

# Bioinformatics Analysis Identifying Key Biomarkers in Bladder Cancer

Chuan Zhang , Mandy Berndt-Paetz and Jochen Neuhaus \* 

Department of Urology, University of Leipzig, 04103 Leipzig, Germany;  
Chuan.Zhang@medizin.uni-leipzig.de (C.Z.); Mandy.Berndt@medizin.uni-leipzig.de (M.B.-P.)

\* Correspondence: jochen.neuhaus@medizin.uni-leipzig.de; Tel.: +49-341-9717688

Received: 13 March 2020; Accepted: 10 April 2020; Published: 16 April 2020



**Abstract:** Our goal was to find new diagnostic and prognostic biomarkers in bladder cancer (BCa), and to predict molecular mechanisms and processes involved in BCa development and progression. Notably, the data collection is an inevitable step and time-consuming work. Furthermore, identification of the complementary results and considerable literature retrieval were requested. Here, we provide detailed information of the used datasets, the study design, and on data mining. We analyzed differentially expressed genes (DEGs) in the different datasets and the most important hub genes were retrieved. We report on the meta-data information of the population, such as gender, race, tumor stage, and the expression levels of the hub genes. We include comprehensive information about the gene ontology (GO) enrichment analyses and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. We also retrieved information about the up- and down-regulation of genes. All in all, the presented datasets can be used to evaluate potential biomarkers and to predict the performance of different preclinical biomarkers in BCa.

**Dataset:** The following are available online at <http://www.mdpi.com/2306-5729/5/2/38/s1>

**Dataset License:** CC-BY-NC

**Keywords:** bioinformatics; bladder cancer; biomarker; data

## 1. Summary

Bladder cancer is one of the most common malignancies [1]. Although new treatment strategies and tools for surgical resection [2], neoadjuvant chemotherapy [3,4], and photodynamic therapy (PDT) [5] have been developed, BCa remains with a high rate of recurrence [1]. Up to now, cystoscopy and bioptic histology are still the gold standards to diagnose bladder cancer (BCa) [6]. No consent about urinary marker or non-invasive screening strategies could be found by the European Association of Urology (EUA). Furthermore, nuclear matrix protein 22 (NMP22) is only recommended by American Urological Association (AUA) under certain conditions [7]. Therefore, it remains a priority to develop reliable, safe, and non-invasive diagnostic/prognostic biomarkers and therapeutic targets for BCa, and considerable efforts are ongoing.

Bioinformatics analysis is necessary for the integration of, e.g., huge amounts of transcriptome, microarray, and RNA-sequencing data to disclose alterations in gene expression, mutational burden, transcriptome, and proteome of cancer compared to non-cancer controls [8]. Of special importance are so-called hub genes, defined as highly connected genes, which can be regarded as a representative of a distinct module in the genes network [9]. Furthermore, hub genes potentially play an important role in the progression of cancer. Therefore, they are good biomarker candidates and may even provide new therapeutic targets [8,10]. Following, we provide supplemental results from our recent investigation

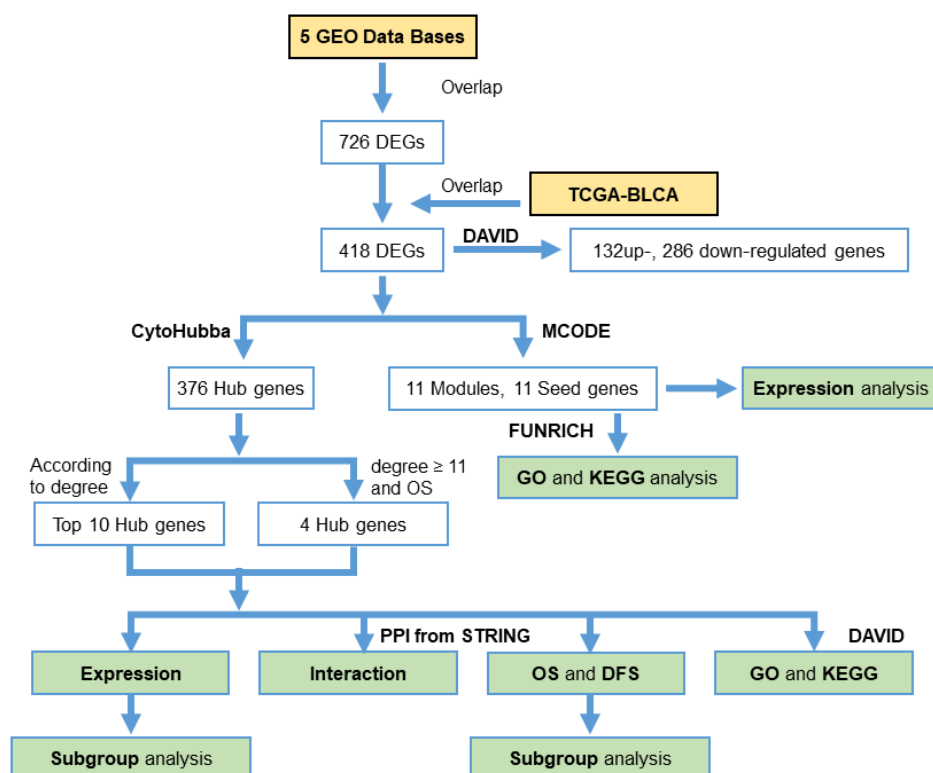
“Identification of key biomarkers in bladder cancer: Evidence from bioinformatics analysis” [11], which we share in a freely available and searchable form for further preclinical research. All in all, the data is the supplement for the previous research, providing selected data from the public The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases for cancer researchers and bioinformaticians to search and download for analysis based on the requisition of the public GDC Data Portal (<https://portal.gdc.cancer.gov/>).

## 2. Data Description

### 2.1. Database Analyses

Our previous study “Zhang et al. 2020, Identification of key biomarkers in bladder cancer: Evidence from bioinformatics analysis” [11] integrated six public datasets. Five datasets were downloaded from GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo>) [12] and the bladder cancer (TCGA-BLCA) dataset was retrieved from TCGA (The Cancer Genome Atlas, <https://portal.gdc.cancer.gov/>) [13]. Below, we detailed the study design and the characteristics of this recent study.

We compared BCa samples and non-cancerous samples (histologically normal tissue adjacent to the tumor) and extracted six sets of differentially expressed genes (DEGs) from six public datasets according to the criteria for defining DEGs [11]. We used FUNRICH software [14] to map the DEGs and detected the overlap between the datasets (Table 1 and Table S1). We considered only DEGs, which were expressed in at least two of the five GEO databases (Table S2). Finally, we identified 418 DEGs (Table S3) overlapping with the 2537 DEGs from the TCGA-BLCA database (Table S3). Of those DEGs, 132 were upregulated and 286 were downregulated (Figure 1).



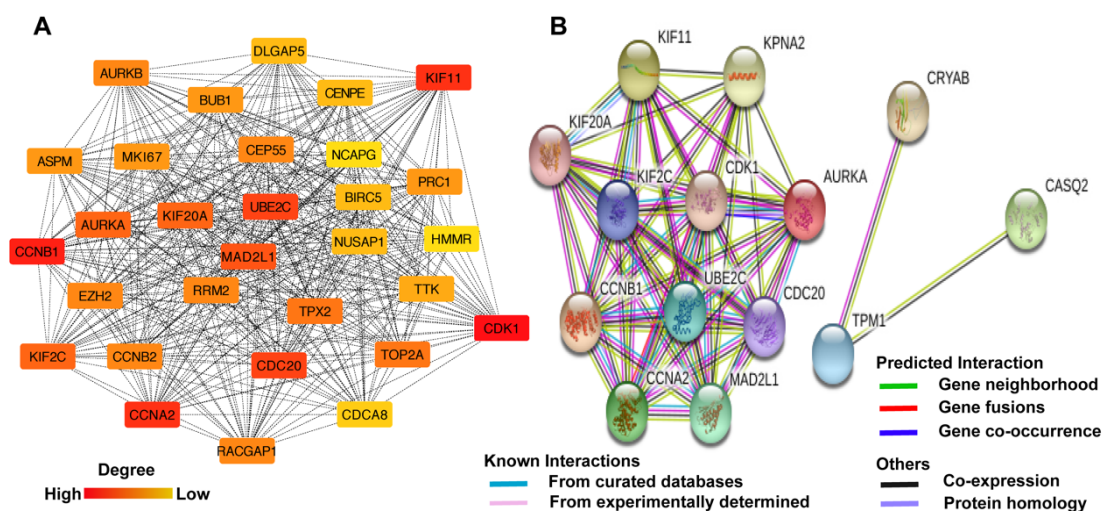
**Figure 1.** Study design and workflow of the analysis. Taken from Zhang et al. 2019 [11]. For detailed information on the analytical methods and the software packages please refer to the Materials and Methods section in the cited paper.

Molecular complex detection (MCODE) [15–17] analysis identified 11 relevant modules (subnetworks) and cytoHubba (based on Cytoscape software) [18] classified 376 of those 418 DEGs as hub genes, which are the gene most interconnected in the networks/modules (Table S4) [8].

To reduce the hub genes to the most promising, we defined 11 seed genes [15,16] for the most important 11 modules and the subsequent analysis yielded 14 hub genes on the basis of correlation to overall survival and degree of interaction (Table S5). The hub genes were ordered by their descending interaction degree: *CDK1*(98), *CCNB1*(92), *CCNA2*(84), *KIF11*(84), *CDC20*(83), *UBE2C*(83), *MAD2L1*(81), *AURKA*(80), *KIF20A*(80), *KIF2C*(80), *KPNA2*(67), *TPM1*(29), *CASQ2*(11), and *CRYAB*(11). Figure 2 depicts the results of the protein–protein interaction (PPI) analysis.

**Table 1.** General information on Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) datasets (adapted from: Zhang et al. 2020 [11]).

Dataset	Number of Noncancerous Bladder Tissue Samples	Number of Cancer Tissue Samples	Number of DEGs Extracted from Dataset	Number of DEGs after FUNRICH Mapping
GSE27448 [19,20]	5	10	5251	4701
GSE52519 [21]	3	9	751	742
GSE61615 [22]	2	2	842	736
GSE76211 [23,24]	3	3	770	658
GSE100926 [25]	3	3	223	194
TCGA-BLCA [13]	19	406	2873	2537



**Figure 2.** Interaction network analysis. (A) The interactions and protein–protein networks of the top 30 hub genes. (B) Protein–protein network and interaction among the 14 hub genes from STRING-db.org, accessed on 11 November 2019. Nodes with different colors represent different query proteins. A different color on the edge means a different interaction (see legend in figure). Adapted from Figure 4 of Zhang et al. 2020 [11].

We then performed a gene ontology (GO) enrichment analysis [26] and used the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [27] to identify the pathways potentially related to these protein-coding genes and to predict the roles of these genes in BCa. The Go and KEGG analyses significantly enriched the ‘Pathways in cancers’, ‘Viral carcinogenesis’, and ‘Cell cycle’, which relate to carcinogenesis and progression of cancer. We listed all significant results of GO and KEGG analysis based on the 418 DEGs and Benjamini–Hochberg value <0.05 in Table S6. The GO and KEGG analysis results based on the 14 hub genes are available and searchable at <http://www.mdpi.com/2075-4418/10/2/66/s1>, Table S5. Moreover, we also listed all significant results of GO and KEGG analyses of up- (Table S7) and downregulated DEGs (Table S8). In addition, we

extracted the clinical meta-data from TCGA-BLCA for the correlation to overall survival (OS) and disease-free survival (DFS) based on the 14 hub genes (Table S5).

In addition, we performed other subgroup analyses, such as the expression levels in different groups based on tumor stage, lymph nodal metastasis, race of patients, gender of patients, histological subtype, and molecular subtypes. We found that *CKD1*, *CCNB1*, *CCNA2*, *KIF11*, *CDC20*, *UBE2C*, *MD2L1*, *AURKA*, *KIF20A*, *KIF2C*, *KPNA2*, *TPM1*, *CASQ2*, and *CRYAB* were significantly higher expression in Caucasian and African American than in the ASI cohort. Except for *TPM1*, *CASQ2*, and *CRYAB*, all the genes were significantly overexpressed in both, male and female bladder cancer patients. However, no significant difference was found between males and females. All the genes were significantly higher expressed in non-papillary tumors than in papillary tumors (Table 2) [28]. In addition, *CKD1*, *CCNB1*, *CCNA2*, *KIF11*, *CDC20*, *UBE2C*, *MD2L1*, *AURKA*, *KIF20A*, *KIF2C*, and *KPNA2* were significantly upregulated in papillary tumors and non-papillary tumors than in non-cancerous tissues; in contrast, *TPM1*, *CASQ2*, and *CRYAB* were significantly downregulated in papillary tumors and non-papillary tumors than in non-cancerous tissues. Intriguingly, except for *CRYAB*, we found that *TPM1* and *CASQ2* were most significantly downregulated in “Luminal Papillary” tumors, while the other genes were most significantly upregulated in subtype of “Neuronal” and “Basal squamous” based on molecular subtyping (Table 2).

## 2.2. Literature Research

We retrieved seven bioinformatics studies for BCa biomarkers based on public database analyses [9,29–34], and we compared the biomarkers in the present study with the ones that were reported in the literature studies. We found that *CRYAB* and *CASQ2* were so far unrecognized as biomarkers in previous studies. On the basis of Oncomine meta-analysis (<https://www.oncomine.org/>), we here present the meta-analysis of the expression levels of the hub genes described, but not been shown in our previous study (Figure 3 and Table S9). The genes compared in the meta-analysis were *CCNB1*, *CCNA2*, *KIF11*, *CDC20*, *UBE2C*, *MAD2L1*, *AURKA*, *KIF2C*, *CASQ2*, *CRYAB*, and *KIF20A*. Furthermore, except for the *CRYAB* and *CASQ2*, which have been shown in our previous paper, we constructed the expression body maps of the other 12 hub genes reported in our previous study using GEPIA (<http://gepia.cancer-pku.cn>, accessed on Nov.11. 2019). Body maps are an impressive way to visualize the differences in gene expression between normal and tumor tissues (Figure 4). Ultimately, our research results were roughly in line with the majority of the retrieved studies.

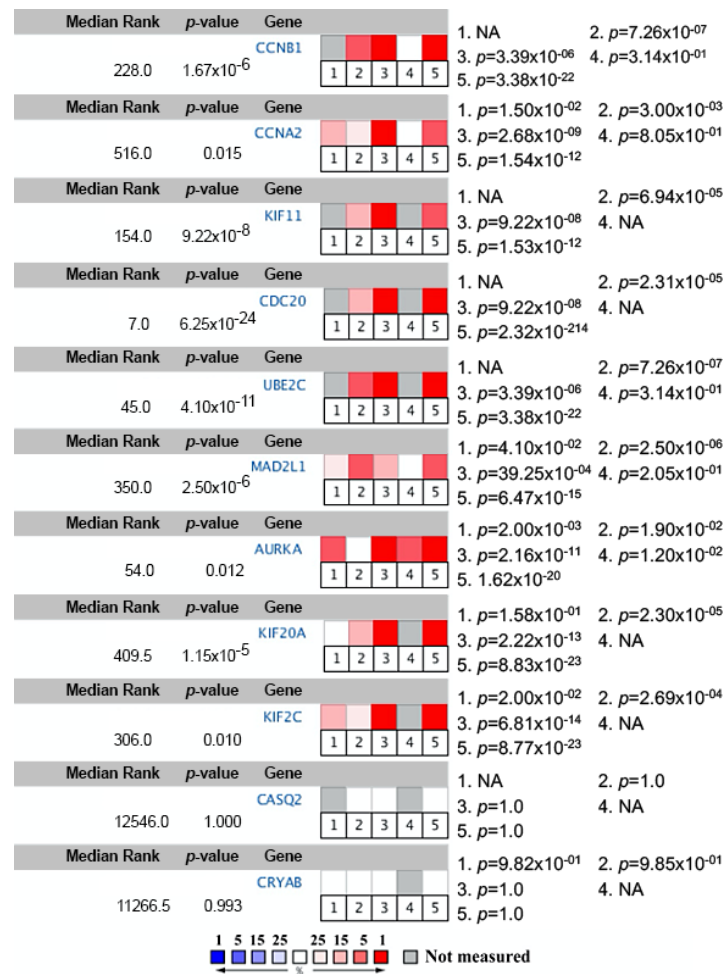
**Table 2.** Statistically significant difference of expression levels of target genes in subgroups.

Target Genes	Race of Patients		Gender of Patients		Histological Subtypes		Molecular Subtypes	
<i>CDK1</i>	CAU (↑) vs. ASI	$p = 8.829 \times 10^{-4}$	N vs. M (↑)	$p < 1.000 \times 10^{-12}$	N vs. PT (↑)	$p = 2.109 \times 10^{-15}$	N vs. NET (↑)	$p = 5.922 \times 10^{-8}$
	AFA (↑) vs. ASI	$p = 3.379 \times 10^{-3}$	N vs. F (↑)	$p = 1.554 \times 10^{-15}$	N vs. NPT (↑)	$p = 1.624 \times 10^{-12}$	N vs. BST (↑)	$p = 1.624 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 3.872 \times 10^{-2}$	N vs. LT (↑)	$p = 5.809 \times 10^{-7}$
							N vs. LIT (↑)	$p = 2.824 \times 10^{-12}$
							N vs. LPT (↑)	$p = 1.625 \times 10^{-12}$
<i>CCNB1</i>	CAU (↑) vs. ASI	$p = 7.479 \times 10^{-7}$	N vs. M (↑)	$p = 5.311 \times 10^{-13}$	N vs. PT (↑)	$p = 2.949 \times 10^{-10}$	N vs. NET (↑)	$p = 1.620 \times 10^{-5}$
	AFA (↑) vs. ASI	$p = 5.221 \times 10^{-4}$	N vs. F (↑)	$p = 2.907 \times 10^{-12}$	N vs. NPT (↑)	$p = 2.631 \times 10^{-14}$	N vs. BST (↑)	$p = 1.624 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 2.944 \times 10^{-3}$	N vs. LT (↑)	$p = 2.202 \times 10^{-4}$
							N vs. LIT (↑)	$p = 4.952 \times 10^{-7}$
							N vs. LPT (↑)	$p = 3.502 \times 10^{-9}$
<i>CCNA2</i>	CAU (↑) vs. ASI	$p = 5.447 \times 10^{-8}$	N vs. M (↑)	$p = 5.311 \times 10^{-9}$	N vs. PT (↑)	$p = 4.309 \times 10^{-7}$	N vs. NET (↑)	$p = 5.481 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 9.961 \times 10^{-4}$	N vs. F (↑)	$p = 3.062 \times 10^{-8}$	N vs. NPT (↑)	$p = 4.395 \times 10^{-10}$	N vs. BST (↑)	$p = 1.863 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 6.169 \times 10^{-4}$	N vs. LT (↑)	$p = 2.176 \times 10^{-4}$
							N vs. LIT (↑)	$p = 6.473 \times 10^{-5}$
							N vs. LPT (↑)	$p = 4.175 \times 10^{-6}$
<i>KIF11</i>	CAU (↑) vs. ASI	$p = 5.620 \times 10^{-7}$	N vs. M (↑)	$p = 5.836 \times 10^{-9}$	N vs. PT (↑)	$p = 8.989 \times 10^{-8}$	N vs. NET (↑)	$p = 4.693 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 1.065 \times 10^{-6}$	N vs. F (↑)	$p = 1.036 \times 10^{-7}$	N vs. NPT (↑)	$p = 9.950 \times 10^{-10}$	N vs. BST (↑)	$p = 2.290 \times 10^{-13}$
					PT vs. NPT (↑)	$p = 6.997 \times 10^{-3}$	N vs. LT (↑)	$p = 1.922 \times 10^{-5}$
							N vs. LIT (↑)	$p = 4.399 \times 10^{-5}$
							N vs. LPT (↑)	$p = 5.848 \times 10^{-7}$
<i>CDC20</i>	CAU (↑) vs. ASI	$p = 5.038 \times 10^{-3}$	N vs. M (↑)	$p < 1.000 \times 10^{-12}$	N vs. PT (↑)	$p = 1.691 \times 10^{-12}$	N vs. NET	$p = 2.644 \times 10^{-8}$
	AFA vs. ASI	$p = 3.077 \times 10^{-3}$	N vs. F (↑)	$p < 1.000 \times 10^{-12}$	N vs. NPT (↑)	$p < 1.000 \times 10^{-12}$	N vs. BST	$p < 1.000 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 9.644 \times 10^{-5}$	N vs. LT	$p = 8.558 \times 10^{-8}$
							N vs. LIT	$p = 2.198 \times 10^{-11}$
							N vs. LPT	$p = 3.194 \times 10^{-14}$
<i>UBE2C</i>	CAU (↑) vs. ASI	$p = 6.099 \times 10^{-3}$	N vs. M (↑)	$p < 1.000 \times 10^{-12}$	N vs. PT (↑)	$p = 1.624 \times 10^{-12}$	N vs. NET (↑)	$p = 1.045 \times 10^{-8}$
			N vs. F (↑)	$p < 1.000 \times 10^{-12}$	N vs. NPT (↑)	$p = 1.624 \times 10^{-12}$	N vs. BST (↑)	$p = 1.624 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 1.429 \times 10^{-2}$	N vs. LT (↑)	$p = 1.664 \times 10^{-10}$
							N vs. LIT (↑)	$p = 1.497 \times 10^{-9}$
							N vs. LPT (↑)	$p = 1.624 \times 10^{-12}$
<i>MAD2L1</i>	CAU (↑) vs. ASI	$p = 1.627 \times 10^{-7}$	N vs. M (↑)	$p = 4.241 \times 10^{-14}$	N vs. PT (↑)	$p = 1.634 \times 10^{-10}$	N vs. NET (↑)	$p = 4.815 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 4.381 \times 10^{-4}$	N vs. F (↑)	$p = 1.488 \times 10^{-12}$	N vs. NPT (↑)	$p = 1.625 \times 10^{-12}$	N vs. BST (↑)	$p < 1.000 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 3.153 \times 10^{-4}$	N vs. LT (↑)	$p = 1.755 \times 10^{-6}$
							N vs. LIT (↑)	$p = 1.364 \times 10^{-7}$
							N vs. LPT (↑)	$p = 1.625 \times 10^{-10}$

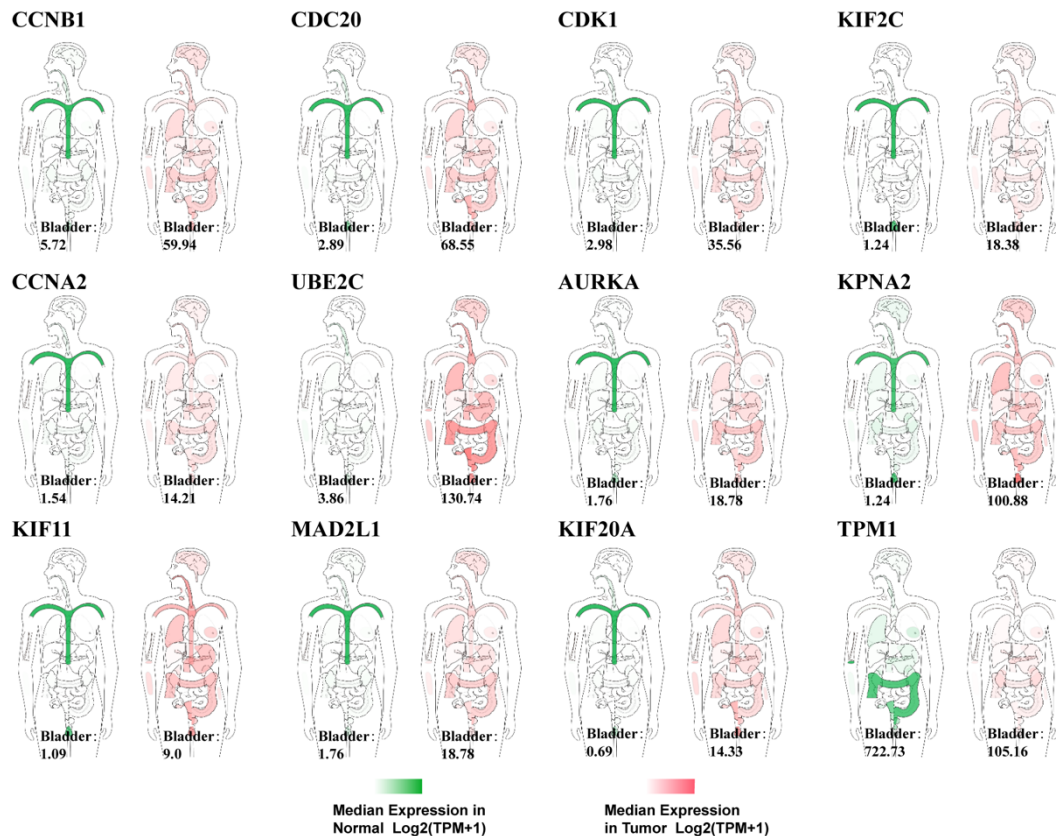
Table 2. Cont.

Target Genes	Race of Patients		Gender of Patients		Histological Subtypes		Molecular Subtypes	
<i>AURKA</i>	CAU (↑) vs. ASI	$p = 1.565 \times 10^{-7}$	N vs. M (↑)	$p < 1.000 \times 10^{-12}$	N vs. PT (↑)	$p = 6.550 \times 10^{-15}$	N vs. NET (↑)	$p = 1.539 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 1.629 \times 10^{-4}$	N vs. F (↑)	$p < 1.000 \times 10^{-12}$	N vs. NPT (↑)	$p < 1.000 \times 10^{-12}$	N vs. BST (↑)	$p < 1.000 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 2.996 \times 10^{-3}$	N vs. LT (↑)	$p = 2.259 \times 10^{-10}$
							N vs. LIT (↑)	$p = 1.487 \times 10^{-11}$
							N vs. LPT (↑)	$p = 1.674 \times 10^{-12}$
<i>KIF20A</i>	CAU (↑) vs. ASI	$p = 6.312 \times 10^{-7}$	N vs. M (↑)	$p = 4.398 \times 10^{-8}$	N vs. PT (↑)	$p = 2.500 \times 10^{-7}$	N vs. NET (↑)	$p = 4.343 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 1.179 \times 10^{-4}$	N vs. F (↑)	$p = 1.725 \times 10^{-6}$	N vs. NPT (↑)	$p = 7.345 \times 10^{-9}$	N vs. BST (↑)	$p = 1.373 \times 10^{-9}$
							N vs. LT (↑)	$p = 7.949 \times 10^{-5}$
							N vs. LIT (↑)	$p = 8.399 \times 10^{-4}$
							N vs. LPT (↑)	$p = 7.846 \times 10^{-7}$
<i>KIF2C</i>	CAU (↑) vs. ASI	$p = 1.844 \times 10^{-5}$	N vs. M (↑)	$p = 1.624 \times 10^{-12}$	N vs. PT (↑)	$p = 7.327 \times 10^{-15}$	N vs. NET (↑)	$p = 1.387 \times 10^{-7}$
	AFA (↑) vs. ASI	$p = 1.828 \times 10^{-5}$	N vs. F (↑)	$p < 1.000 \times 10^{-12}$	N vs. NPT (↑)	$p = 1.624 \times 10^{-12}$	N vs. BST (↑)	$p = 1.624 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 2.143 \times 10^{-3}$	N vs. LT (↑)	$p = 1.563 \times 10^{-10}$
							N vs. LIT (↑)	$p = 4.494 \times 10^{-13}$
							N vs. LPT (↑)	$p = 1.625 \times 10^{-12}$
<i>KPNA2</i>	CAU (↑) vs. ASI	$p = 2.028 \times 10^{-8}$	N vs. M (↑)	$p = 1.364 \times 10^{-11}$	N vs. PT (↑)	$p = 2.543 \times 10^{-9}$	N vs. NET (↑)	$p = 1.269 \times 10^{-8}$
	AFA (↑) vs. ASI	$p = 1.019 \times 10^{-4}$	N vs. F (↑)	$p = 2.920 \times 10^{-11}$	N vs. NPT (↑)	$p = 2.059 \times 10^{-12}$	N vs. BST (↑)	$p = 1.626 \times 10^{-12}$
					PT vs. NPT (↑)	$p = 3.995 \times 10^{-4}$	N vs. LT (↑)	$p = 6.761 \times 10^{-7}$
							N vs. LIT (↑)	$p = 9.536 \times 10^{-8}$
							N vs. LPT (↑)	$p = 5.829 \times 10^{-8}$
<i>TPM1</i>	CAU (↑) vs. ASI	$p = 2.199 \times 10^{-11}$	N vs. M (↓)	$p = 3.825 \times 10^{-3}$	N vs. PT (↓)	$p = 3.180 \times 10^{-3}$	N vs. NET (↓)	$p = 3.638 \times 10^{-3}$
	AFA (↑) vs. ASI	$p = 7.136 \times 10^{-3}$	N vs. F (↓)	$p = 4.022 \times 10^{-3}$	N vs. NPT (↓)	$p = 4.757 \times 10^{-3}$	N vs. BST (↓)	$p = 4.404 \times 10^{-3}$
					PT vs. NPT (↑)	$p = 1.321 \times 10^{-6}$	N vs. LT (↓)	$p = 3.798 \times 10^{-3}$
							N vs. LIT (↓)	$p = 7.343 \times 10^{-3}$
							N vs. LPT (↓)	$p = 2.404 \times 10^{-3}$
<i>CASQ2</i>	CAU (↑) vs. ASI	$p = 2.406 \times 10^{-7}$	N vs. M (↓)	$p = 4.142 \times 10^{-3}$	N vs. PT (↓)	$p = 3.699 \times 10^{-3}$	N vs. NET (↓)	$p = 3.638 \times 10^{-3}$
	AFA (↑) vs. ASI	$p = 4.344 \times 10^{-2}$	N vs. F (↓)	$p = 3.979 \times 10^{-3}$	N vs. NPT (↓)	$p = 4.741 \times 10^{-3}$	N vs. BST (↓)	$p = 4.404 \times 10^{-3}$
	CAU vs. AFA (↑)	$p = 2.058 \times 10^{-2}$			PT vs. NPT (↑)	$p = 1.379 \times 10^{-3}$	N vs. LT (↓)	$p = 3.798 \times 10^{-3}$
							N vs. LIT (↓)	$p = 7.343 \times 10^{-3}$
							N vs. LPT (↓)	$p = 2.404 \times 10^{-3}$
<i>CRYAB</i>	CAU (↑) vs. ASI	$p = 4.492 \times 10^{-8}$			PT vs. NPT (↑)	$p = 4.366 \times 10^{-4}$		
	AFA (↑) vs. ASI	$p = 1.972 \times 10^{-2}$						

The findings with  $p$ -value  $< 0.05$  (Benjamini–Hochberg) were shown above. CAU (Caucasian), AFA (African American); ASI (Asian); male (M),  $n = 297$ ; female (F),  $n = 105$ ; Normal (N),  $n = 19$ ; papillary tumors (PT),  $n = 132$ ; non-papillary tumors (NPT),  $n = 271$ ; molecular subgroups: neuronal tumors (NET),  $n = 20$ ; basal squamous tumors (BST),  $n = 142$ ; luminal tumors (LT),  $n = 26$ ; luminal-infiltrated tumors (LIT):  $n = 78$ ; luminal-papillary tumors (LPT)  $n = 142$ .  $p$  ( $p$ -value).



**Figure 3.** Oncomine meta-analysis of Hub genes in bladder cancer (BCa) vs. non-cancerous tissue. Left hand: median rank (median rank of the gene across each of the analyses); p-value (p-value for the median-ranked analysis); color of the boxes indicate the percentile of the z-transformed expression level of the gene in the particular study; right hand: p = (p-value reported in each of the studies); NA (not measured in the study); (1) Blaveri et al., Clin Cancer Res, 2005 [35], invasive cancer samples n = 51, normal bladder samples n = 3; (2) Dyrskjot et al., Cancer Res, 2004 [36], invasive cancer samples n = 13, normal bladder samples n = 14; (3) Lee et al., J Clin Oncol, 2010 [37], invasive cancer samples n = 62, normal bladder samples n = 10; (4) Modlich et al. Clin Cancer Res, 2004 [38], invasive cancer samples n = 20, normal bladder samples n = 4; and (5). Sanchez-Carbayo et al., J Clin Oncol, 2006 [39], invasive cancer samples n = 72, normal bladder samples n = 52.



**Figure 4.** The expression body map of Hub genes. The median expressions of hub genes in tumors were marked in red and normal tissues were marked in green. The map is based on the GEPIA database (<http://gepia.cancer-pku.cn>), transcript per million (TPM).

### 3. Methods

The workflow of the current study is depicted in Figure 1, which has been published before [11]. For more detailed information of the datasets, materials and methods please refer to this article.

#### 3.1. Data Source Identification and Data Mining

The quality control of microarray data was conducted by relative log expression (RLE) box plot through R studio (version 1-1-463). Criteria were made for defining DEGs, compared the expression levels between the non-cancerous tissues and cancer samples, where  $|\text{Log FC (fold change)}| > 1$  and a  $p$ -value  $< 0.05$  were considered statistically significant [15,40].

#### 3.2. Acquisition of the Hub Genes

DEGs should at least be expressed in two different GEO datasets. The overlap between DEGs in different datasets was determined by FUNRICH software (version 3.1.3) and 418 DEGs were identified. Based on the interaction degree of 418 DEGs extracted from STRING database [14], cytoHubba analysis [18,41] reported 376 hub genes, of which 135 hub genes fulfilled the criterion of degree  $\geq 11$ . However, to find the most important hub genes, we focused on the top 10 hub genes, all showing a degree  $\geq 80$ . Additionally, we included another four hub genes *KPNA2*, *TPM1*, *CASQ2*, and *CRYAB*, which not only significantly correlated with overall survival based on the results from Human Protein Atlas, but also showed a degree  $\geq 11$  [15,40].



### 3.3. Functional and Clinical Analysis of the Data

FUNRICH software concerned on DEGs consisted of 11 important modules, and DAVID (Database for Annotation, Visualization and Integrated Discovery; version 6.7; <http://david.ncifcrf.gov>) emphasized on the 418 DEGs and 14 Hub genes [42]. Both software packages were used to annotate, visualize, and integrate the discoveries, and to extract the crucial biological information. We also used DAVID to analyze the up- and downregulated genes, separately. A Benjamini–Hochberg FDR <0.05 was considered significant in DAVID and FUNRICH analyses. In addition, we performed GO and KEGG analysis to identify the critical biological process (BP), cellular component (CC), molecular function (MP), and essential pathways potentially related to the initiation and development of BCa.  $p < 0.05$  was considered statistically significant.

Clinical information was extracted from TCGA-BLCA using R software and, subsequently, the expression levels of 14 hub genes in subgroups were analyzed based on tumor stage, lymph nodal metastasis, race of patients, gender of patients, histological subtype, and molecular subtypes. To evaluate the prognostic value of the identified DEGs, we did Kaplan–Meier survival analyses of overall survival (OS) and disease-free survival (DFS).  $p < 0.05$  was considered statistically significant.

### 3.4. Literature Retrieval and Oncomine Meta-Analysis

PubMed, EMBASE, Science Direct, and Google Scholar databases were used to search and identify published results about bioinformatics analysis in BCa, via the Google search engine. The data collection process was undertaken and ended in November 2019. The retrieval criteria were rigorous, filtering with bioinformatics analysis and BCa, only 7 full-text papers were returned [9,29–34]. On the basis of Oncomine database, 5 studies were relevant [35] and we gathered the information of the 14 hub genes from the 5 previous studies.

## 4. User Notes

The present report describes the character of the research “Identification of key biomarkers in bladder cancer: Evidence from bioinformatics analysis” [11]. Furthermore, the present report provides a convenient way to use extended datasets for biomarker discovery and hypothesis generation.

**Supplementary Materials:** We provide the supplementary Tables S1–S3 listing step by step the identification process of the most promising DEGs. The DEGs listed in Table S3 were used during the following analysis steps. We also indicated up- or downregulation of those DEGs (↑; ↓) in bladder cancer. Table S4 provides the 376 hub genes defined by cytoHubba. Table S5 summarizes the clinical and gene expression data of the  $n = 406$  TCGA-BLCA patients for the 14 hub genes, defined from their degree of network interaction and the 11 seed genes, defined from the 11 most important modules. This table is the basis for the gene expression analyses. Tables S6–S8 summarize the results of the GO and KEGG pathway analyses. Table S9 provides the data of the Oncomine meta-analysis. The data tables may be used to construct a complete data set after applying different normalization strategies as cross-platform normalization or batch effects removal. This data set can be used as a benchmarking data set for machine learning-based feature selection in data-driven biomarker research. The following supplemental data are available online at <http://www.mdpi.com/2306-5729/5/2/38/s1>, Table S1: DEGs extracted from different dataset, Table S2: 726 DEGs extracted from 5 GEO datasets, Table S3: 418 DEGs in BCa, Table S4: 376 Hub genes in BCa, Table S5: Clinical meta-data of 11 seed genes and 14 Hub genes, Table S6: GO and KEGG results of 418 DEGs, Table S7: GO and KEGG results of 132 upregulated DEGs, Table S8: GO and KEGG results of 286 downregulated DEGs, Table S9: Oncomine meta-analysis results of previous results.

**Author Contributions:** Conceptualization, C.Z. and J.N.; Data curation, C.Z.; Supervision, J.N.; Writing—original draft, C.Z.; Writing—review and editing, M.B.-P. and J.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the Leipzig University for Open Access Publishing.

**Acknowledgments:** We thank the research team of our laboratory for discussion and advice with the data analysis. We acknowledge support from Leipzig University for Open Access Publishing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kulkarni, G.S.; Black, P.C.; Sridhar, S.S.; Kapoor, A.; Zlotta, A.R.; Shayegan, B.; Rendon, R.A.; Chung, P.; van der Kwast, T.; Alimohamed, N.; et al. Canadian Urological Association guideline: Muscle-invasive bladder cancer. *Can. Urol. Assoc. J.* **2019**, 230–238. [[CrossRef](#)] [[PubMed](#)]
2. Flaig, T.W.; Spiess, P.E.; Agarwal, N.; Bangs, R.; Boorjian, S.A.; Buyyounouski, M.K.; Downs, T.M.; Efsthathiou, J.A.; Friedlander, T.; Greenberg, R.E.; et al. NCCN Guidelines Insights: Bladder Cancer, Version 5.2018. *J. Natl. Compr. Cancer Netw. JNCCN* **2018**, 16, 1041–1053. [[CrossRef](#)] [[PubMed](#)]
3. Babjuk, M.; Burger, M.; Comperat, E.M.; Gontero, P.; Mostafid, A.H.; Palou, J.; van Rhijn, B.W.G.; Roupert, M.; Shariat, S.F.; Sylvester, R.; et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) 2019 Update. *Eur. Urol.* **2019**, 76, 639–657. [[CrossRef](#)] [[PubMed](#)]
4. Zaghloul, M.S.; Christodouleas, J.P.; Smith, A.; Abdallah, A.; William, H.; Khaled, H.M.; Hwang, W.T.; Baumann, B.C. Adjuvant Sandwich Chemotherapy Plus Radiotherapy vs. Adjuvant Chemotherapy Alone for Locally Advanced Bladder Cancer After Radical Cystectomy: A Randomized Phase 2 Trial. *JAMA Surg.* **2018**, 153, e174591. [[CrossRef](#)]
5. Berndt-Paetz, M.; Weimann, A.; Sieger, N.; Schastak, S.; Riyad, Y.M.; Griebel, J.; Arthanareeswaran, V.K.A.; Stolzenburg, J.U.; Neuhaus, J. Tetrahydroporphyrin-tetratosylat (THPTS): A near-infrared photosensitizer for targeted and efficient photodynamic therapy (PDT) of human bladder carcinoma. An in vitro study. *Photodiagnosis Photodyn. Ther.* **2017**, 18, 244–251. [[CrossRef](#)]
6. Kutwin, P.; Konecki, T.; Cichocki, M.; Falkowski, P.; Jablonowski, Z. Photodynamic Diagnosis and Narrow-Band Imaging in the Management of Bladder Cancer: A Review. *Photomed. Laser Surg.* **2017**, 35, 459–464. [[CrossRef](#)]
7. Power, N.E.; Izawa, J. Comparison of Guidelines on Non-Muscle Invasive Bladder Cancer (EAU, CUA, AUA, NCCN, NICE). *Bladder Cancer (Amst. Neth.)* **2016**, 2, 27–36. [[CrossRef](#)]
8. Gao, Y.; Liu, S.; Guo, Q.; Zhang, S.; Zhao, Y.; Wang, H.; Li, T.; Gong, Y.; Wang, Y.; Zhang, T.; et al. Increased expression of TRIP13 drives the tumorigenesis of bladder cancer in association with the EGFR signaling pathway. *Int. J. Biol. Sci.* **2019**, 15, 1488–1499. [[CrossRef](#)]
9. Gao, X.; Chen, Y.; Chen, M.; Wang, S.; Wen, X.; Zhang, S. Identification of key candidate genes and biological pathways in bladder cancer. *PeerJ* **2018**, 6, e6036. [[CrossRef](#)]
10. Stein, J.P.; Lieskovsky, G.; Cote, R.; Groshen, S.; Feng, A.C.; Boyd, S.; Skinner, E.; Bochner, B.; Thangathurai, D.; Mikhail, M.; et al. Radical cystectomy in the treatment of invasive bladder cancer: Long-term results in 1,054 patients. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2001**, 19, 666–675. [[CrossRef](#)]
11. Zhang, C.; Berndt-Paetz, M.; Neuhaus, J. Identification of Key Biomarkers in Bladder Cancer: Evidence from a Bioinformatics Analysis. *Diagnostics* **2020**, 10, 66. [[CrossRef](#)] [[PubMed](#)]
12. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, 30, 207–210. [[CrossRef](#)] [[PubMed](#)]
13. Tomczak, K.; Czerwinska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. (Pozn. Pol.)* **2015**, 19, A68–A77. [[CrossRef](#)] [[PubMed](#)]
14. Pathan, M.; Keerthikumar, S.; Ang, C.S.; Gangoda, L.; Quek, C.Y.; Williamson, N.A.; Mouradov, D.; Sieber, O.M.; Simpson, R.J.; Salim, A.; et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **2015**, 15, 2597–2601. [[CrossRef](#)]
15. Pan, Y.; Lu, L.; Chen, J.; Zhong, Y.; Dai, Z. Identification of potential crucial genes and construction of microRNA-mRNA negative regulatory networks in osteosarcoma. *Hereditas* **2018**, 155, 21. [[CrossRef](#)]
16. Assenov, Y.; Ramírez, F.; Schelhorn, S.E.; Lengauer, T.; Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics (Oxf. Engl.)* **2008**, 24, 282–284. [[CrossRef](#)]
17. Wang, J.; Zhong, J.; Chen, G.; Li, M.; Wu, F.X.; Pan, Y. ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, 12, 815–822. [[CrossRef](#)]
18. Smoot, M.E.; Ono, K.; Ruscheinski, J.; Wang, P.L.; Ideker, T. Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics (Oxf. Engl.)* **2011**, 27, 431–432. [[CrossRef](#)]
19. Zaravinos, A.; Lambrou, G.I.; Boulalas, I.; Delakas, D.; Spandidos, D.A. Identification of common differentially expressed genes in urinary bladder cancer. *PLoS ONE* **2011**, 6, e18135. [[CrossRef](#)]

20. Zaravinos, A.; Lambrou, G.I.; Volanis, D.; Delakas, D.; Spandidos, D.A. Spotlight on differentially expressed genes in urinary bladder cancer. *PLoS ONE* **2011**, *6*, e18255. [[CrossRef](#)]
21. Borisov, N.; Tkachev, V.; Suntsova, M.; Kovalchuk, O.; Zhavoronkov, A.; Muchnik, I.; Buzdin, A. A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency. *Cell Cycle (Georget. Tex.)* **2018**, *17*, 486–491. [[CrossRef](#)] [[PubMed](#)]
22. Zhao, F.; Lin, T.; He, W.; Han, J.; Zhu, D.; Hu, K.; Li, W.; Zheng, Z.; Huang, J.; Xie, W. Knockdown of a novel lincRNA AATBC suppresses proliferation and induces apoptosis in bladder cancer. *Oncotarget* **2015**, *6*, 1064–1078. [[CrossRef](#)]
23. Lu, M.; Ge, Q.; Wang, G.; Luo, Y.; Wang, X.; Jiang, W.; Liu, X.; Wu, C.L.; Xiao, Y.; Wang, X. CIRBP is a novel oncogene in human bladder cancer inducing expression of HIF-1alpha. *Cell Death Dis.* **2018**, *9*, 1046. [[CrossRef](#)]
24. Hu, Q.; Wang, G.; Peng, J.; Qian, G.; Jiang, W.; Xie, C.; Xiao, Y.; Wang, X. Knockdown of SIRT1 Suppresses Bladder Cancer Cell Proliferation and Migration and Induces Cell Cycle Arrest and Antioxidant Response through FOXO3a-Mediated Pathways. *Biomed. Res. Int.* **2017**, *2017*, 3781904. [[CrossRef](#)] [[PubMed](#)]
25. He, W.; Zhong, G.; Jiang, N.; Wang, B.; Fan, X.; Chen, C.; Chen, X.; Huang, J.; Lin, T. Long noncoding RNA BLACAT2 promotes bladder cancer-associated lymphangiogenesis and lymphatic metastasis. *J. Clin. Investig.* **2018**, *128*, 861–875. [[CrossRef](#)] [[PubMed](#)]
26. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
27. Kanehisa, M. The KEGG database. *Novartis Found. Symp.* **2002**, *247*, 91–101.
28. Amin, M.B. Histological variants of urothelial carcinoma: Diagnostic, therapeutic and prognostic implications. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc.* **2009**, *22* (Suppl. 2), S96–S118. [[CrossRef](#)]
29. Han, Y.; Jin, X.; Zhou, H.; Liu, B. Identification of key genes associated with bladder cancer using gene expression profiles. *Oncol. Lett.* **2018**, *15*, 297–303. [[CrossRef](#)]
30. Hu, J.; Zhou, L.; Song, Z.; Xiong, M.; Zhang, Y.; Yang, Y.; Chen, K.; Chen, Z. The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets. *J. Cell. Physiol.* **2019**. [[CrossRef](#)]
31. Jia, Z.; Ai, X.; Sun, F.; Zang, T.; Guan, Y.; Gao, F. Identification of new hub genes associated with bladder carcinoma via bioinformatics analysis. *Tumori* **2015**, *101*, 117–122. [[CrossRef](#)] [[PubMed](#)]
32. Han, Y.; Zheng, Q.; Tian, Y.; Ji, Z.; Ye, H. Identification of a nine-gene panel as a prognostic indicator for recurrence with muscle-invasive bladder cancer. *J. Surg. Oncol.* **2019**, *119*, 1145–1154. [[CrossRef](#)] [[PubMed](#)]
33. Jiang, W.D.; Yuan, P.C. Molecular network-based identification of competing endogenous RNAs in bladder cancer. *PLoS ONE* **2019**, *14*, e0220118. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, Y.; Fang, L.; Zang, Y.; Xu, Z. Identification of Core Genes and Key Pathways via Integrated Analysis of Gene Expression and DNA Methylation Profiles in Bladder Cancer. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2018**, *24*, 3024–3033. [[CrossRef](#)]
35. Blaveri, E.; Simko, J.P.; Korkola, J.E.; Brewer, J.L.; Baehner, F.; Mehta, K.; Devries, S.; Koppie, T.; Pejavar, S.; Carroll, P.; et al. Bladder cancer outcome and subtype classification by gene expression. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2005**, *11*, 4044–4055. [[CrossRef](#)]
36. Dyrskjot, L.; Kruhoffer, M.; Thykjaer, T.; Marcussen, N.; Jensen, J.L.; Moller, K.; Orntoft, T.F. Gene expression in the urinary bladder: A common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res.* **2004**, *64*, 4040–4048. [[CrossRef](#)]
37. Lee, J.S.; Leem, S.H.; Lee, S.Y.; Kim, S.C.; Park, E.S.; Kim, S.B.; Kim, S.K.; Kim, Y.J.; Kim, W.J.; Chu, I.S. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2010**, *28*, 2660–2667. [[CrossRef](#)]
38. Modlich, O.; Prisack, H.B.; Pitschke, G.; Ramp, U.; Ackermann, R.; Bojar, H.; Vogeli, T.A.; Grimm, M.O. Identifying superficial, muscle-invasive, and metastasizing transitional cell carcinoma of the bladder: Use of cDNA array analysis of gene expression profiles. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2004**, *10*, 3410–3421. [[CrossRef](#)]
39. Sanchez-Carbayo, M.; Socci, N.D.; Lozano, J.; Saint, F.; Cordon-Cardo, C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2006**, *24*, 778–789. [[CrossRef](#)]

40. Li, L.; Lei, Q.; Zhang, S.; Kong, L.; Qin, B. Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis. *Oncol. Rep.* **2017**, *38*, 2607–2618. [[CrossRef](#)]
41. Chin, C.H.; Chen, S.H.; Wu, H.H.; Ho, C.W.; Ko, M.T.; Lin, C.Y. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8* (Suppl. 4), S11. [[CrossRef](#)] [[PubMed](#)]
42. Huang, D.W.; Sherman, B.T.; Tan, Q.; Collins, J.R.; Alvord, W.G.; Roayaei, J.; Stephens, R.; Baseler, M.W.; Lane, H.C.; Lempicki, R.A. The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **2007**, *8*, R183. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).