Background: Bladder cancer (BLCA) is the fifth most common type of cancer worldwide, with high recurrence and progression rates. Although considerable progress has been made in the treatment of BLCA through accurate typing of molecular characteristics, little is known regarding the various genetic and epigenetic changes that have evolved in stem and progenitor cells. To address this issue, we have developed a novel stem cell typing method.

Methods: Based on six published genomic datasets, we used 26 stem cell gene sets to classify each dataset. Unsupervised and supervised machine learning methods were used to perform the classification.

Results: We classified BLCA into three subtypes—high stem cell enrichment (SCE_H), medium stem cell enrichment (SCE_M), and low stem cell enrichment (SCE_L)—based on multiple cross-platform datasets. The stability and reliability of the classification were verified. Compared with the other subtypes, SCE_H had the highest degree of cancer stem cell concentration, highest level of immune cell infiltration, and highest sensitivity not only to predicted anti-PD-1 immunosuppressive therapy but also to conventional chemotherapeutic agents such as cisplatin, sunitinib, and vinblastine; however, this group had the worst prognosis. Comparison of gene set enrichment analysis results for pathway enrichment of various subtypes reveals that the SCE_H subtype activates the important pathways regulating cancer occurrence, development, and even poor prognosis, including epithelial-mesenchymal transition, hypoxia, angiogenesis, KRAS signal upregulation, interleukin 6-mediated JAK-STAT signaling pathway, and inflammatory response. Two identified pairs of transcription factors, GRHL2 and GATA6 and IRF5 and GATA3, possibly have opposite regulatory effects on SCE_H and SCE_L, respectively.

Conclusions: The identification of BLCA subtypes based on cancer stem cell gene sets revealed the complex mechanism of carcinogenesis of BLCA and provides a new direction for the diagnosis and treatment of BLCA.

Keywords: Bladder cancer, Stem cells, Immune microenvironment, Hypoxia, Epithelial-mesenchymal transition, Immunotherapy
**Background**

Bladder cancer (BLCA), generally occurs in bladder intraepithelial cells, is the fifth most common type of cancer worldwide. Approximately 151,000 new cases of BLCA and more than 52,000 related deaths worldwide are reported annually [1–3]. Urothelial cancer is the most common type of BLCA, accounting for approximately 90% of all BLCA cases [1]. Most BLCA cases can be diagnosed at an early stage, but the rate of recurrence and progression remains high, approximately 78% of patients relapse within 5 years [4]. Using various biological detection technology, molecular typing of BLCA through genetic analysis has shown differences in drug reactivity and prognosis in patients with BLCA based on their biological heterogeneity, for example, molecular classification of the Cancer Genome Atlas Quartile [5], University of North Carolina Dichotomy [6], MD Anderson Cancer Center Trisection [7], and Lund University Quintiles [8]. Although these classification methods reveal the pathogenic mechanism of BLCA at the molecular level, they do not fundamentally demonstrate the origin of heterogeneity in tumors. New evidence suggests that cancer stem cell (CSC) subpopulations are characterized by a mixture of stem cells and cancer cells. In addition to the abilities of self-renewal and differentiation, CSCs can also act as tumors' seeds [9, 10] and are thus considered as the driving force of heterogeneity.

Tumors are complex integrated systems composed of relatively differentiated tumor cells, infiltrating immune cells, CSCs, tumor-associated endothelial cells, stromal cells, and other cell types [11, 12]. The function and plasticity of CSCs are induced by specific signals and cell interactions in the tumor niche. Studies have shown that stem cells in melanoma can preferentially inhibit T cell activation and influence the induction of regulatory T cells, thereby evading recognition by the immune system [13]. In glioblastoma, CSCs suppress T cell responses by generating immunosuppressive cytokines through the STAT3 pathway and inducing T cell apoptosis, leading to an increase in cancer stemness and carcinogenic potential [14]. Additionally, some molecular signal transduction pathways, which control stem cell balance, are abnormally activated or inhibited to contribute to the self-renewal, proliferation, survival, and differentiation characteristics of CSCs. In a study using an experimental model of colon cancer, elevated inflammatory nuclear factor κB signal transduction enhanced Wnt activation and induced dedifferentiation of non-stem cells, which acquired tumor-initiating ability [15]. These results indicate that immune cells and their related cytokines and signal transduction pathways can directly regulate and enhance the CSC phenotype.

CD44 is a CSC surface marker [16], including BLCA stem cells [17], and its overexpression is positively correlated with BLCA tumor aggressiveness. IL-6 can regulate CD44 which is essential for the maintenance of normal stem cells. In addition, abnormal activation of the JAK-STAT signaling pathway induces tumors, while the IL6/JAK/STAT3 signaling pathway helps to maintain the plasticity of breast CSCs. Meanwhile, upon its activation, the mTORC1-STAT3 signaling pathway also helps to maintain the stemness of BLCA stem cells [18, 19]. MYC influences somatic cell reprogramming and controls embryonic stem cell self-renewal. Following MYC inactivation, tumors undergo various proliferative arrests, cell differentiation, and apoptosis, thus inhibiting tumor occurrence. In liver tumor cells, Shachaf et al. have demonstrated that MYC inactivation triggers stem cell differentiation, while its reactivation can restore their tumor characteristics. Therefore, although the inactivation of oncogenes restores normal cells, some cells retain their potential of becoming cancerous, possibly existing in a tumor dormant state [20]. GATA3 influences the maintenance of BLCA stem cells. Yang et al. [21] were the first to show that the KMT1A-GATA3-STAT3 signaling pathway promotes BLCA stem cell self-renewal. KMT1A protein directly catalyzes the tri-methylation modification (H3K9me3) of the 9th lysine of histone H3 in the promoter region of the GATA3 gene (−1351 to approximately −1172), thereby inhibiting its transcription. The GATA3 protein can directly bind to the promoter region (−1710 to approximately −1530) of the STAT3 gene inhibiting its transcription. Therefore, the transcriptional repression of the GATA3 gene, mediated by histone methyltransferase KMT1A, promotes the upregulation of STAT3 expression and activation, ultimately achieving the maintenance of BLCA stem cells.

In our study, we divided BLCA into high stem cell enrichment (SCE_H), medium stem cell enrichment (SCE_M), and low stem cell enrichment (SCE_L) subtypes, using stem cell gene set collection. We have demonstrated the stability and reliability of this classification with six independent datasets using the unsupervised clustering method. Importantly, we systematically examined the prognostic significance of BLCA stem cell subtypes, relationship between immune cells and genes, sensitivity of immune checkpoint inhibitor treatment, and possible changes in the biological pathways and important transcriptional regulation factors/networks (Fig. 1). Our study provides an insight into and a basis of BLCA stem cells and helps to improve clinical diagnosis and treatment of BLCA.

**Methods**

**Stem cell signature collection**

The 26 human stem cell gene sets used in this study were obtained from StemChecker (http://stemchecker.
expression checks (18), RNAi screens (1), literal curation (2), computationally derived (2), and TF target genes (3).

**Data processing**

The datasets used to identify the BLCA stem cell subtypes were from three different platforms: The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and ArrayExpress databases. TCGA’s RNA-seq data (fragments per kilobase of transcript per million mapped reads (FPKM)) of 19 normal samples and 414 cancer samples, variant data of VarScan, and clinical information were downloaded from TCGA Knowledge Base (https://portal.gdc.cancer.gov/repository). Gene annotation was performed using the Ensemble database. The ArrayExpress database contains RNA-seq and clinical data \((n = 476)\) for 476 cases of early urothelial carcinoma (E-MTAB-4321) FPKM from the European Genome-phenome Archive. The expression matrices of four GEO datasets, GSE13507 \((n = 165)\), GSE32548 \((n = 131)\), GSE31684 \((n = 93)\), and GSE32894 \((n = 308)\), were all quantile-normalized, and the genes were annotated in their respective platform files Illumina human -6 v2.0 expression beadchip, Illumina HumanHT-12 V3.0 expression beadchip, [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, and Illumina HumanHT-12 V3.0 expression beadchip.

**Identification of BLCA subtypes based on stem cell gene sets**

For each BLCA dataset, we used the GSVA package to perform a single-sample gene set enrichment analysis (ssGSEA) to quantify the enrichment level of each BLCA sample in the 26 stem cell gene sets. The ConsensusClusterPlus package was used for consensus clustering and stem cell subtype screening of the ssGSEA scores. Briefly, \(k\)-means clustering was performed using 50 iterations (each using 80% of samples). The best cluster number was determined by the clustering score for the cumulative distribution function (CDF) curve, and the relative changes in the area under the CDF curve were evaluated.

**Survival analysis**

The Kaplan–Meier curve was used to describe the differences in survival of patients with BLCA in different
datasets for classifying stem cell subtypes. We compared the survival prognosis of patients with BLCA (overall survival (OS), relapse-free survival (RFS), and progression-free survival (PFS)). The log-rank test used \( P < 0.05 \) as the threshold to detect significant differences in survival time.

**Immune checkpoint inhibitor treatment response prediction**

Tumor immune dysfunction and exclusion is a calculation method for simulating tumor immune escape primarily by examining how the expression of each gene in the tumor interacts with the level of cytotoxic T cell (CTL) infiltration to affect patient survival [23]. We used TCGA’s FPKM RNA_seq expression profile combination subclass mapping method to predict the clinical response of BLCA stem cell subtypes to immune checkpoint blockade [24].

**Chemical response prediction**

We used TCGA’s FPKM RNA seq expression profile to predict the chemotherapy response of each sample based on the largest publicly available pharmacogenomics database (Genomics of Drug Sensitivity in Cancer (GDSC), https://www.cancerrxgene.org/); six commonly used chemotherapeutic agents were selected, namely, cisplatin, doxorubicin, gemcitabine, sunitinib, methotrexate, and vinblastine. The prediction process was conducted using the R package “pRRophetic” where the half-maximum inhibitory concentration IC50 of the sample was estimated using ridge regression, and the accuracy of the prediction was evaluated using 10-fold cross-validation, according to the GDSC training set. All parameters were set to the default values, and the repeated gene expression was averaged.

**Pathway enrichment analysis**

We compared the biological changes in every two subtypes in TCGA dataset and used h.all.v7.1.symbols.gmt as the reference gene set for the gene set enrichment analysis (GSEA). The analysis was performed using 1000 permutations, a \( < 0.05 \) false discovery rate (FDR) as the screening threshold, and GSEA version 4.0.1.

**Evaluation of immune cell infiltration level, tumor purity, and stromal content in BLCA**

ESTIMATE was used to evaluate the level of immune cell infiltration, tumor purity, and stromal content in the BLCA stem cell typing [25].

**Comparison of immune cell fraction between BLCA stem cell subtypes**

CIBERSORT is an algorithm that deconvolves the expression matrix of 22 human immune cell subgroups and can be used to estimate the proportion of immune cells [26]. We set the permutations to 1000 and used \( P < 0.05 \) as the screening threshold. The Kruskal–Wallis test was used to compare the differences in immune cell components of each BLCA stem cell subtype.

**Gene co-expression network analysis**

To identify key genes or gene networks that characterize various stem cell subtypes in BLCA, we performed weighted correlation network analysis (WGCNA) [27] to detect gene modules associated with stem cell subtypes. The gene matrix is composed of 4876 differential genes in BLCA control normal tissues (the difference is generated by limma package in R, \(|\log 2 \text{ fold change}| > 1, P < 0.05\)). WGCNA network construction and module detection used the unsigned topological overlap matrix; the best soft threshold (power) was set to 3, the minimum number of genes in the module was 50, and the branch merge interception height was 0.25. The hub gene was defined as that which has a Pearson correlation (due to the generally low value of the connection weight, the Pearson correlation was used) of greater than 0.30, with connections to at least 10 genes. The gene co-expression network was visualized using the CytoScape 3.7.1 software. Wilcoxon tests were used to examine the expression differences of hub genes between stem cell subtypes. The results of survival analysis were divided into high and low groups based on the median expression of the transcription factor using the GEPIA (http://gepia.cancer-pku.cn/) database [28]. The log-rank test was used for survival distribution. Top 20 enrichment pathways were obtained using Metascape (http://metascape.org/gp/index.html#/main/step1).

**Statistical analysis**

A comparison of the estimated IC50 of BLCA stem cell subtypes was performed using the Kruskal–Wallis test. \( CD274 \) expression differences between stem cell subtypes were evaluated using ANOVA in R. All tests were two-tailed, and \( P < 0.05 \) was considered as statistically significant.

**Results**

**BLCA subtypes identified based on stem cell gene sets**

We collected 26 stem cell gene sets representing unique self-regenerating properties (Supplementary Table 1) and quantified the scores of 26 stem cell gene sets in each sample using ssGSEA. We used the ConsensusClusterPlus package to divide all tumor samples into \( k \) \( (k = 2–9) \) different subtypes. The CDF curve based on the consensus scores achieves the best division when \( k = 3 \). Additionally, the principal component analysis results indicated that the ssGSEA scores, based on the 26 stem cell gene sets, were divided into 3 subtypes.
(Fig. 2a–d), which were defined as SCE_H, SCE_M, and SCE_L (Fig. 2e). Similarly, we performed the same clustering and subtyping for the remaining datasets E-MTAB-4321, GSE13507, GSE31684, GSE32548, and GSE32894 (Supplementary Figure 1).

When using ESTIMATE to evaluate the level of immune infiltration in all datasets, we found that the SCE_H immune score in all 6 datasets was much higher than that for other subtypes, and SCE_L showed the lowest immune score (Supplementary Figure 2). The comparison of the stromal content showed the same trend (SCE_H > SCE_M > SCE_L). However, the comparison of tumor purity of the BLCA stem cell subtypes showed opposite results. SCE_H and SCE_L showed the lowest and highest tumor purity (SCE_L > SCE_M > SCE_H), respectively. This is consistent with the results observed for most HLA genes and immune cell marker genes evaluated, such as CD8A (CD8 T cells), GZMA (cytotoxic cells), IFNG (Th1 cells), PMCH (Th2 cells), CD68 (macrophages), and IL17A (Th17 cells), among others, which were significantly upregulated and downregulated in SCE_H and SCE_L, respectively (Supplementary Figure 3).

Due to the close association between BLCA stem cell subtypes and immunity, we focused on the differential expression of CD274 (PD-L1) in each subtype (Fig. 3). In all six datasets, SCE_H and SCE_L showed the highest and lowest expression levels, respectively, indicating that the BLCA subtype SCE_H was more sensitive to anti-PD-1 immunotherapy than the remaining subtypes. Subsequent immune checkpoint inhibitor treatment response prediction and survival analysis have both confirmed these results.

**Survival of patients with different BLCA stem cell subtypes**

Since BLCA is a heterogeneous disease with a high recurrence rate, exploring the association between subtype classification and clinical prognosis is beneficial for prognosis assessment and clinical management of BLCA. We performed OS, RFS, and PFS analysis on the six datasets. Unexpectedly, all datasets showed consistent

![Fig. 2](image_url)

**Fig. 2** Identification of three stem cell subtypes in the BLCA TCGA cohort. 

a Consensus score matrix for BLCA samples when \( k = 3 \). A higher consensus score between two samples indicates that they are more likely to be assigned to the same cluster in different iterations. 

b CDF describes a real random variable of its probability distribution based on consensus scores for different subtype numbers \( (k = 2–9) \). 

c Delta area curve of all samples when \( k = 3 \). 

d Three-dimensional plot for ssGSEA scores in three stem cell subtypes; each dot indicates a sample, and different colors represent different subtypes. 

e Heatmap of ssGSEA scores for three subtypes. BLCA, bladder cancer; ssGSEA, single-sample gene set enrichment analysis; CDF, cumulative distribution function; TCGA, The Cancer Genome Atlas.
trends (Fig. 4). SCE_H and SCE_L showed the worst and best survival in the prognostic analysis (SCE_L > SCE_M > SCE_H), respectively. The P values of the log-rank for the OS of TCGA, GSE31684, GSE32548, and GSE32849 were 5.631e−4, 0.038, 2.158e−4, and 8.755e−6, respectively. The P values for the log-rank of PFS for TCGA, E-MTAB-4321, and GSE13507 were 0.004, 3.976e−9, and 7.046e−4, respectively, and the log-rank P value of TCGA RFS was 0.032.

Prediction of therapeutic response of BLCA stem cell subtypes to immune checkpoint inhibitors

Based on the above results, we further evaluated the responses of the three subtypes to immunotherapy. At present, 5 PD-1/PD-L1 immunotherapy drugs have been approved by the Food and Drug Administration for treating BLCA. This includes nivolumab and pembrolizumab (both PD-1 inhibitors) approved in 2016 and 2017 for treating patients with locally advanced
or metastatic urothelial cancer who were administered first-line platinum-containing chemotherapy for 1 year [29–34]. We used the tumor immune dysfunction and exclusion algorithm to predict the likelihood of a response to immunotherapy. The results showed significant differences in the responses to immunotherapy among the SCE_H (20%, 32/158), SCE_M (42%, 71/168), and SCE_L groups (61%, 54/88) ($P = 2.951 \times 10^{-10}$).

We further performed subclass mapping to compare the expression profiles of the three stem cell subtypes which were defined using another published dataset containing 47 patients with melanoma who responded to immunotherapy [35]. In a pairwise comparison of the three subtypes, more promising results were observed in SCE_H for the anti-PD1 and anti-CTLA4 treatments compared to the other subtypes (Fig. 5a–c) (anti-PD1 therapy: SCE_H vs SCE_L, FDR = 0.036; SCE_H vs SCE_M, $P = 0.046$; SCE_M vs SCE_L, FDR = 0.048; anti-CTLA4 therapy: SCE_H vs SCE_L, FDR = 0.036; SCE_H vs SCE_M, FDR = 0.008). We further correlated the BLCA stem cell typing results with the published molecular typing and immunotyping results in TCGA cohort. SCE_H primarily corresponded to the molecular subtypes luminal-infiltrated and basal squamous and C1 and C2 for immune subtypes; SCE_L primarily corresponded to luminal-papillary and C1–C4, and SCE_M showed a wide distribution of molecular and immune subtypes (Fig. 5d).

**Differences in sensitivity of stem cell subtypes to chemotherapy**

Since chemotherapy is a common treatment strategy for patients with BLCA, we selected six chemotherapeutic agents (cisplatin, doxorubicin, gemcitabine, sunitinib, methotrexate, and vinblastine) and evaluated the response of the three subtypes. We designed the prediction model on the GDSC cell line dataset using ridge regression and evaluated the satisfactory prediction accuracy using 10-fold cross-validation. We estimated the IC50 of each sample in TCGA dataset based on the prediction models of these six chemotherapeutic agents. For cisplatin, sunitinib, and vinblastine, SCE_L was the least sensitive while SCE_H was the most sensitive compared to the other subtypes. For doxorubicin, SCE_M was the most sensitive, while for gemcitabine and methotrexate, SCE_M was the most sensitive and SCE_L was the least sensitive relative to the other subtypes (Fig. 6).

---

**Fig. 5** BLCA stem cell subtype immunotherapy response prediction. a SCE_H vs SCE_L sensitive response to PD1 and CTLA4 inhibitors (Benjamini and Hochberg corrected $P = 0.036$, $P = 0.036$). b SCE_H vs SCE_M sensitive response to PD1 and CTLA4 inhibitors (Benjamini and Hochberg corrected $P > 0.05$, $P = 0.008$). c SCE_M vs SCE_L sensitive response to PD1 and CTLA4 inhibitors (Benjamini and Hochberg corrected $P = 0.008$, $P > 0.05$). d Sankey chart showing the distribution of BLCA stem cell subtypes in C1–C6 (C5 was not available for BLCA) and molecular subtypes. BLCA, bladder cancer; SCE_H, high stem cell enrichment; SCE_M, medium stem cell enrichment; SCE_L, low stem cell enrichment.
Differences among 22 human immune cell subgroups of BLCA stem cell subtypes in CIBERSORT
To explain the difference in survival of patients with different BLCA stem cell subtypes, we used the CIBERSORT algorithm to calculate the proportions of 22 immune cells in each subtype of the six datasets, with $P < 0.05$ as the threshold for screening. The results showed that the proportions of macrophages M0, M1, and M2 had an upward trend in the SCE_H subtype (except for GSE13507) and the proportion of regulatory T cells (Tregs) was significantly ($P < 0.05$) increased in the SCE_H subtype (Supplementary Figure 4A–F). We also used the ssGSEA scores of immune cells in each cohort as continuous variables and performed univariate Cox analysis (Supplementary Table 2). Further, we divided the patients according to the optimal expression cutoff of CD274 and each stemness gene into risk groups I, II, III, and IV (for example, CD274 and CD44 corresponded to CD274_low CD44_low, CD274_high CD44_low, CD274_low CD44_high, and CD274_high CD44_high). According to the scatter plot, for each pair of risk groups divided by CD274 and the optimal threshold of stemness gene expression, patients with stem cell subtypes primarily belonged to groups I and III (87–88%), while the...
SCE_L subtype was primarily concentrated in risk group III of the CD274 and GATA3, ID1, MYC, SOX9, CXCL8 gene pairs, and in the CD274 and CD44, finally, the HIF1A, MYC, SOX9, CXCL8 gene pairs were concentrated in risk group I. Further survival analysis of groups I and III of each gene pair showed that patients with the higher SCE_L subtype had better survival than those with a lower SCE_L subtype (Fig. 7i–o). This is consistent with the observation that patients in the SCE_L group had the longest survival duration.

**GSEA for BLCA stem cell subtypes**

To explore the biological changes caused by differences in the enrichment of stem cells, we conducted a pairwise comparison of the GSEA results for each subtype.
By selecting at least one pathway with an FDR < 0.05, we found that as the enrichment of stem cells increased, the epithelial-mesenchymal transition (EMT) became more significant. EMT is considered as a signal of malignant transformation in all cancers, giving cells the ability to metastasize and invade, by imparting stem cell characteristics, reducing apoptosis and aging, and resisting chemical and immunotherapy [36, 37] (Supplementary Table 3). EMT can also activate multiple pathways; regulate cell metabolism, angiogenesis, proliferation, and migration; and enable cells to respond to hypoxic environments. Pathways are also significantly enriched, for example, during hypoxia, angiogenesis, inflammatory response, IL6-mediated JAK-STAT signaling pathway, and KRAS signal upregulation (Fig. 8). These pathways together constitute a vicious circle of cancer occurrence, proliferation, invasion, and metastasis.

**Somatic mutation landscape of BLCA stem cell subtypes with identified pathways**

The tumor suppressor genes, TP53 and RB1, play an important role in regulating cell division [38]. Inactivation, mutations, and deletions of TP53 and RB1 are one of the primary causes of BLCA [39]. The mutation frequency of TP53 and RB1 in SCE_L (31% and 4%) was much lower than that in SCE_M (56% and 25%) and SCE_H (47% and 18%). However, the mutation frequency of STAG2, one of the most commonly mutated genes in BLCA [40], in SCE_L (24%) was significantly higher than that of SCE_M (11%) and SCE_H (9%). In the identified EMT pathways, the mutation frequency of COL6A3, LRP1, and FBN2 in SCE_L (12%, 12%, and 13%, respectively) was significantly higher than that of SCE_M (3%, 6%, and 5%, respectively) and SCE_H (7%, 4%, and 4%, respectively). In the hypoxia pathway, no MYH9 mutation was observed in SCE_L (0%), while the mutation frequencies in SCE_M and SCE_H were 5% and 8%, respectively. In addition, the mutation frequencies of CDKNIA in SCE_L, SCE_M, and SCE_H were 13%, 11%, and 6%, respectively, and in the KRAS signaling pathway, the mutation frequencies of RELN in SCE_L, SCE_M, and SCE_H were 12%, 6%, and 5%, respectively (Fig. 9). High-frequency gene mutations during angiogenesis and inflammation and in the IL6/JAK/STAT signaling pathway were not obtained.

**Key gene networks identified in BLCA stem cell subtypes**

WGCNA is used to describe correlation patterns between genes. Using microarray gene expression data, or RNA-seq gene expression data, WGCNA can be used to identify highly correlated gene sets (module), which are randomly assigned with different colors. The colors are
only used to distinguish different modules with no practical meaning or associated value. WGCNA promotes a network-based genetic screening method that can be used to identify candidate biomarkers or therapeutic targets. A total of 14 gene modules were generated based on WGCNA. Among them, brown and black modules showed the strongest association with stem cell subtypes. The brown module was positively correlated with SCE\_L and negatively correlated with SCE\_H and SCE\_M, while the black module was positively correlated with SCE\_H and negatively correlated with SCE\_M and SCE\_L (Fig. 10a). We intersected the genes in the black module most correlated with SCE\_H and the genes in the brown module most correlated with SCE\_L with human transcription factors identified using Cistrome (http://cistrome.org/). Two pairs of transcription factor regulators were identified in the brown and black modules (Fig. 10b, c), IRF5 and GATA3 and GRHL2 and GATA6. IRF5 is a key transcription factor regulating the differentiation of M1 macrophages into M2, enabling its anti-inflammatory role; these cells also influence tissue repair and reconstruction as well as cancer occurrence [41–43]. GATA3 is a type 2 helper T cell (Th2) cytokine-specific transcription factor and a key stemness gene that regulates cell differentiation. It enables Th2 to express IL-4 and other cytokines, promotes antibody production, mediates humoral immunity, and suppresses anti-tumor immunity [21, 44, 45]. Therefore, GATA3 may act as a tumor suppressor gene in BLCA. GRHL2 and GATA6 play various regulatory roles during embryonic development, damage repair, epidermal barrier formation, tracheal epithelial formation, and neural tube development. They also play an important role in the occurrence and development of tumors, cell proliferation, invasion, and metastasis, which were verified by the relevant pathways identified in the black module [46–52] (Fig. 10d–e). Thus, we identified GATA3 and GATA6 as important transcription factors with opposite expression and effects on tumor prognosis in patients with different BLCA stem cell subtypes (Fig. 10f–i).

**Discussion**

BLCA is one of the primary malignant tumors that endangers human health. Although several molecular genotyping schemes for BLCA have been proposed, they remain in their infancy stage compared with those of breast cancer [53–56]. A unified, well-developed, and highly feasible molecular typing scheme is required for better diagnosis and treatment of BLCA. Additionally, to date, no studies have classified BLCA based on stem cell gene sets, and thus, we used specific stem cell gene sets to identify and verify our new classification for BLCA. BLCA can be divided into three stable subtypes: SCE\_H, SCE\_M, and SCE\_L. Among them, the SCE\_H subtype showed the highest degree of immune infiltration and lowest tumor purity relative to the other subtypes. Patients with this subtype have the worst prognosis. This appears to contradict the previous suggestions that a higher degree of tumor immune infiltration is associated with a better prognosis. Using the CIBERSORT analysis of the immune cell fraction of stem cell subtypes, we found that the proportion of various cells associated with cytotoxicity in the SCE\_H subtype was significantly lower than that in the other subtypes, such as resting NK cells, CD8 T cells, and CD4 T cells. The number of macrophages M0, M1, and M2 was significantly increased. The composition of immune cells in the tumor microenvironment is complex and has different roles in various stages of tumor progression. Among them, macrophages show the highest content in tumor tissues and had the most significant regulatory effect on tumors. These cells, which can promote the
proliferation, invasion, and metastasis of tumor cells and induce tumor cells to develop immune tolerance, are known as tumor-associated macrophages (TAMs) (most studies have suggested that TAMs are primarily the M2 type) [57]. TAMs are often distributed around CSCs, and the amount of infiltration is closely correlated to the tumor histological grade and number of CSCs. Jinushi et al. [58, 59] found that the growth factor and inflammatory cytokine MFG-E8 and IL-6 secreted by TAM activate the STAT3 and sonic hedgehog signaling pathways, thus inducing CSC formation and enhance CSC tumorigenesis and resistance to chemotherapy. This is consistent with our survival analysis showing high ssGSEA scores for macrophage M0 and M2 and predicting a poor prognosis for patients with BLCA.

Fig. 10 Identification of key genes/networks based on WGCNA for BLCA stem cell subtypes. a Correlation heat map of 14 modules in 3 stem cell subtypes obtained from 4876 differential genes compared to normal tissues and BLCA of TCGA cohort. b In the brown module, the co-expression network is composed of RFS and GATA3. c In the black module, the co-expression network is composed of GRHL2 and GATA6. d Top 20 enrichment pathways obtained by using Metascape in the black module. e Top 20 enrichment channels arranged by -Log10P value in the black module. f Differential expression of GATA3 and h GATA6 in BLCA stem cell subtypes, Wilcoxon tests, *P < 0.05, **P < 0.01, ***P < 0.001. g Kaplan–Meier survival curve according to the median expression level using GEPIA, log-rank test. BLCA, bladder cancer; TCGA, The Cancer Genome Atlas; WGCNA, weighted correlation network analysis.
In addition, the immune checkpoint molecule, CD274 (PD-L1), was significantly upregulated in the SCE_H subtype. This molecule suppresses the proliferation and differentiation of T lymphocytes, promotes the differentiation of Tregs, and induces the secretion of cytokines, thereby suppressing the immune response [57]. Prediction of the anti-PD-1 treatment response showed that SCE_H is more sensitive to anti-PD-1 than other subtypes, indicating that CD274 is highly expressed in tumor/tumor stem cells and may be involved in the tumor immune escape process. SCE_H primarily corresponds with the luminal-infiltrated and basal-squamous molecular subtypes of BLCA. These two types of tumors exhibit high levels of immune infiltration and respond well to immune checkpoint therapy (PD-1, PD-L1, and CTLA4). This demonstrates that the stem cell classification we defined is closely related with the existing molecular typing of BLCA. For locally advanced/metastatic patients, the standard first-line treatment strategy is combination chemotherapy (MVAC) consisting of methotrexate, vinblastine, doxorubicin, and cisplatin, and dual therapy (GC) consisting of gemcitabine and cisplatin [60, 61]. Compared with other subtypes, SCE_H had the highest sensitivity to cisplatin, sunitinib, and vinblastine, while SCE_L was more sensitive to methotrexate and gemcitabine than the other subtypes. Methotrexate, an anti-folate chemotherapeutic agent, inhibits tumor cell DNA synthesis by inhibiting dihydrofolate reductase, thus halting tumor growth and reproduction. This effect may be attributed to the lower mutation rate of TP53 and RB1 during cell cycle in the SCE_L subtypes than other subtypes. Patients of SCE_H and other subtypes of BLCA may benefit from a combination of chemotherapy and immunotherapy.

Next, we explored the biological changes caused by different levels of BLCA stem cell enrichment and showed that SCE_H was positively correlated with EMT, hypoxia, angiogenesis, and inflammatory response activation in the tumor microenvironment. Studies have confirmed that early tumor cells are in an epithelioid state, and as the tumor progresses, more mesenchymal features are gradually obtained, such mesenchymal cells are resistant to therapy. In addition, activation of EMT in tumor cells induces the initial stages of tumors, also known as the CSC state, suggesting that EMT is an integral process in the progression of all types of malignant tumors [62]. In a breast tumor progression model, Morel et al. [63, 64] showed that following activation by the Ras-mitogen-activated protein kinase pathway, EMT induction can drive breast epithelial cells to obtain stem cell and tumorigenic properties of CSC. In addition, COL6A3 may be involved during the EMT process induced by TGF-β/Smad. COL6A3 silencing inhibits cell proliferation and angiopoiesis [65], which is consistent with COL6A3 having a lower mutation rate in SCE_H and SCE_M, and a higher mutation rate in SCE_L. Additionally, activation of the hypoxic pathway helps cancer cells to be more adaptable to the hypoxic environment. Under hypoxic conditions, the hypoxia-inducible factor (HIF-1α) pathway is activated to promote the release of vascular endothelial growth factor and platelet-derived growth factor, inducing endothelial cells from the original tumor blood vessels to proliferate, bud, and generate new tumor blood vessels, allowing tumors to invade and metastasize. Notably, immune cells infiltrating the tumor microenvironment can secrete a large number of cytokines and chemokines to promote EMT in tumor cells. Further, uncontrollable inflammatory lesions can regulate EMT in tumor cells, and a positive feedback loop can be formed between the inflammatory lesions and EMT, allowing the EMT process and uncontrollable inflammatory state to continue. These common pathways constitute a vicious circle of tumorgenesis, development, drug resistance, and poor prognosis.

**Conclusion**

By identifying BLCA subtypes based on stem cell gene sets, we systematically analyzed the relationship between these subtypes in the tumor microenvironment and immune cells, immunotherapy/chemotherapy response, corresponding pathways, and key genes. These results provide a basis and reference for the clinical diagnosis and treatment of BLCA.

**Supplementary information**

Supplementary information accompanies this paper at https://doi.org/10.1186/s13287-020-01973-4.

Additional file 1: Table S1. The 26 stem gene sets used for identification of BLCA subtype. BLCA: Bladder cancer. Figure S1. Clustering heat map of stem cell subtype in (A) E-MTAB-4321, (B) GSE13507, (C) GSE13684, (D) GSE32548, and (E) GSE32894. Figure S2. Evaluation of immune cell infiltration level, tumor purity, and stromal content in BLCA. (A–F) Immune score, (G–L) stromal score (stromal content), and (M–R) tumor purity in all six datasets. * P < 0.05, ** P < 0.01, *** P < 0.001; ns means not significant. BLCA: bladder cancer. Figure S3. Comparisons of the expression levels of immune-related genes between BLCA subtypes. (A–C) Expression levels of HLA genes between BLCA subtypes in TCGA, E-MTAB-4321 and GSE32894. (D–E) Expression levels of immune cell subgroup marker genes between BLCA subtypes: Kruskal–Wallis test, * P < 0.05, ** P < 0.01, *** P < 0.001; ns means not significant. BLCA: bladder cancer. Figure S4. Difference analysis of 22 human immune cell subgroups. (A–J) Fraction of different immune cell subgroups in (A) TCGA, (B) E-MTAB-4321, (C) GSE13507, (D) GSE13684, (E) GSE32548, and (F) GSE32894. (G–L) stromal score (stromal content), and (M–R) tumor purity in all six datasets. * P < 0.05, ** P < 0.01, *** P < 0.001; ns means not significant. BLCA: bladder cancer. Table S2. Univariate Cox analysis for all six datasets. Table S3. GSEA for BLCA stem cell subtypes.
Abbreviations

SCE_H: High stem cell enrichment; SCE_M: Medium stem cell enrichment; SCE_L: Low stem cell enrichment; BLCA: Bladder cancer; CSC: Cancer stem cell; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; FPKM: Fragments per kilobase of transcript per million mapped reads; ssGSEA: Single-sample gene set enrichment analysis; CDF: Cumulative distribution function; OS: Overall survival; RFS: Relapse-free survival; PFS: Progression-free survival; CTL: Cytotoxic T cell; WGCNA: Weighted correlation network analysis; ANOVA: Analysis of variance; Tregs: Regulatory T cells; OCLR: One-class logistic regression; EMT: Epithelial-mesenchymal transition; NES: Normalized enrichment score; FDR: False discovery rate; Th2: Type 2 helper T cell; TAMs: Tumor-associated macrophages; nf-κB: Nuclear factor-kappa B; HIF-α: Hypoxia-inducible factor; GSEA: Gene set enrichment analysis

Acknowledgements

We would like to thank Editage (www.editage.cn) for the English language editing.

Authors’ contributions

CT analyzed and interpreted the patient data and was a major contributor in writing the manuscript. JM proposed the conception and performed the data analyses. XL was responsible for the critical work of the manuscript. ZL corrected the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (2016JANSF-A0380306), a self-generated project of the Guangxi Zhuang Autonomous Region Health Department (Z20170816), China.

Availability of data and materials

The datasets generated or analyzed for this study can be found in the TCGA Knowledge Base (https://portal.gdc.cancer.gov/repository), GEO (https://www.ncbi.nlm.nih.gov/geo/), and ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) databases. Twenty-six stem cell gene sets were obtained from StemChecker (http://stemchecker.sysbiolab.eu/).

Ethics approval and consent to participate

All data were downloaded from public databases and therefore did not require approval and review by the ethics committee.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

1Department of Urology, The First Affiliated Hospital of China Medical University, Shenyang 110001, China. 2Department of Oncology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou 450014, China. 3Department of Oncology, Affiliated Hospital of Guilin Medical University, Guilin 541001, China. 4Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin 541001, China.

Received: 20 May 2020 Accepted: 12 October 2020

Published online: 28 October 2020

References

1. Alifrangis C, et al. Molecular and histopathology directed therapy for advanced bladder cancer. Nat Rev Urol. 2019;16(8):465–83.
2. Martinez Rodriguez RH, et al. Bladder cancer: present and future. Med Clin (Barc). 2017;149(10):449–55.
3. Antoni S, et al. Bladder cancer incidence and mortality: a global overview and recent trends. Eur Urol. 2017;71(1):96–108.
4. Pfannstiel C, et al. The tumor immune microenvironment drives a prognostic relevance that correlates with bladder cancer subtypes. Cancer Immunol Res. 2019;7(6):923–38.
5. Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507(7492):315–22.
6. Damrauer JS, et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. Proc Natl Acad Sci U S A. 2014;111(8):3110–5.
7. Choi W, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell. 2014;25(2):152–65.
8. Spodahl G, et al. A molecular taxonomy for urethelial carcinoma. Clin Cancer Res. 2012;18(12):3377–86.
9. Yu Z, et al. Cancer stem cells. Int J Biochem Cell Biol. 2012;44(12):2144–51.
10. Cabrera MC, et al. Cancer stem cell plasticity and tumor hierarchy. World J Stem Cells. 2015;7(1):27–36.
11. Matsui WH. Cancer stem cell signaling pathways. Medicine (Baltimore). 2016;95(Suppl 1):S8–S19.
12. Plaks V, et al. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? Cell Stem Cell. 2015;16(3):225–38.
13. Shackleton M, et al. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009;138(5):822–9.
14. Wei J, et al. Globoblastoma cancer-initiating cells inhibit T-cell proliferation and effector responses by the signal transducers and activators of transcription 3 pathway. Mol Cancer Ther. 2010;9(1):67–78.
15. Santisteban M, et al. Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. Cancer Res. 2009;69(7):2887–95.
16. Wu CT, et al. Predictive value of CD44 in muscle-invasive bladder cancer and its relationship with IL-6 signalling. Ann Surg Oncol. 2018;25(12):3518–26.
17. Li Y, et al. Bladder cancer stem cells: clonal origin and therapeutic perspectives. Oncotarget. 2017;8(39):6668–70.
18. Zhou J, et al. Activation of the PTEN/mTOR/STAT3 pathway in breast cancer correlates with a poor outcome. Clin Cancer Res. 2014;20(10):77.
19. Wang Z, et al. Exploitation of the Notch signaling pathway as a novel target for cancer therapy. Anticancer Res. 2008;28(6A):3621–30.
20. Shachaf CM, et al. Tumor dormancy and MYC inactivation: pushing cancer to the brink of normalcy. Cancer Res. 2005;65(11):4471–4.
21. Yang Z, et al. The KMT1A-GATA3-STAT3 circuit is a novel self-renewal signaling of human bladder cancer stem cells. Clin Cancer Res. 2017;23(21):6673–85.
22. Pinto JP, et al. StemChecker: a web-based tool to discover and explore stemness signatures in gene sets. Nucleic Acids Res. 2015;43(W1):W72–7.
23. Jiang P, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–8.
24. Hoshida Y, et al. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One. 2007;2(11):e1195.
25. Yoshiiara K, et al. Inferred tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2621.
26. Newman AM, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773–82.
27. Langfelder P, et al. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
28. Tang Z, et al. GEPA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W96–W102.
29. Inman BA, et al. Atezolizumab: a PD-L1-blocking antibody for bladder cancer. Clin Cancer Res. 2017;23(8):1886–90.
30. Sidaway P. Bladder cancer: atezolizumab: an alternative to cisplatin? Nat Rev Urol. 2017;14(2).67.
31. Cattini C, et al. Atezolizumab and bladder cancer: facing a complex disease. Lancet. 2018;391(10118):305–6.
32. Gourd E. Neoadjuvant pembrolizumab in bladder cancer. Lancet Oncol. 2018;19(12).e669.
33. Safaty M, et al. Cost-effectiveness of pembrolizumab in second-line advanced bladder cancer. Eur Urol. 2018;74(1):57–62.
34. Sidaway P. Bladder cancer: pembrolizumab is superior to chemotherapy. Nat Rev Urol. 2017;14(5):261.
35. Roh W, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci Transl Med. 2017;9(379):eaah3560.
36. McConkey DJ, et al. Molecular subtypes of bladder cancer. Curr Oncol Rep. 2018;20(10):77.
40. Balbas-Martinez C, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet. 2013;45(12):1464–9.

41. Ban T, et al. Regulation and role of the transcription factor IRF5 in innate immune responses and systemic lupus erythematosus. Int Immunol. 2018;30(11):529–36.

42. Almuttaqi H, et al. Advances and challenges in targeting IRFs, a key regulator of inflammation. FEBS J. 2019;286(9):1624–37.

43. Krausgruber T, et al. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. Nat Immunol. 2011;12(3):231–8.

44. Li Y, et al. GATA3 in the urinary bladder: suppression of neoplastic transformation and down-regulation by androgens. Am J Cancer Res. 2014;4(5):461–73.

45. Li Y, et al. Loss of GATA3 in bladder cancer promotes cell migration and invasion. Cancer Biol Ther. 2014;15(4):428–35.

46. Martinielli P, et al. GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. Gut. 2017;66(9):1665–76.

47. Peng T, et al. The interaction of LOXL2 with GATA6 induces VEGFA expression and angiogenesis in cholangiocarcinoma. Int J Oncol. 2019;55(3):657–70.

48. Kamijo H, et al. Aberrant CD137 ligand expression induced by GATA6 overexpression promotes tumor progression in cutaneous T-cell lymphoma. Blood. 2018;132(18):1922–35.

49. Chia NY, et al. Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. Gut. 2016;65(5):701–19.

50. Rosas M, et al. The transcription factor Gata6 links tissue macrophage phenotype and proliferative renewal. Science. 2014;344(6184):645–8.

51. Chen AF, et al. GRHL2-dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naive pluripotency. Cell Stem Cell. 2018;23(2):226–38 e4.

52. Nishino H, et al. Grainyhead-like 2 (GRHL2) regulates epithelial plasticity in pancreatic cancer progression. Cancer Med. 2017;6(11):2686–96.

53. Goldhirsh A, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206–23.

54. Coates AS, et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26(8):1533–46.

55. Curigliano G, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. Ann Oncol. 2017;28(8):1700–12.

56. Burstein HJ, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. Ann Oncol. 2019;30(10):1541–57.

57. Liu C, et al. Treg cells promote the SREBP1-dependent metabolic fitness of tumor-promoting macrophages via repression of CD8(+) T cell-derived interferon-gamma. Immunity. 2019;51(2):381–97 e6.

58. Tan Y, et al. MFG-E8 is critical for embryonic stem cell-mediated T cell immunomodulation. Stem Cell Rep. 2015;5(3):741–52.

59. Jinushi M, et al. Tumor-associated macrophages regulate tumorigenicity and anticancer drug responses of cancer stem/initiating cells. Proc Natl Acad Sci U S A. 2011;108(30):12425–30.

60. Cheetham PJ, et al. New agents for the treatment of advanced bladder cancer. Oncology (Williston Park). 2016;30(6):571–9 88.

61. Knowles MA, et al. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer. 2015;15(1):25–41.

62. Dongre A, et al. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019;20(2):69–84.

63. Morel AP, et al. Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS One. 2008;3(8):e2888.

64. Mani SA, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704–15.

65. Huang Y, et al. Collagen type VI alpha 3 chain promotes epithelial-mesenchymal transition in bladder cancer cells via transforming growth factor β (TGF-β)/Smad pathway. Med Sci Monit. 2018;24:5346–54.