **11**. 632387. doi:10.3389/fonc.2021.632387.
6. J. Sun *et al.* (2021). Identification of a Novel Ferroptosis-Related Gene Prognostic Signature in Bladder Cancer. *Frontiers in oncology*. **11**. 730716. doi:10.3389/fonc.2021.730716.

7. J. Fu, Y. Wang. (2021). Identification of a novel pyroptosis-related gene signature for predicting prognosis in bladder cancer. *Cancer investigation*. 1-24. doi:10.1080/07357907.2021.1991944.
8. S. Bensinger, H. Christofk. (2012). New aspects of the Warburg effect in cancer cell biology. *Seminars in cell & developmental biology*. **23**. 352-361. doi:10.1016/j.semcdb.2012.02.003.
9. J. Doherty, J. Cleveland. (2013). Targeting lactate metabolism for cancer therapeutics. *The Journal of clinical investigation*. **123**. 3685-3692. doi:10.1172/jci69741.
10. M. Certo, C. Tsai, V. Pucino, P. Ho, C. Mauro. (2021). Lactate modulation of immune responses in inflammatory versus tumour microenvironments. *Nature reviews. Immunology*. **21**. 151-161. doi:10.1038/s41577-020-0406-2.
11. M. Bissell, W. Hines. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nature medicine*. **17**. 320-329. doi:10.1038/nm.2328.
12. A. Brandt *et al.* (2016). LDHA-Associated Lactic Acid Production Blunts Tumor Immunosurveillance by T and NK Cells. *Cell metabolism*. **24**. 657-671. doi:10.1016/j.cmet.2016.08.011.
13. O. Colegio *et al.* (2014). Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature*. **513**. 559-563. doi:10.1038/nature13490.
14. K. Goetze, S. Walenta, M. Ksiazkiewicz, L. Kunz-Schughart, W. Mueller-Klieser. (2011). Lactate enhances motility of tumor cells and inhibits monocyte migration and cytokine release. *International journal of oncology*. **39**. 453-463. doi:10.3892/ijo.2011.1055.
15. P. Chen *et al.* (2017). Gpr132 sensing of lactate mediates tumor-macrophage interplay to promote breast cancer metastasis. *Proceedings of the National Academy of Sciences of the United States of America*. **114**. 580-585. doi:10.1073/pnas.1614035114.
16. P. Ranganathan *et al.* (2018). GPR81, a Cell-Surface Receptor for Lactate, Regulates Intestinal Homeostasis and Protects Mice from Experimental Colitis. *Journal of immunology (Baltimore, Md. : 1950)*. **200**. 1781-1789. doi:10.4049/jimmunol.1700604.
17. R. Haaset *et al.* (2015). Lactate Regulates Metabolic and Pro-inflammatory Circuits in Control of T Cell Migration and Effector Functions. *PLoS biology*. **13**. e1002202. doi:10.1371/journal.pbio.1002202.
18. Z. Husain, P. Seth, V. Sukhatme. (2013). Tumor-derived lactate and myeloid-derived suppressor cells: Linking metabolism to cancer immunology. *Oncoimmunology*. **2**. e26383. doi:10.4161/onci.26383.
19. J. Leek, W. Johnson, H. Parker, A. Jaffe, J. Storey. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)*. **28**. 882-883. doi:10.1093/bioinformatics/bts034.
20. A. Liberzon *et al.* (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. **1**. 417-425. doi:10.1016/j.cels.2015.12.004.
21. A. Mayakonda, D. Lin, Y. Assenov, C. Plass, H. Koeffler. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. **28**. 1747-1756. doi:10.1101/gr.239244.118.

22. M. Ritchie *et al.* (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. **43**. e47. doi:10.1093/nar/gkv007.
23. V. Gligorijevic, Y. Panagakis, S. Zafeiriou. (2019). Non-Negative Matrix Factorizations for Multiplex Network Analysis. *IEEE transactions on pattern analysis and machine intelligence*. **41**. 928-940. doi:10.1109/tpami.2018.2821146.
24. W. Shao *et al.* (2021). The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. *Frontiers in cell and developmental biology*. **9**. 676485. doi:10.3389/fcell.2021.676485.
25. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. **44**. D457-462. doi:10.1093/nar/gkv1070.
26. M. Kanehisa, S. Goto. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. **28**. 27-30. doi:10.1093/nar/28.1.27.
27. A. Subramanian *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. **102**. 15545-15550. doi:10.1073/pnas.0506580102.
28. A. Newman *et al.* (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. **12**. 453-457. doi:10.1038/nmeth.3337.
29. K. Yoshihara *et al.* (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. **4**. 2612. doi:10.1038/ncomms3612.
30. Y. Chen, Z. Li, G. Zhou, Y. Sun. (2021). An Immune-Related Gene Prognostic Index for Head and Neck Squamous Cell Carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. **27**. 330-341. doi:10.1158/1078-0432.Ccr-20-2166.
31. Z. Zhang, M. Kattan. (2017). Drawing Nomograms with R: applications to categorical outcome and survival data. *Annals of translational medicine*. **5**. 211. doi:10.21037/atm.2017.04.01.
32. A. Iasonos, D. Schrag, G. Raj, K. Panageas. (2008). How to build and interpret a nomogram for cancer prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. **26**. 1364-1370. doi:10.1200/jco.2007.12.9791.
33. V. Balachandran, M. Gonen, J. Smith, R. DeMatteo. (2015). Nomograms in oncology: more than meets the eye. *The Lancet. Oncology*. **16**. e173-180. doi:10.1016/s1470-2045(14)71116-7.
34. A. Vickers, E. Elkin. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making*. **26**. 565-574. doi:10.1177/0272989x06295361.
35. G. Bader, C. Hogue. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. **4**. 2. doi:10.1186/1471-2105-4-2.
36. J. Brunet, P. Tamayo, T. Golub, J. Mesirov. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*. **101**. 4164-4169. doi:10.1073/pnas.0308531101.

37. M. Chen *et al.* (2019). m6A RNA methylation regulators can contribute to malignant progression and impact the prognosis of bladder cancer. *Bioscience reports*. **39**. doi:10.1042/bsr20192892.
38. F. Zhan *et al.* (2021). Development and Validation of a Hypoxia-Related Signature for Predicting Survival Outcomes in Patients With Bladder Cancer. *Frontiers in genetics*. **12**. 670384. doi:10.3389/fgene.2021.670384.
39. B. Dong *et al.* (2021). Tumor Expression Profile Analysis Developed and Validated a Prognostic Model Based on Immune-Related Genes in Bladder Cancer. *Frontiers in genetics*. **12**. 696912. doi:10.3389/fgene.2021.696912.
40. F. Harrell, K. Lee, D. Mark. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. **15**. 361-387. doi:10.1002/(sici)1097-0258(19960229)15:4<361::Aid-sim168>3.0.Co;2-4.
41. P. Blanche, J. Dartigues, H. Jacqmin-Gadda. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in medicine*. **32**. 5381-5397. doi:10.1002/sim.5958.
42. H. Wei *et al.* (2021). Structural insight into the molecular mechanism of p53-mediated mitochondrial apoptosis. *Nature communications*. **12**. 2280. doi:10.1038/s41467-021-22655-6.
43. S. Hamet *et al.* (2019). TP53 gain-of-function mutation promotes inflammation in glioblastoma. *Cell death and differentiation*. **26**. 409-425. doi:10.1038/s41418-018-0126-3.
44. L. Tan *et al.* (2021). TP53-induced glycolysis and apoptosis regulator alleviates hypoxia/ischemia-induced microglial pyroptosis and ischemic brain damage. *Neural regeneration research*. **16**. 1037-1043. doi:10.4103/1673-5374.300453.
45. A. Necchi *et al.* (2018). Pembrolizumab as Neoadjuvant Therapy Before Radical Cystectomy in Patients With Muscle-Invasive Urothelial Bladder Carcinoma (PURE-01): An Open-Label, Single-Arm, Phase II Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. **36**. 3353-3360. doi:10.1200/jco.18.01148.
46. M. Eckstein, D. Sikic, P. Strissel, F. Erlmeier. (2018). Evolution of PD-1 and PD-L1 Gene and Protein Expression in Primary Tumors and Corresponding Liver Metastases of Metastatic Bladder Cancer. *European urology*. **74**. 527-529. doi:10.1016/j.eururo.2018.06.028.
47. A. Leiblich *et al.* (2006). Lactate dehydrogenase-B is silenced by promoter hypermethylation in human prostate cancer. *Oncogene*. **25**. 2953-2960. doi:10.1038/sj.onc.1209262.
48. J. Cui *et al.* (2015). Suppressed expression of LDHB promotes pancreatic cancer progression via inducing glycolytic phenotype. *Medical oncology (Northwood, London, England)*. **32**. 143. doi:10.1007/s12032-015-0589-8.
49. L. Shi *et al.* (2019). SIRT5-mediated deacetylation of LDHB promotes autophagy and tumorigenesis in colorectal cancer. *Molecular oncology*. **13**. 358-375. doi:10.1002/1878-0261.12408.
50. F. Hirschhaeuser, U. Sattler, W. Mueller-Klieser. (2011). Lactate: a metabolic key player in cancer. *Cancer research*. **71**. 6921-6925. doi:10.1158/0008-5472.Can-11-1457.

51. M. Vander Heiden, L. Cantley, C. Thompson. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N.Y.)*. **324**. 1029-1033.
doi:10.1126/science.1160809.
52. C. Dai *et al.* (2020). Lactate Dehydrogenase A Governs Cardiac Hypertrophic Growth in Response to Hemodynamic Stress. *Cell reports*. **32**. 108087. doi:10.1016/j.celrep.2020.108087.
53. C. Yuet *et al.* (2018). LDHA upregulation independently predicts poor survival in lung adenocarcinoma, but not in lung squamous cell carcinoma. *Future oncology (London, England)*. **14**. 2483-2492.
doi:10.2217/fon-2018-0177.
54. H. Xie *et al.* (2014). Targeting lactate dehydrogenase—a inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell metabolism*. **19**. 795-809.
doi:10.1016/j.cmet.2014.03.003.
55. Y. Zhao *et al.* (2009). Upregulation of lactate dehydrogenase A by ErbB2 through heat shock factor 1 promotes breast cancer cell glycolysis and growth. *Oncogene*. **28**. 3689-3701.
doi:10.1038/onc.2009.229.
56. J. Zhu *et al.* (2020). Identifying Apoptosis-Related Transcriptomic Aberrations and Revealing Clinical Relevance as Diagnostic and Prognostic Biomarker in Hepatocellular Carcinoma. *Frontiers in oncology*. **10**. 519180. doi:10.3389/fonc.2020.519180.
57. W. Kang, J. Hu, Q. Zhao, F. Song. (2021). Identification of an Autophagy-Related Risk Signature Correlates With Immunophenotype and Predicts Immune Checkpoint Blockade Efficacy of Neuroblastoma. *Frontiers in cell and developmental biology*. **9**. 731380.
doi:10.3389/fcell.2021.731380.
58. J. Bi, F. Bi, X. Pan, Q. Yang. (2021). Establishment of a novel glycolysis-related prognostic gene signature for ovarian cancer and its relationships with immune infiltration of the tumor microenvironment. *Journal of translational medicine*. **19**. 382. doi:10.1186/s12967-021-03057-0.
59. L. Ippolito, A. Morandi, E. Giannoni, P. Chiarugi. (2019). Lactate: A Metabolic Driver in the Tumour Landscape. *Trends in biochemical sciences*. **44**. 153-166. doi:10.1016/j.tibs.2018.10.011.
60. G. Sgarbi, G. Gorini, F. Liuzzi, G. Solaini, A. Baracca. (2018). Hypoxia and IFN- γ Expression Promote ROS Decrease in Cancer Cells. *Cells*. **7**. doi:10.3390/cells7070064.
61. A. Balgi *et al.* (2011). Regulation of mTORC1 signaling by pH. *PloS one*. **6**. e21549.
doi:10.1371/journal.pone.0021549.
62. D. Xie, S. Zhu, L. Bai. (2016). Lactic acid in tumor microenvironments causes dysfunction of NKT cells by interfering with mTOR signaling. *Science China. Life sciences*. **59**. 1290-1296.
doi:10.1007/s11427-016-0348-7.
63. P. Bosshart, D. Kalbermatter, S. Bonetti, D. Fotiadis. (2019). Mechanistic basis of L-lactate transport in the SLC16 solute carrier family. *Nature communications*. **10**. 2649. doi:10.1038/s41467-019-10566-6.
64. N. Price, V. Jackson, A. Halestrap. (1998). Cloning and sequencing of four new mammalian monocarboxylate transporter (MCT) homologues confirms the existence of a transporter family with

- an ancient past. *The Biochemical journal*. 321-328. doi:10.1042/bj3290321.
65. V. Ganapathy, M. Thangaraju, P. Prasad. (2009). Nutrient transporters in cancer: relevance to Warburg hypothesis and beyond. *Pharmacology & therapeutics*. **121**. 29-40. doi:10.1016/j.pharmthera.2008.09.005.
66. P. Sonveaux *et al.* (2008). Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *The Journal of clinical investigation*. **118**. 3930-3942. doi:10.1172/jci36843.
67. M. van Dijk, S. Göransson, S. Strömblad. (2013). Cell to extracellular matrix interactions and their reciprocal nature in cancer. *Experimental cell research*. **319**. 1663-1670. doi:10.1016/j.yexcr.2013.02.006.
68. W. Sullivan *et al.* (2018). Extracellular Matrix Remodeling Regulates Glucose Metabolism through TXNIP Destabilization. *Cell*. **175**. 117-132.e121. doi:10.1016/j.cell.2018.08.017.
69. K. Garber. (2006). Energy deregulation: licensing tumors to grow. *Science (New York, N.Y.)*. **312**. 1158-1159. doi:10.1126/science.312.5777.1158.
70. C. Changet *et al.* (2015). Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell*. **162**. 1229-1241. doi:10.1016/j.cell.2015.08.016.
71. J. Zhang, C. Wolfgang, L. Zheng. (2018). Precision Immuno-Oncology: Prospects of Individualized Immunotherapy for Pancreatic Cancer. *Cancers*. **10**. doi:10.3390/cancers10020039.

Figures

Figure 1

The research workflow

Figure 2

Characterization of lactate-associated genes (LAGs) in tumors at the biological level. **(A)** The copy number variation (CNV) frequency of 24 LAGs, more than half of which have copy number deletion. **(B)** The position of LAGs at which CNV occurs on the chromosome. **(C)** Genomic changes from 412 BLCA samples from the TCGA with the waterfall plot representing mutation information for different LAGs. **(D)** Correlation between 24 LAGs, where the blue line indicates a negative correlation, and the red line

indicates a positive correlation. (E) The expression levels of 13 out of 24 LAGs were dysregulated between cancer and paracancerous tissue (Mann-Whitney U test, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

Figure 3 Subgroup of BLCA patients and their tumor microenvironment based on 13 LAGs. (A) Non-negative matrix factorization (NMF) Consensus clustering matrix for $k = 2-10$, (B) and the cophenetic index shows

that the best grouping is $k = 2$. **(C)** The K-M survival analysis showed that the survival rate of cluster 1 was significantly lower than cluster 2 (log-rank test, $p=0.04$). **(D)** The degree of immune cells' infiltration in different clusters based on the "CIBERSORT" algorithm (Mann-Whitney U test, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). **(E)** Tumor microenvironment scoring in different clusters based on the "ESTIMATE" algorithm (Mann-Whitney U test, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). **(F)** Distribution of BLCA subtypes based on LAGs in known immunophenotypes.

Figure 4

Construction of LAG-score signature by Lasso-Cox regression analysis. **(A)** The distribution of the partial likelihood deviation of the Lasso coefficient preserved 33 variables when the partial likelihood deviation reached the minimum (Log Lambda = -3.35). Survival risk profile **(B)** and K-M survival curve (log-rank test, $P < 0.001$) **(C)** of patients based on the LAG-score in the TCGA cohort. Survival risk profile **(D)** and K-M survival curve **(E)** (log-rank test, $P=0.005$) of patients based on the LAG-score in the GSE32894 cohort.

Figure 5

Validation of the accuracy of LAG-score signature and model comparison. ROC curve in TCGA cohort (A) and GSE32894 cohort (B) at 1, 3, and 5 years. The t-SNE analysis demonstrated that LAG-score could well group the TCGA cohort (C) and the GSE32894 cohort (D) into two subgroups. The ROC curve (E) and C-index (F) of the LAG-score signature and the other four prognostic model.

Figure 7

Different scoring subgroups participated in different biological pathways. GSEA enrichment analysis for high LAG-score subgroup (A) and low LAG-score subgroup (B).

Figure 8

Mutation analysis based on LAG-score. Top 20 gene mutation frequencies in the high LAG-score subgroup (A) and the low LAG-score subgroup (B). (C) K-M survival curves for patients by TP53 mutation and LAG-score subgroups (log-rank test, $P < 0.001$). Tumor mutation burden (D) difference in high and low LAG-score subgroups and the correlation with the LAG-score (E).

Figure 9

Tumor microenvironment, immune checkpoints and immunotherapy based on LAG-score. (A) The lollipop plot shows the correlation between LAG-score and each component in TME (Spearman, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). Differences in expression levels of PD-1 (B), PD-L1 (C), PD-L2 (D), CTLA4 (E) and TIDE scores (F) between the high and low LAG-score subgroups (Mann-Whitney U test, $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). (G) Correlation between TIDE score and LAG-score. (H) The distribution of BLCA subgroups based on LAG-score in the aforementioned LAGs- based BLCA subtypes and known immunophenotypes.

Figure 10

Nomogram scoring system. Nomogram **(A)** combined with age, T stage, and LAG-score and calibration diagrams of 1-, 3-, and 5-year **(B)**. **(C)** The decision curve further proves the accuracy of the nomogram.

Supplementary Files

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

- [figS1.pdf](#)
- [FigS2.pdf](#)
- [SupplementaryFigureLegends.docx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTableLegends.docx](#)