Identification of key genes and microRNA regulatory network in development and progression of urothelial bladder carcinoma

Miao Mo, Xiheng Hu, Wei He, Xiongbing Zu, Long Wang, Yangle Li

Department of Urology, Xiangya Hospital, Central South University, Changsha, China

Contribution: (I) Conception and design: M Mo, Y Li; (II) Administrative support: Y Li; (III) Provision of study materials or patients: X Hu, W He, X Zu; (IV) Collection and assembly of data: X Hu, W He, X Zu, L Wang; (V) Data analysis and interpretation: M Mo, Y Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yangle Li, MD. Department of Urology, Xiangya Hospital, Central South University, Changsha, China. Email: lyle2003@csu.edu.cn.

**Background:** Bladder cancer as other cancers contains multiple dynamic alterations in progression. Theoretically, large number of genes participates in cancer progression. In the present study, the interconnections of genesets defined by Gene Set Enrichment Analysis (GSEA) and tumor histopathological stages were characterized. In addition, the outcomes with genesets were discussed in bladder cancer.

**Methods:** Transcriptome data from 411 tissues of urothelial bladder carcinoma and 19 samples from adjacent tissues were retrieved from The Cancer Genome Atlas (TCGA) database. Single-sample GSEA (ssGSEA), cluster analysis of geneset enrichment scores and genesets as indicators in prognosis were applied to elucidate the correlations between genesets and bladder cancer progression.

**Results:** Chemical and genetic perturbations (CGP), canonical pathways (CP), CP:BIOCARTA (BioCarta gene sets), CP:KEGG (KEGG gene sets) and CP:REACTOME (Reactome gene sets) in C2 collection, upstream cis-regulatory motifs serum response factor (SRF) in C3 collection, KRAS in C6 collection and C8+ T cells in C7 collection were observed as enriched by ssGSEA. The cluster 2 identified from cluster analysis shows a more immune active microenvironment which tended to increase in stage II and decreased in stage IV indicating the crucial role in bladder cancer progression. miR-450, miR-518s, transcription factor PAX3, KRAS and PTEN were potential markers for outcomes of urothelial bladder carcinoma. Activating tumor immune microenvironment had deteriorated prognosis of patients with bladder cancer.

**Conclusions:** Our findings demonstrated that activating tumor immune microenvironment is a negative factor for outcomes of urothelial bladder carcinoma. These data provided a potential combination strategy for patients with bladder cancer.

**Keywords:** Bladder cancer; immune microenvironment; The Cancer Genome Atlas (TCGA); prognosis; genesets

doi: 10.21037/tau-20-1124

**View this article at:** http://dx.doi.org/10.21037/tau-20-1124

**Introduction**

Bladder cancer is the ninth most common malignant tumor and about 430,000 new cases occurs per year. Unfortunately, the treatment for bladder cancer is still limited and thus new strategies to obtain clinical benefit are developed. Immunotherapy has been introduced in bladder cancer as early as 1940s by using Bacillus Calmette–Guérin and developed in recent years (1). In preliminary Phase I trial, improved survival rate with 50% response rate was found in bladder cancer patients treated with a monoclonal antibody that targets programmed death-ligand 1 (PD-L1) MPDL3280A (2). Another immunotherapy is that combined Bacillus Calmette–Guérin and IL-10 antibody mediated by Th1 response (3). These immunotherapies hint that changes of tumor microenvironment in bladder cancer is crucial (4,5). Thus, immune factors are valuable as prognostic indicators which still need development and
evaluation. On the other hand, immune signatures would promise new combination therapies to improve therapeutic effects for bladder cancer via regulating immune cell infiltration (6).

Besides the tumor microenvironment, large information about pathogeny and markers of bladder cancer are still scarce (7). Fortunately, recent progresses have been made based on multi-omics data and numerous databases facilitate the association studies to identify valuable information (8). For instance, the Cancer Genome Atlas (TCGA) includes over 2.5 petabytes data of multi-omics and corresponding clinical indicators from patients which provides an unprecedented chance to illustrate comprehensive association between cancer development stages and molecular signatures (9). The association in gastric cancer between angiogenesis and cytotoxic signatures were constructed via elucidation by analyzing the data from TCGA which identified key characteristics for potential combination strategy and further clinical investigations (10). Another report using TCGA data to establish association between DNA methylation and gene expression in cancers showing the effects of location of the CpG site in the gene body on gene expression (11). These reports contribute to the understanding of cancer makers for diagnosis and prognosis. However, the global information on correlation between developmental stages of bladder cancer and signatures of transcript expression are still limited. Thus, we presumed that high correlation could be identified which would provide possible markers for aggressive disease and therapeutic information for improved characterization in bladder cancer patients.

In the present study, we performed a direct global single-sample Gene Set Enrichment Analysis (ssGSEA) of gene expression in bladder cancer based from TCGA data. Previous NGS data analysis based on single gene level may missing biological meaningful changes due to gene-level analysis are less obvious than geneset. Biological processes are realized through sets of genes functionally related. By investigating expression alterations on genesets may help to uncover meaningful functional pathways and regulation mechanisms at higher resolutions than single gene levels. Our analyzes indicated that several immunological signatures as well as gene and microRNA (miRNA) expressions were highly correlated with progression of bladder cancer. Furthermore, the characteristics of specific signatures and their prognostic implications were assessed. Collectively, these clues revealed the association of transcript expression signatures and bladder cancer which promise possible prognostic markers. We present the following article in accordance with the STROBE reporting checklist (available at: http://dx.doi.org/10.21037/tau-20-1124).

Methods

Data sources

The RSEM-normalized RNA sequencing data were download from TCGA BLCA Urothelial Bladder Carcinoma (TCGA-BLCA) [https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Bladder%20Carcinoma%20(BLCA)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443]. In total, 411 samples from urothelial bladder carcinoma patients and 19 samples from adjacent tissues from urothelial bladder carcinoma. Additional data sets including status of patients were downloaded from TCGA database. Genesets data were collected from Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/index.jsp, version MSigDB 6.2).

ssGSEA analysis

To build the biological association changes in molecular signaling pathways from urothelial bladder carcinoma patients, the ssGSEA analysis was employed to identify significantly enriched genesets based on TCGA data. The ssGSEA determined genesets in different groups according to functions. The pairwise comparisons were performed in all the experimental conditions and ranked according to changes of expressions. The geneset enrichment score was calculated which represented extent of overrepresentation in each geneset. The ssGSEA analysis was conducted with 1,000 random gene set membership assignments. When P<0.01 and FDR <25%, the genesets were confirmed as significant enrichment. MSigDB was used to identify the gene set definitions as 4 major collections including C2 (curated gene sets), C3 (motif gene sets), C6 (oncogenic gene sets) and C7 (immunologic gene sets). The significant differences between urothelial bladder carcinoma and adjacent tissues were determined by R package: limma with |logFC|>0.01 and adjusted P<0.01.

Determination of association between genesets and phenotypes

To evaluate the association between genesets and phenotypes, we analyzed the survival differences in patients
from different clusters including 410 patients from TCGA database. In addition, the tumor stage information from these patients was employed to analyze the correlation between frequency of tumor stages and different clusters.

### Effects of genesets on prognosis

The prognosis with expression changes of genesets was assayed using multivariate Cox regression analysis. The multivariate Cox regression analysis was performed from SD top 500 geneset by R. P<0.01 as cutoff for significant.

### Statistical analysis

All statistical analysis was performed under R version 3.5.3. P<0.01 and FDR <25% were used as cutoffs for significant geneset enrichment outcomes. |logFC| >0.01 and adjusted P<0.01 was considered as statistically significant for differential ssGSEA test between urothelial bladder carcinoma and adjacent tissues. P<0.01 was used as cutoff for significant Cox regression survival analysis.

### Results

#### Correlation between genesets and urothelial bladder carcinoma

We first evaluated the correlation between MSigDB gene sets and urothelial bladder carcinoma incidence by mathematical method, which consisted of 411 cancer tissues and 19 adjacent tissues. The C2 collection includes chemical and genetic perturbations (CGP), canonical pathways (CP), CP:BIOCARTA (BioCarta gene sets), CP:KEGG (KEGG gene sets) and CP:REACTOME (Reactome gene sets) showed that differentiated genesets are mostly cancer related (Figure 1A). In C3 collection, we found that 3 genesets were enriched. Among them, the genes shared upstream cis-regulatory motifs serum response factor (SRF) showed a significant drop in tissues from bladder carcinoma (Figure 1B).

Only 1 gene set as KRAS was enriched in C6 collection. Similar to SRF, the “Genes up-regulated in epithelial kidney cancer cell lines when over-expressing an oncogenic form of KRAS gene” was significantly down regulated in urothelial bladder carcinoma (Figure 1C). Five genesets were determined as enrichments as gene sets related to the immune system in C7 collection. In these genesets, 3 of the 5 genesets were associated with CD8+ T cells showing significant stimulation in carcinoma (Figure 1D).

### Association of genesets clusters with bladder cancer

Considered that several genesets from collections of C2, C3, C6 and C7, we used geneset enrichment score to determine crucial genesets clusters which participated in progression of bladder cancer by distance matrix. We used 500 genesets with the largest standard deviation from the four collections. Four clusters were mathematically established and presented in the symmetric heatmaps with pair-wise pearson correlations (Figure 2A). The principal component analysis (PCA) suggested that the 4 clusters could be clearly distinguished (Figure 2B).

To analyze the survival prognostic probability of 4 clusters, the survival analysis on 410 patients dder carcinoma. The survival analysis proved to be significant (P=0.047). The cluster 3 showed the poorest prognostic while cluster 4 had higher survival prognostic probability (Figure 2C). In addition, the clusters were co-analyzed with tumor histopathological stages. In stage I, only cluster 1 and 4 were found while cluster 2 and cluster 3 were only observed in stage II–IV. Also, we found that the frequency of cluster 4 tended to decrease with tumor stage increase (Figure 2D). The analysis of clusters to tumor stages showed that cluster 3 had the highest frequency in stage IV while cluster 4 had the highest frequency in stage II (Figure 2E).

The heatmap showed the expression significance (adjusted P value <0.05) and the 20 genesets with the largest |log(fold change)| from each cluster. Samples in cluster 3 showed more immune active microenvironments, while samples in cluster 4 exhibited the opposite pattern (Figure 3).

### Outcomes with genesets

The cox hazard regression analysis was performed to predict outcomes with genesets and adjusted by age. In C2 collection, activation of Cytosolic Sulfonation and Protease seems prolong the patients’ survival rate (Figure 4A). MiR-450, miR-518s and transcription factor PAX3 appeared as a significant predictor for worse outcome of bladder carcinoma (Figure 4B). In C6 collection, upregulation of KRAS and downregulation of PTEN may predict worse outcomes (Figure 4C). In contrast, active immune microenvironment may predict worse outcomes in C7 collection (Figure 4D).
Figure 1 Single-sample Gene Set Enrichment Analysis (ssGSEA) analysis reveals functional differences between cancer tissues and adjacent tissues. Signature enrichment scores were subjected to differential analysis of C2 (A), C3 (B), C6 (C) and C7 (D) collection sets.
Discussion

A systems biology approach was conducted to build the correlation between MsigDB collection and progression of urothelial bladder carcinoma. This included several genesets that significantly expressed in urothelial bladder carcinoma tissues compared to adjacent tissues, including SRF, KRAS and CD8+ T cells. Lower expression of SRF was unexpected in bladder cancer as present result. Regarding that SRF participates in urinary bladder smooth muscle formation, we proposed that high expression of SRF is attributed to maintaining tissues of bladder (12). CD8+ T cells were significantly activated in bladder cancer deduced from TCGA data. In a previous study, high density of CD8+ T cells were observed and the CD8+ T cells distributed widely in bladder cancer tissues (4,13,14). In addition, the density of CD8+ T cells in bladder cancer had positive correlation with poor prognosis (2). Our result supported that as a marker for prognosis in patients with urothelial bladder carcinoma.

Considering that large amount of genesets that may mediate progression of urothelial bladder carcinoma, we used 500 genesets to determine crucial function cluster. Four clusters were identified and showed different prognosis in patients. This result demonstrated that the method of cluster analysis to determine genesets for cancer progression is a potential approach to reflect global expression patterns of genes mathematically. Similar studies also suggested that analysis of massive data may provide valuable information for understanding molecular basis of cancer development (15-17). Interestingly, the cluster 3 which was highly correlated with the progression had been defined as more immune active microenvironment. Associated with cancer stages, this cluster showed higher frequencies in stage
II, III and IV while had the lowest frequency in stage I. Meanwhile, based on our findings, the immune active microenvironment plays a crucial role in stage II and stage III which participates the tumor deterioration especially in invasion and migration. This is similar to previous studies that tumor immune active microenvironment regulates tumor growth and invasion in glioblastoma (18). Barar concluded that tumor microenvironment including immune active microenvironment infiltrates into cancer cells and promotes migration (19). TILs in the tumor immune microenvironment (TIME) have been researched for over 40 years (20). TILs can consist of T, B, and natural killer (NK) cells. CD8+ cytotoxic T-effector cells (TEFF) play a critical role in restraining tumor development, whereas CD4+ Th cells can have pro- or anti-tumor effects (21). TILs have long been proved to be prognosis indicators and predict response to immuno-oncologic treatments (22). Our findings further validated that strategies to treatment on immune active microenvironment influences the success for cancer therapy and understanding of immune subtypes of bladder cancer can help with better personalized treatments.

We then extended the analysis to effect of single geneset on prognosis. In this analysis, we found that miR-450, miR-518s and transcription factor PAX3 in C3 collection. miR-450 associated with human kidney cancer subtypes and cancer cell apoptosis had reported before (23). Overexpression of miR-450a were shown to suppressing multiple genes involved in the epithelial-to-mesenchymal transition (EMT) and reducing tumor migration, invasion and increased anoikis in ovarian tumor cell lines as well as reducing tumor growth in an ovarian tumor xenographic model (24). It has been shown that miR-518a regulated chemokine receptor CCR6 expression in colorectal cancer cells. Low expression of miR-518a-5p was proved to upregulate PIK3C2A and affect the cellular response to the drug, causing resistance to imatinib in GISTs (25). Pax genes consist of a family of transcription factors that are essentially required for the genesis of a variety of tissues and organs. PAX3 is a well-known gene that affects solid tumor alveolar rhabdomyosarcoma (26). Similar to C3 collection, KRAS and PTEN were two factors for prognosis. It is not surprising that upregulation of KRAS had poor outcomes since KRAS is a famous oncogene while downregulation of PTEN was factor for worse outcomes since KRAS is a well-known tumor suppressor gene (27,28). These data suggested that the findings of the present study are similar to major

Figure 3 Heatmap of expression significance from the 20 genesets with the largest |log(fold change)| from each cluster.
Figure 4 Cox hazard regression analysis of outcomes with genesets in C2 (A), C3 (B), C6 (C) and C7 (D) collection sets.
of cancer mechanisms (28-33). Thus, the bioinformatic method for identifying geneset from TCGA data is consistent with result from experimental medicine. But functions of above genes and microRNAs in bladder cancer are still lack of evidence. Our analysis tried to fill the gap and indicated bioinformatic method for identifying geneset from TCGA data is consistent with result from experimental medicine. This approach promises an accuracy and inexpensive method for finding potential markers in cancer. Still, we found that active immune microenvironment had predict worse outcomes in urothelial bladder carcinoma. As found in ssGSEA and cluster analysis, activating immune microenvironment indeed was a signature in bladder cancer. Recently, anti-PD1 immunotherapy is introduced to several cancer and has been proven to be an effective therapy including bladder cancer via regulating immune landscape and gene expressions (34-36). PD-L1 expression had strong associated T-cell infiltration and high expression of genes associated to Th1 and cytotoxic cell function was found in cancers (37-40). Thus, changing the status of immune active microenvironment by anti-PD1 immunotherapy may be a key process.

Conclusions

Our study identified the significant factors infiltrated in urothelial bladder carcinoma. CD8+ T cells, miR-450, miR-518s, transcription factor PAX3, KRAS and PTEN were found highly correlated with urothelial bladder carcinoma as targetable markers. The present data suggested that activating tumor immune microenvironment deteriorated prognosis of patients with bladder cancer. These clues suggested that ameliorate the immune niche may be an effective way for combination therapy. Further experimental and clinical research is required to determine the potential value of controlling tumor immune microenvironment in bladder cancer.

Acknowledgments

Funding: Natural Science Foundation of Hunan Province (2020JJ5915).

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at http://dx.doi.org/10.21037/tau-20-1124

References


Peer Review File: Available at http://dx.doi.org/10.21037/tau-20-1124

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tau-20-1124). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.
38. Mlecnik B, Bindea G, Angell HK, et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a