Controversial T1G3 bladder cancer is the key to revealing the changes in the biological functions of bladder cancer cells

Shen Pan

Shengjing Hospital of China Medical University

Yunhong Zhan

Shengjing Hospital of China Medical University

Xiaonan Chen

Shengjing Hospital of China Medical University

Bin Wu

Shengjing Hospital of China Medical University

Bitian Liu (✉ liu_bitian@163.com)

Shengjing Hospital of China Medical University <https://orcid.org/0000-0002-3747-9567>

Research

Keywords: T1G3, bladder cancer, GEO, TCGA, WGCNA, GSEA, marker

Posted Date: January 9th, 2020

DOI: <https://doi.org/10.21203/rs.2.20498/v1>

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

[Read Full License](#)

Abstract

Background T1G3 shows a higher chance of recurrence and progression among early bladder cancer types and the available treatment option is controversial. High recurrence and progression are the problems that need to be explored and solved. Changes in the internal signals of bladder cancer cells and differential genes may be the root cause of these problems.

Methods GSE120736, GSE19915, GSE19423, GSE32548 and GSE37815 datasets were obtained from Gene Expression Omnibus (GEO) to identify differentially expressed genes (DEGs). Bladder cancer transcript data from The Cancer Genome Atlas (TCGA) were clustered into different cell-specific gene sets according to weighted gene co-expression network analysis (WGCNA). Multiple sets of databases were used for gene expression comparison, functional enrichment, and protein interaction analysis, including The Human Protein Atlas, Cancer Dependency Map, Metascape, Gene set enrichment analysis, and DisNor.

Results DEGs were obtained through GEO data comparison and intersection. After WGCNA was proven to recognise cell-specific gene sets, candidate DEGs were selected and shown to be specifically expressed in cancer cells. Candidate DEGs were related to mitosis and cell cycle. Further, 12 functional candidate markers were identified from the sequencing data of 30 bladder cancer cell lines. These genes were all up-regulated and previously shown to be closely related to bladder cancer progression.

Conclusions Twelve functional genes with specific differential expression in bladder cancer cells were identified. WGCNA can identify the relatively specific expression sets of different cells in bladder cancer with greater tumour heterogeneity, which provides new perspectives for future cancer research.

Background

Bladder cancer (BC) is the ninth most frequently diagnosed cancer worldwide and 13th in terms of cancer-related deaths[1]. Most cases (75%) are diagnosed with non-muscle-invasive bladder cancer (NMIBC), yet 70% of NMIBCs typically recur and 25% progress to muscle-invasive disease[2]. In NMIBC, the best post-operative treatment of high-grade T1 (T1HG) or T1G3 has always been controversial, and no consensus has been reached yet. T1G3 with less differentiated cancer cells are more likely to relapse and progress than other T1 stage BCs[3]. Bladder preservation with bacillus Calmette-Guerin (BCG) induction and maintenance, second transurethral resection (TUR), or radical cystectomy are the current treatments for T1G3[4, 5]. Progression rate according to the stage at re-TUR in T1 BC patients treated with BCG can reach approximately 25%[6]. As a result, the search for the optimal management of T1G3 continues[7].

There is, therefore, a need to identify biomarkers that are capable of distinguishing between progressive and non-progressive T1 BC or identify multiple targets as the basis for combination therapy. A retrospective multicentre study, in which a total of 96 patients with T1G3 urothelial carcinoma of bladder were included, used Biomark Fluidigm Arrays based on the quantitative polymerase chain reaction

(qPCR), allowing validation of differentially expressed genes in a larger series. ANXA10, DAB2, HYAL2, SPOCD1, and MAP4K1 were identified to predict the progression in T1G3 BC patients[8]. Another study containing specimens from TUR of 92 patients with T1HG BC identified TOPO-2 α , p16, survivin, and E-CAD to assist the accurate prediction of survival in single T1 high-grade urothelial carcinoma, that is less than 3 cm in diameter[9]. In the last few years, marker research has been conducted not only in pathological tissues but also in liquid biopsies from blood and urine collected from patients with T1G3[10, 11].

Based on the computational principles, weighted gene co-expression network analysis (WGCNA) is not only used for topology network construction to identify highly co-expressed genes[12] but is also well suited to decompose high-throughput sequencing data into subsets of genes with cell-specific expression from high heterogeneity tissues. Fluctuations in the proportion of various cells within the tumour, as well as differences in the collection of pathological tissue sites, can result in significant differences in the sequencing results. The change in gene expression is mainly affected by a specific dominant cell population within the tumour. Genes are not specifically expressed in cells that are easily affected by the proportion of other cells. The expression level of the cell-specific genes is not easily affected by the proportion of other types of cells, changing with the change in cell ratio and having a high co-expression feature. Based on this, WGCNA used tumour heterogeneity to find specific gene subsets derived from cancer cells in BC.

In order to resolve the controversy surrounding T1G3 BC, it is important to understand the abnormal signals and find aberrantly expressed genes in cancer cells, rather than finding the differentially expressed genes (DEGs) obtained after comparison of samples. With the increased use of the high-throughput sequencing and microarray data, the application of bioinformatics in BC is increasing[13–15]. However, few studies are focused on T1G3 either due to the insufficient pathology or the absence of normal tissue for comparison. In Gene Expression Omnibus (GEO), we selected 5 items that met the requirements, including T1HG/T1G3 and low-grade T1 (T1LG)/ T1G2. By comparing and using the intersection to obtain DEGs, we aimed to bring DEGs into the gene sets obtained by WGCNA to identify DEGs in cancer cells and used BC cell lines to predict functional candidate markers.

Materials And Methods

GEO data

The GSE120736, GSE19915, GSE19423, GSE32548, and GSE37815 gene expression profile matrix files were obtained from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database. The platform of the GSE120736 dataset is the GPL10558 Illumina Human HT-12V4.0 expression beadchip, and this dataset contains 27 T1 high-grade BC tissues and 14 T1 low-grade BC tissues. The platform of GSE19915 is the GPL3883 Swegene Human 27K RAP UniGene188 array, and it contains 7 T1G3 BC tissues and 10 T1G2 BC tissues. The platform of GSE19423 is the GPL6102 Illumina human-6 v2.0 expression beadchip, and it contains 9 T1 high BC tissues and 39 T1 low BC tissues. The platform of

GSE32548 is the GPL6947 Illumina HumanHT-12 V3.0 expression beadchip, and it contains 38 T1G3 BC tissues and 14 T1G2 BC tissues. The platform of GSE37815 is the GPL6102 Illumina human-6 v2.0 expression beadchip, and it contains 5 T1G3 BC tissues and 13 T1G2 BC tissues.

GSE97768, RNA sequencing data for 30 BC cell lines, can be seen as the result of the sequencing of different monoclonal cancer cells. We used this as a cancer cell isolated from a complex tumour microenvironment and performed a Gene Set Enrichment Analysis (GSEA) in a single gene to identify the biological functions involved in BC cells.

Microarray data processing

GEO2R online software was used to analyse the raw data of microarray in order to recognise the DEGs. P-value < 0.05 and $|FC| \geq 1$ were used as the cut-off criteria. Subsequently, Bioinformatics & Evolutionary Genomics (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) was used to identify overlapping DEGs in five gene expression microarrays. The up-regulated and down-regulated genes were measured.

TCGA data download and pre-processing

RNA-seq and related clinical information for human bladder transitional cell carcinoma and papilloma samples were obtained from the TCGA database website (portal.gdc.cancer.gov), containing 430 tissues and 405 cases. These data were updated on August 26, 2019. RNA-seq data containing 19 normal samples and 411 cancer samples were combined into a matrix file using the Perl language merge script (www.perl.org). The gene name is then converted from Ensembl id to a matrix of gene symbols via the Ensembl database (asia.ensembl.org/index.html). Using the edgeR R package, the genes that had the same name were combined, and the genes with an average expression of more than 1 were selected and normalised.

Cell fraction of various cells in the tumour environment (TME)

EPIC (gfellerlab.shinyapps.io/EPIC_1-1/) application is designed to estimate the proportion of immune and cancer cells from the bulk tumour gene expression data[16]. This is done by fitting gene expression reference profiles from the main non-malignant cell types and simultaneously accounting for an uncharacterized cell type without prior knowledge about it (e.g. cancer cells in solid tumours samples). EPIC established reference gene expression profiles for major tumour invasive immune cell types (CD4T, CD8T, B, NK, macrophages) and further deduced reference spectra of cancer-associated fibroblasts (CAFs) and endothelial cells. We used EPIC to calculate the cell fraction of cells in TME.

WGCNA and module preservation

WGCNA was performed using the WGCNA R package[17]. Because some genes with no significant expression changes between samples were highly correlated in WGCNA, the genes with the most differential expression were used in subsequent WGCNA. Genes with the highest 25% of DEG variance were chosen, guaranteeing the heterogeneity and accuracy of bioinformatics statistics for further co-expression network analysis[18, 19]. First, RNA-seq data were filtered to reduce outliers. The co-expression similarity matrix consisted of the absolute values of the correlation between transcript expression levels. A Pearson correlation matrix was constructed for paired genes. We constructed a weighted adjacency matrix using the power function $a_{mn} = |c_{mn}|^\beta$ (c_{mn} = Pearson correlation between gene m and gene n; a_{mn} = adjacency between gene m and gene n). The parameter β emphasized a strong correlation between genes and penalized a weak correlation. Next, an appropriate β value was selected to increase the similarity matrix and achieve a scale-free co-expression network. The adjacency matrix was then converted into a topological overlap matrix (TOM), which measures the network connectivity of genes defined as the sum of adjacent genes generated by all other networks[20]. Average linkage hierarchical clustering was performed on TOM-based dissimilarity measurements, and the minimum size (genome) of the gene dendrogram was 30. Through further analysis of modules, we calculated their dissimilarity and constructed module dendrograms.

Identify cell-related modules

Gene significance (GS) was calculated to measure the correlation between genes and sample traits and determine the significance of each module. Module eigengenes were considered the main components in the principal component analysis of each gene module, and the expression patterns of all genes were summarized as a single feature expression profile within a given module. Next, GS was defined as the $\log_{10}$ conversion of the p-value in the linear regression between gene expression and clinical data ($GS = \lg P$). Module significance (MS) was defined as the average GS within the module and calculated to measure the correlation between the module and sample traits[21]. Statistical significance was determined using the relevant P values. In order to increase the capacity of the modules, we selected a cutoff value (< 0.25) to merge modules with similar heights. Here, we choose cell fraction of various cells calculated by EPIC as phenotype.

Immunohistochemistry (IHC)

In order to confirm the protein expression, immunohistochemistry data are provided on The Human Protein Atlas (www.proteinatlas.org), which aims to map all the human proteins in cells, tissues and organs[22]. After finding the pathology of urothelial carcinoma, we compared the expression of genes in cancer cells and stroma.

Expression of genes in cell lines

The Cancer Dependency Map Project at Broad Institute ([depmap portal, depmap.org/portal/](http://depmap.org/portal/)) systematically identifies genetic and pharmacologic dependencies and the predictive biomarkers. It contains 56202 gene expression data of 1201 cell lines from the Cancer Cell Line Encyclopedia (CCLE). We searched and compared gene expression between 36 urinary tract cell lines and 39 fibroblast cell lines.

Pathway, process and protein-protein interaction enrichment analysis

Metascape (metascape.org) provides a variety of functions, such as gene enrichment analysis and protein interaction network analysis[23]. For each given gene list, pathway and process enrichment analysis were carried out with the following ontology sources: KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways and CORUM. All genes in the genome were used as the enrichment background. Terms with a P-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 were collected and grouped into clusters based on their membership similarities. A subset of enriched terms was selected and rendered as a network plot, where terms with a similarity > 0.3 were connected by edges to capture the relationships between the terms further. Protein-protein interaction enrichment analysis was carried out to construct the subset of proteins that form physical interactions with at least one other member in the list. If the network contained between 3 and 500 proteins, the Molecular Complex Detection (MCODE) algorithm[24] was applied to identify densely connected network components.

DisNor (<https://disnor.uniroma2.it/>) was used to generate and investigate a protein interaction network, linking disease genes from the causal interaction information annotated in SIGNOR and protein interaction data in Mentha. The Multi-Protein Search module was selected. The first neighbour with physical interactions (score > 0.4) was chosen as the complexity level, which analysed the causal relationship between proteins.

Gene set enrichment analysis (GSEA)

Data from GSE97768 30 BC cell lines were divided into two groups according to the median expression of cancer cell candidate DEGs. GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was used to identify the potential functions correlated with key genes by assessing whether a series of previously defined biological processes were enriched in the gene rank derived from whole genes between the two groups. In the molecular signature database v6.2, curated hallmark gene sets were used. Terms with |normalised enrichment score| > 1, nominal P-value < 0.05, and false discovery rate of q-value < 0.25 were identified.

Results

Identification of DEGs

In total, 86 T1G3/T1HG samples and 90 T1G2/T1LG samples were considered in this study. The online GEO2R tool was utilized to determine the DEGs based on cut-off values: P values <0.05 and $|\log FC| \geq 1$. We obtained overlapping DEGs of the five or random four datasets, including 19 up-regulated and 13 down-regulated genes (Figure 1) to confirm the reliability of DEGs in BC.

DEGs in modules constructed by WGCNA

Although the tumour purity was high in the T1 tumour tissue, we do not think that the identified DEGs are not specifically or highly expressed in cancer cells. We used TCGA urothelial carcinoma data with high tumour heterogeneity (mainly stage ii/iii /iv) to cluster previously identified DEGs by WGCNA. After clustering, we found that DEGs were mainly concentrated in several modules (Figure 2A). Among them, the brown and green modules had the most genes, with 9 and 8 genes, respectively (Figure 2B). In order to further prove that the modules recognised by WGCNA are cell-specific expression gene sets, the different cell fractions calculated by EPIC (Table S1) were correlated to the modules obtained by WGCNA. The green module had a correlation with CD4 T cell components of more than 50% (Figure 2C). Therefore, these genes are specifically expressed in CD4 T cells, and not in cancer cells. The brown, red, sky-blue, steel-blue, and tan modules were positively correlated (Figure 2A). These modules and candidate DEGs are highly likely to be derived from BC cells and may have relatively high expression specificity (Figure 2D).

WGCNA identifying cell-specific expressing modules

After analysing the calculation principles, we thought that WGCNA was suitable for identifying cell-specifically expressed genes. In order to find genes specifically expressed in

cancer cells in pathological tissues, it was needed to demonstrate that WGCNA can mine cell-specific expression genes. WGCNA-identified modules correlated with various cellular fractions, the correlation between B cells and the cyan module reached 0.9, and the correlation between CAFs and the pink module reached 0.8 (Figure 2C). The constructed modules were highly correlated with B cells and CAFs, indicating that both B cells and CAFs have specifically expressed genes. Since there is no B cell-related cell line sequencing data, we started with fibroblasts to prove our inference. CAFs are the main components of the stroma surrounding cancer cells[25], which can easily be distinguished from cancer cells after tissue staining. Fibroblasts also have corresponding cell lines, which can be compared with urothelial cell lines to observe gene-specific expression. The correlation between GS and Module Membership (MM) values in the pink module, were as high as 0.96, and the maximum values of each were over 0.8, indicating that pink module has very specific expressed genes in CAFs (Figure 3A). Using 310 genes in the pink module (307 genes queried in the Cancer Dependency Map), the average expression levels in fibroblasts and urothelial cell lines were calculated and compared. Genes with larger MM values in the pink module were specific for fibroblasts or CAFs (Figure 3B). We selected 10 genes with the largest MM values, all greater than 0.87; SPARC, COL5A1, COL6A3, COL1A1, COL1A2, FBN1, TIMP2, CTHRC1, COL5A2, and THY1. These genes were expressed more in the BC stroma (Figure 3C), excluding COL5A2, which was not retrieved in The Human Protein Atlas. RNA expression of the 10 genes was significantly higher in fibroblasts than in urothelial cells (Figure 3D) according to the data downloaded from the Cancer Dependency Map (Table S2). The genes within a module recognised by WGCNA were highly co-expressed and specifically or relatively specifically expressed by certain types of cells. Based on this, we can identify relatively specifically expressed DEGs in cancer cells (Figure 4A).

Expression of candidate DEGs in cancer cell

The 16 candidate DEGs focus on several highly relevant modules (Figure 2A). These genes are likely to be expressed relatively specifically by cancer cells. In a comparison of IHC and the cell line RNA-seq, we confirmed the relative specific expression of these genes at the protein and RNA levels. In IHC images, the staining of urothelial cancer cells was significantly higher than the surrounding tissues or cells (Figure 4B). The comparison of urothelial cancer cells with fibroblasts (the main component in the matrix) reflected the specific degree of expression of these DEGs in cancer cells better (Figure 4C). In addition to GSDME, the other 15 genes had higher RNA expression in urothelial cancer cell lines than in surrounding tissues or cells (Table S3).

Function annotation of cancer cell candidate DEGs

The functional enrichment analysis of candidate DEGs reflects changes in the biological function of cancer cells within cancer tissues. According to Metascape results, the 16 candidate DEGs are enriched in top nine clusters with their representative enriched terms (one per cluster), containing mitotic cell cycle phase transition, cell division, cell cycle, chromosome condensation, meiotic nuclear division, Mitotic G1-G1 S phases, intrinsic apoptotic signalling pathway, cellular component disassembly, and developmental process involved in reproduction (Figure 5A). Detailed enrichment results can be found in Table S4.

After analysing the enrichment function of candidate DEGs, GSEA helped us to enrich the biological function changes caused by the high or low expression of each candidate DEGs in BC cells (GSE97768). The 12 functionally enriched genes were all up-regulated. Half or more of the DEGs were enriched in E2F targets, G2M checkpoint and mitotic spindle (Figure 5B). This result reflected the functional enrichment of each DEGs and their common enrichment functions. Both GSEA results and Metascape results were enriched in terms of mitosis and cell cycle.

Protein-protein Interaction Analysis

Based on Metascape, the MCODE networks identified for 16 candidate DEGs were gathered, and components of only one densely connected network (AURKB, CCNB2, CDC20, and CDCA5) were identified (Figure 5C). Pathway and process enrichment analysis were applied to each MCODE component, and the three best-scoring terms by P-value were retained as the functional description of the corresponding components, shown in the tables underneath corresponding network plots. In DisNor, we imported 16 candidate DEGs and found that there was no regulatory relationship or protein interaction between 11 genes identified (Figure 6).

BC and candidate markers

We searched Pubmed for studies of candidate DEGs in BC, and the role of 11 genes in BC was clarified. We summarized the 11 genes in terms of chemotherapy resistance, stem cell characteristics, progression/poor prognosis/relapse, migration/invasion/proliferation (Table 1). These 11 genes and SAPCD2 were functionally enriched in single-gene GSEA. Therefore, we defined the 12 functional genes as candidate markers (AURKB, CCNB2, CDC20, CDCA5, CKS2, E2F2, MCM2, MELK, NUSAP1, SAPCD2, TOP2A, and TRIP13) and speculated that these 12 up-regulated genes interfere with biological function signals in BC cells to jointly promote cancer progression.

Table 1. The role of 11 candidate DEGs in bladder cancer

	Gene	Chemotherapy resistance	Stem cell characteristics	Progression / poor prognosis /relapse	migration / invasion / proliferation
1	AURKB	Burgess EF[26]	Pan S[18]	Luo Y[27]	
2	CCNB2				Lei CY[28]
3	CDC20				Kidokoro T[29]
4	CDCA5		Pan S[18]	Chang IW[30]	
5	CKS2			Chen R[31], Kawakami K[32]	
6	E2F2				Hayami S[33]
7	MCM2			Burger M[34], Kruger S[35]	
8	MELK			Shi J[36]	
9	NUSAP1		Pan S[18]		
10	TOP2A			Koren R[37]	Zeng S[38]
11	TRIP13	Lu S[39]		Niu L[40]	Lu S[39]

Discussion

The high recurrence rate in early BC and low survival rate in advanced BC continue to be problems with poor scope for improvement[41]. With the occurrence and progress of tumours, the components and signal exchange in tumours will become more and more complicated, and the possibility of a cure will gradually decrease. Abnormal signals in cancer cells causing tumorigenesis and development have not been identified. Our study aimed to find DEGs in cancer cells, and the poor prognosis of early BC T1G3 is the key to solving the problem.

At present, many studies compare the differences between tissue samples and directly use DEGs as the difference between cancer cells for experiments and analysis. However, most of these DEGs may not be caused by signal differences in BC cells, nor are they relatively specific or highly expressed in BC cells. The signal changes in cancer cells are not singular, and targeted therapy targeting a single gene may not be effective. To date, no studies have identified DEGs inside BC cells. We compared T1 stage tumours with higher tumour purity, and the tumour components of T1G3/T1HG and T1G2/T1LG were very similar, but the DEGs obtained at this time could not be used as candidate genes. The candidate DEGs screened

by the WGCNA construction module are relatively specific and have high expression in the cancer cell. Most of candidate markers' functions have been verified in BC, but no known studies have been undertaken to analyse them.

WGCNA is based on the construction of scale-free networks and overlapping topological networks, clustering highly co-expressed genes into one module. The highly co-expressed genes recognised by WGCNA can also be recognised as a class of cell-specifically expressed genes. In highly heterogeneous tumours, the expression levels of these specifically expressed genes will change with the proportion of such cells, showing a high degree of co-expression. The high correlation between the cell ratio of CAFs and the pink module and the differential expression of genes within the pink module in cancer cells and CAFs confirmed our inference. Based on this inference, we continued to identify 16 candidate DEGs that are relatively specifically expressed by cancer cells.

The 12 candidate markers were relatively specific and functional in cancer cells in BC. Genes relatively specific and differentially expressed in BC cells are important to explore changes in biological functions in BC. The function of the 12 candidate markers was related to mitosis and the cell cycle (Fig. 5B), suggesting a close relationship with cancer progression. The 12 markers had similar functional enrichment and were related to BC progression or drug resistance in previous studies, excluding SAPCD2 (Table 1). There may be protein-protein interactions between AURKB, CCNB2, CDC20, and CDCA5 (Fig. 5C). However, according to the results in DisNor, there is no direct relationship between them (Fig. 6), which means that it is difficult to control BC by targeting a single gene effectively, and the relationship between them should be clarified and then combined with targeted therapy.

This study takes advantage of the characteristics of different clinical stages of BC to find out the DEGs inside BC cells with a new perspective, based on the characteristics of WGCNA. Next, data from 30 BC cell lines with monoclonal characteristics were used to predict biologically functional DEGs as candidate markers, excluding interference from TME. Through clever design, we can find out the relatively specific differentially expressed functional genes in BC cells. Perhaps, this can start the era of joint targeted therapy for BC. Such an approach may also provide some help for other cancers. However, we did not further explore the relationship between these 12 genes through experiments and needed to study this further.

Conclusions

Our study aimed to provide insight into the difficult problem of T1G3 BC. We identified relatively specific differentially expressed functional genes in BC cells. Our results can be used as a foundation for exploring the biological function changes in BC cells and in the development of combined targeted therapy for T1G3 BC. WGCNA can identify cell-specific or relatively specifically expressed gene sets, which may provide new perspectives to the research of various cancers. However, in the future, we need to clarify the relationship of these genes and which genes are jointly targeted to reduce the recurrence and progression of BC.

Abbreviations

AURKB

Aurora Kinase B

BC

Bladder cancer

BCG

Bacillus Calmette-Guerin

CAFs

Cancer-associated fibroblasts

CCLE

Cancer Cell Line Encyclopedia

CCNB2

Cyclin B2

CDC20

Cell Division Cycle 20

CDCA5

Cell Division Cycle Associated 5

CKS2

CDC28 Protein Kinase Regulatory Subunit 2

COL17A1

Collagen Type XVII Alpha 1 Chain

CTSV

Cathepsin V

DEGs

Differentially expressed genes

E2F2

E2F Transcription Factor 2

GEO

Gene Expression Omnibus

GS

Gene significance

GSDME

Gasdermin E

GSEA

Gene Set Enrichment Analysis

IHC

Immunohistochemistry

MCM2

Minichromosome Maintenance Complex Component 2

MCODE
Molecular Complex Detection
MELK
Maternal Embryonic Leucine Zipper Kinase
MM
Module Membership
MS
Module significance
NMIBC
Non-muscle-invasive bladder cancer
NUSAP1
Nucleolar And Spindle Associated Protein 1
qPCR
Quantitative polymerase chain reaction
SAPCD2
Suppressor APC Domain Containing 2
SOX15
SRY-Box Transcription Factor 15
T1HG
High-grade T1
T1LG
Low-grade T1
TCGA
The Cancer Genome Atlas
TME
Tumour environment
TOM
Topological overlap matrix
TOP2A
DNA Topoisomerase II Alpha
TRIP13
Thyroid Hormone Receptor Interactor 13
TUR
Transurethral resection
WGCNA
Weighted gene co-expression network analysis

Declarations

Acknowledgements

Not applicable

Authors' contributions

BL proposed the study concept and design, and drafted the manuscript. BL and SP collected, analysed, and interpreted the data. YZ, XC, and BW participated in revising the manuscript. All authors read and approved the final manuscript.

Funding

The present study was supported by 345 Talent Project (Shengjing Hospital of China Medical University).

Availability of data and materials

The datasets used and/or analysed during the current study are available from GEO (www.ncbi.nlm.nih.gov/geo/) and TCGA (portal.gdc.cancer.gov).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Antoni, S., J. Ferlay, I. Soerjomataram, et al. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol.* 2017;71:96-108. doi: 10.1016/j.eururo.2016.06.010.

2. Kwan, M.L., L.H. Kushi, K.N. Danforth, et al. The Be-Well Study: a prospective cohort study of lifestyle and genetic factors to reduce the risk of recurrence and progression of non-muscle-invasive bladder cancer. *Cancer Causes Control*. 2019;30:187-193. doi: 10.1007/s10552-019-1130-8.
3. Jordan, B. and J.J. Meeks. T1 bladder cancer: current considerations for diagnosis and management. *Nat Rev Urol*. 2019;16:23-34. doi: 10.1038/s41585-018-0105-y.
4. Klaassen, Z., A.M. Kamat, W. Kassouf, et al. Treatment Strategy for Newly Diagnosed T1 High-grade Bladder Urothelial Carcinoma: New Insights and Updated Recommendations. *Eur Urol*. 2018;74:597-608. doi: 10.1016/j.eururo.2018.06.024.
5. Ayati, M., E. Amini, R. Shahrokhi Damavand, et al. Second Transurethral Resection of Bladder Tumor: Is it Necessary in All T1 and/or High-Grade Tumors? *Urol J*. 2018. doi: 10.22037/uj.v0i0.4670.
6. Palou, J., F. Pisano, R. Sylvester, et al. Recurrence, progression and cancer-specific mortality according to stage at re-TUR in T1G3 bladder cancer patients treated with BCG: not as bad as previously thought. *World J Urol*. 2018;36:1621-1627. doi: 10.1007/s00345-018-2299-2.
7. Mostafid, H., J. Palou, M. Burger, et al. T1 High-grade Bladder Cancer: The Search for the Optimal Management Continues. *Eur Urol*. 2018;74:609-610. doi: 10.1016/j.eururo.2018.08.016.
8. van der Heijden, A.G., L. Mengual, J.J. Lozano, et al. A five-gene expression signature to predict progression in T1G3 bladder cancer. *Eur J Cancer*. 2016;64:127-36. doi: 10.1016/j.ejca.2016.06.003.
9. Raspollini, M.R., R.J. Luque, C.L. Menendez, et al. T1 high-grade bladder carcinoma outcome: the role of p16, topoisomerase-IIalpha, survivin, and E-cadherin. *Hum Pathol*. 2016;57:78-84. doi: 10.1016/j.humpath.2016.06.022.
10. Nicolazzo, C., G.M. Busetto, A. Gradilone, et al. Circulating Tumor Cells Identify Patients with Super-High-Risk Non-Muscle-Invasive Bladder Cancer: Updated Outcome Analysis of a Prospective Single-Center Trial. *Oncologist*. 2019. doi: 10.1634/theoncologist.2018-0784.
11. Siracusano, S., B. Niccolini, R. Knez, et al. The simultaneous use of telomerase, cytokeratin 20 and CD4 for bladder cancer detection in urine. *Eur Urol*. 2005;47:327-33. doi: 10.1016/j.eururo.2004.10.007.
12. Giulietti, M., G. Occhipinti, A. Righetti, et al. Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data. *Front Oncol*. 2018;8:450. doi: 10.3389/fonc.2018.00450.
13. Gao, X., Y. Chen, M. Chen, et al. Identification of key candidate genes and biological pathways in bladder cancer. *PeerJ*. 2018;6:e6036. doi: 10.7717/peerj.6036.
14. Hu, J., L. Zhou, Z. Song, et al. The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets. *J Cell Physiol*. 2019. doi: 10.1002/jcp.28208.
15. Liu, Y., X. Wu, G. Wang, et al. CALD1, CNN1, and TAGLN identified as potential prognostic molecular markers of bladder cancer by bioinformatics analysis. *Medicine (Baltimore)*. 2019;98:e13847. doi: 10.1097/md.00000000000013847.
16. Racle, J., K. de Jonge, P. Baumgaertner, et al. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. 2017;6. doi: 10.7554/eLife.26476.

17. Langfelder, P. and S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi: 10.1186/1471-2105-9-559.
18. Pan, S., Y. Zhan, X. Chen, et al. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. *Front Oncol*. 2019;9:613. doi: 10.3389/fonc.2019.00613.
19. Pan, S., Y. Zhan, X. Chen, et al. Bladder Cancer Exhibiting High Immune Infiltration Shows the Lowest Response Rate to Immune Checkpoint Inhibitors. *Front Oncol*. 2019;9:1101. doi: 10.3389/fonc.2019.01101.
20. Botia, J.A., J. Vandrovcova, P. Forabosco, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol*. 2017;11:47. doi: 10.1186/s12918-017-0420-6.
21. Fuller, T.F., A. Ghazalpour, J.E. Aten, et al. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome*. 2007;18:463-72. doi: 10.1007/s00335-007-9043-3.
22. Uhlen, M., C. Zhang, S. Lee, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357. doi: 10.1126/science.aan2507.
23. Zhou, Y., B. Zhou, L. Pache, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523. doi: 10.1038/s41467-019-09234-6.
24. Bader, G.D. and C.W. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2.
25. Erdogan, B. and D.J. Webb. Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis. *Biochem Soc Trans*. 2017;45:229-236. doi: 10.1042/bst20160387.
26. Burgess, E.F., C. Livasy, S. Trufan, et al. High aurora kinase expression identifies patients with muscle-invasive bladder cancer who have poor survival after neoadjuvant chemotherapy. *Urol Oncol*. 2019;37:900-906. doi: 10.1016/j.urolonc.2019.09.009.
27. Luo, Y., M. Barrios-Rodiles, G.D. Gupta, et al. Atypical function of a centrosomal module in WNT signalling drives contextual cancer cell motility. *Nat Commun*. 2019;10:2356. doi: 10.1038/s41467-019-10241-w.
28. Lei, C.Y., W. Wang, Y.T. Zhu, et al. The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer. *Urol Oncol*. 2016;34:237.e1-10. doi: 10.1016/j.urolonc.2015.11.011.
29. Kidokoro, T., C. Tanikawa, Y. Furukawa, et al. CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene*. 2008;27:1562-71. doi: 10.1038/sj.onc.1210799.
30. Chang, I.W., V.C. Lin, H.L. He, et al. CDCA5 overexpression is an indicator of poor prognosis in patients with urothelial carcinomas of the upper urinary tract and urinary bladder. *Am J Transl Res*. 2015;7:710-22.
31. Chen, R., C. Feng, and Y. Xu. Cyclin-dependent kinase-associated protein Cks2 is associated with bladder cancer progression. *J Int Med Res*. 2011;39:533-40. doi: 10.1177/147323001103900222.

32. Kawakami, K., H. Enokida, T. Tachiwada, et al. Identification of differentially expressed genes in human bladder cancer through genome-wide gene expression profiling. *Oncol Rep.* 2006;16:521-31.
33. Hayami, S., M. Yoshimatsu, A. Veerakumarasivam, et al. Overexpression of the JmJc histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol Cancer.* 2010;9:59. doi: 10.1186/1476-4598-9-59.
34. Burger, M., S. Denzinger, A. Hartmann, et al. Mcm2 predicts recurrence hazard in stage Ta/T1 bladder cancer more accurately than CK20, Ki67 and histological grade. *Br J Cancer.* 2007;96:1711-5. doi: 10.1038/sj.bjc.6603784.
35. Kruger, S., C. Thorns, W. Stocker, et al. Prognostic value of MCM2 immunoreactivity in stage T1 transitional cell carcinoma of the bladder. *Eur Urol.* 2003;43:138-45. doi: 10.1016/s0302-2838(02)00580-8.
36. Shi, J., P. Zhang, L. Liu, et al. Weighted gene coexpression network analysis identifies a new biomarker of CENPF for prediction disease prognosis and progression in nonmuscle invasive bladder cancer. *Mol Genet Genomic Med.* 2019;7:e982. doi: 10.1002/mgg3.982.
37. Koren, R., V. Kugel, Y. Dekel, et al. Human DNA topoisomerase-IIalpha expression as a prognostic factor for transitional cell carcinoma of the urinary bladder. *BJU Int.* 2003;91:489-92. doi: 10.1046/j.1464-410x.2003.04118.x.
38. Zeng, S., A. Liu, L. Dai, et al. Prognostic value of TOP2A in bladder urothelial carcinoma and potential molecular mechanisms. *BMC Cancer.* 2019;19:604. doi: 10.1186/s12885-019-5814-y.
39. Lu, S., M. Guo, Z. Fan, et al. Elevated TRIP13 drives cell proliferation and drug resistance in bladder cancer. *Am J Transl Res.* 2019;11:4397-4410.
40. Niu, L., Z. Gao, Y. Cui, et al. Thyroid Receptor-Interacting Protein 13 is Correlated with Progression and Poor Prognosis in Bladder Cancer. *Med Sci Monit.* 2019;25:6660-6668. doi: 10.12659/msm.917112.
41. Kamat, A.M., N.M. Hahn, J.A. Efsthathiou, et al. Bladder cancer. *Lancet.* 2016;388:2796-2810. doi: 10.1016/s0140-6736(16)30512-8.

Figures

19	Upregulated genes
3	RARRES1, TOP2A, TRIP13
7	AURKB, CCNB2, CDC20, CDCA5, E2F2, MCM2, NUSAP1
5	CD36, CKS2, CTSV, GGH, MELK
3	HIST1H2BD, HIST1H2BK, PYCR1
1	SAPCD2

13	Downregulated genes
2	BTBD16, GSDME
4	ALDH1L1, CRTAC1, NSG1, SOX15
3	BMP7, CAPN8, MST1R
2	CFH, TESC
2	COL17A1, CTSE

Figure 2

Overlapping DEGs of the five or random four datasets.

Figure 4

Candidate DEGs in cancer cells. (A) Eigengene dendrogram and Eigengene adjacency heatmap. (B) Distribution of up- and down-regulated genes within the module. (C) Correlation between WGCNA building modules and cell fraction. Modules containing DEGs and related correlation parameters are marked with black boxes. (D) Candidate DEGs in the bladder cancer cell.

Figure 6

The Pink module is a specific expression module of CAFs. (A) Correlation with MM and GS related to CAFs in Pink module. (B) The higher the MM value, the greater the difference between fibroblasts and urothelial cancer cell lines. (C) Immunohistochemical images of 10 genes (excluding COL5A2) with a maximum MM value in the pink module. The stroma surrounding the cancer cells was stained more

deeply. (D) 10 genes with the highest MM values in the pink module have the highest expression in fibroblasts and much lower expression in urinary tract cell lines.

A

B

C

Figure 8

Relative specific expression of candidate DEGs in cancer cells. (A) Recognition of candidate DEGs in bladder cancer cells requires cell-specific expression modules constructed by WGCNA. (B) Immunohistochemical images of 16 candidate DEGs in bladder cancer. Staining of cancer cells is higher

than the surrounding stroma. (C) In 16 candidate DEGs excluding GSDME, the expression was higher in urothelial cells than in fibroblasts. Statistics using T-test, all $P < 0.001$.

Figure 10

Function and protein-protein interaction in 16 candidate DEGs. (A) The top 9 clusters with enriched representative terms (one per cluster) of 16 candidate DEGs. (B)Single-gene GSEA predicts that 12 up-

regulated candidate DEGs have a function in bladder cancer cells. (C) AURKB, CCNB2, CDC20 and CDCA5 may have protein-protein interaction relationship.

Figure 12

Entering 16 genes in DisNor, only 11 genes were identified. There is no direct relationship between them.

Supplementary Files

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

- [ExpressionPublic19Q3of16candidateDEGs.xls](#)
- [ExpressionPublic19Q3of16candidateDEGs.xls](#)
- [EPICcellfraction.csv](#)
- [ExpressionPublic19Q3of10genes.xls](#)
- [GOAllLists.csv](#)

- [EPICcellfraction.csv](#)
- [ExpressionPublic19Q3of10genes.xls](#)
- [GOAllLists.csv](#)