Gene Expression and Regulatory Webwork of POLR2K in Bladder Carcinogenesis by Integrated Bioinformatics Approaches

Liliang Yang
China Pharmaceutical University

Kaizhen Wang
China Pharmaceutical University

Wenjing Guo
China Pharmaceutical University

Xian Chen
China Pharmaceutical University

Qinglong Guo
China Pharmaceutical Univeristy

Libin Wei
China Pharmaceutical University

Yuxin Zhou (✉ zyxspain@hotmail.com)
China Pharmaceutical University

Research article

Abstract

Background RNA polymerase II subunit K (POLR2K) belongs to one of the multiple subunits of RNA polymerase II (Pol II), whose biological function is to synthesize mRNA. Aberrant *POLR2K* expression is related to carcinogenesis. However, *POLR2K*'s underlying role in bladder cancer has not been explored. In the current study, we intend to analyze the function of *POLR2K* and its regulatory network within bladder cancer.

Methods: Public sequencing data was obtained from GEO and TCGA to investigate *POLR2K* expression and regulatory network within bladder cancer (BLCA) by using GEPIA and Oncomine as well as cBioPortal online tool. LinkedOmics was employed to identify genes displaying significantly differential expression patterns and to perform GO and KEGG analyses. After differential genes was assigned and ranked, GSEA analyses was performed to obtain target networks for transcription factors, miRNAs, and kinases that could regulate *POLR2K*–associated gene network. Subsequent functional webwork analyses were used to identify cancer-relevant pathways Moreover, *POLR2K* gene is verified, by ChIP-seq in MCF-7 cell line , with transcription factor binding evidence in the ENCODE Transcription Factor Binding Site Profiles dataset.

Conclusions: The current study implies that *POLR2K* gene is overexpressed and often amplified in BLCA, providing the first evidence that *POLR2K* deregulation, in particular increased transcription, may promote BLCA. These findings uncover a unique expression patterns of POLR2K and its potential regulatory networks in BLCA, contributing greatly to study of the role of *POLR2K* in cancer development.

1. Introduction

Bladder cancer is the 11th most prevalently diagnosed cancer worldwide [1]. About 70% of patients initially diagnosed with non-muscle-invasive bladder cancer (NMIBC), needs post-surgical management and appropriated surveillance, which could be not only costly but also invasive [2]. Despite high recurrence and metastasis, the development of various targeted drugs , including Erdafitinib and Nivolumab[3], has contributed to the significant improvement in the survival of patients and brought about a revolutionary breakthrough in the treatment of advanced BLCA[2, 4], but these targeted drugs have also shown unsatisfactory results. Erdafitinib is prescribed only when abnormal fibroblast growth factor receptor (FGFR) gene is detected in advanced BLCA patients. And Nivolumab may cause serious side effects since it could mislead the immune system to attack healthy cells. The etiology of BLCA is particularly multiplex, including these processes involved in signal transduction and transcriptional regulation, which reflects the accumulated physiological effects of multiple genes [5]. Screening gene networks for alterations associated with carcinogenesis and cancer proliferation could be an encouraging method to identify novel drug target for BLCA.

RNA polymerase II subunit K (*POLR2K*), also known as RPABC4, which is a gene encoding one of multiple subunits of RNA polymerase II (Pol II), whose biological function is to synthesize mRNA. This subunit could be also found in another two polymerases, such as RNA polymerase III (Pol III) and RNA polymerase I (Pol I), respectively. *POLR2K* could be utilized to synthesize diverse function non coding RNAs, small RNAs, mRNA progenitors, and ribosomal RNA progenitors. *POLR2K* is related to DNA replication, translation synthesis and transcription [6, 7] while it participates in many critical cell signaling pathways, where promoting tumor formation and progression[8]. Regulation of *POLR2K* was first explored in chronic hepatitis, where its association with IRF1 was clarified[9], and then its over-expression was confirmed in multiple malicious cancers, containing bladder/breast/ovarian cancer[10, 11]. However, what role *POLR2K* could play has not been investigated in BLCA. We intend to clarify the influences on cancer development, expression patterns and regulatory networks of *POLR2K* in BLCA.

In this studies, we found that *POLR2K* was overexpressed in BLCA patient samples from The Cancer Genome Atlas compared to normal liver samples. In addition, Kaplan-Meier survival analysis displayed that poor progression-free survival and poor overall survival of BLCA patients were related to *POLR2K* overexpression (Supplementary Figure 1). Therefore, we further explore the role of *POLR2K* in BLCA from various public databases. Genomic alterations and functional networks relating to *POLR2K* in BLCA were analyzed by using multi-dimensional analysis methods. Thus, these results reveal a novel function of *POLR2K* as a regulator of cancer proliferation and provide a potential strategy of treatment for BLCA. *POLR2K* expression might function as a novel therapeutic target for bladder cancer.

2. Methods

2.1 Oncomine analysis

Oncomine 4.5 (www.oncomine.org), one of the largest existing oncogene chip database which currently includes 264 independent datasets, containing various cancer types analysis methods[12], was used for analyzing DNA copy number and mRNA expression of *POLR2K* within bladder cancer whose publicly-accessible online database could be derived from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo). *POLR2K* expression was evaluated in bladder carcinoma tissues relative to normal bladder tissues where significant differences were determined as $p \leq$ 0.01.

2.2 Pathway Commons

Pathway Commons is a collective database of biological pathway database derived from multiple sources associated with interactions from various organisms and functional correlations between genes in signaling pathways[13], whose data can be derived from its partner databases and the pathway data are represented in the BioPAX file.*POLR2K* was used as seed gene to acquire the neighboring associated genes.

2.3 UALCAN analysis

UALCAN is a comprehensive transcriptome database whose data is collected from the TCGA, and correlation between clinicopathological features and gene expression could be retrieved [14]. UALCAN (ualcan.path.uab.edu) is used to analyze cancer transcriptome data based on TCGA database.

2.4 cBioPortal analysis

The platform(cbioportal.org), termed as cBioPortal, has hosted more than 200 cancer genomics studies , which is often used to explore, visualize, and analyze multi-dimensional cancer genomic and clinical data information based on its multiple analysis features[15]. The mRNA expression, CNVs, and mutation of *POLR2K* were chosen as search parameters and then the tab Network was used for visualizing its biological interaction network based on the alteration frequencies. Cytoscope was further introduced to analyze GO and KEGG pathway.

2.5 LinkedOmics analysis

With abundance in multi-omics and clinical data information of various cancers from TCGA-associated multidimensional datasets[16], LinkedOmics (www.linkedomics.org/ login.php) was utilized to investigate mRNA sequence data from 408 BLCA patients within the Cancer Genome Atlas database. The gene expression difference correlated with *POLR2K* was studied using LinkFinder module and Pearson's correlation coefficients was chosen as the method to investigate the results. Individual gene's statistical plot are presented in heat maps or scatter plots. After the results were ranked, the pathway enrichment analysis of GO and KEGG as well as target enrichment analysis of transcription factor, miRNA and kinase was performed by GSEA. Target enrichment of the transcription factor and miRNA was primarily derived from Molecular Signatures Database.

2.6 GEPIA analysis

GEPIA, an interactive web application providing and analyzing the RNA sequencing expression data based on the GTEx and TCGA projects [17], enables users to perform various of gene expression analyses and survival analyses including overall survival and progression-free survival. For the survival analysis, *POLR2K* was input as a gene symbol; BLCA was added as the datasets; other setting, except Methods tab, was unchanged by default.

2.7 GeneMANIA analysis

GeneMANIA is used for generating gene function hypotheses, analyzing gene list and then prioritizing genes for functional category assays [18]. When given a query gene list, GeneMANIA (www.genemania.org), by utilizing publically-available proteomics and genomics data information, could extend this list with functional similar and interacting genes to produce a functional related networks. GeneMANIA was also used for visualizing the gene network and predicting gene functions.

2.8 Enrichr analysis

Enrichr is a collective annotation resource for curated human gene datasets and a search engine of accumulated multiple biological knowledge about billions of biological discoveries [19]. We use 50 neighboring genes of POLR2K to inquire Enrichr in Cell Types tab (amp.pharm.mssm.edu/Enrichr).

3. Results
3.1 *POLR2K* expression in bladder cancer

Using TCGA and the Gene Expression Omnibus (GEO) database, we firstly assessed *POLR2K* transcription levels in BLCA studies. Analysis of the Oncomine database revealed an increase in DNA copy number variation (CNV) and mRNA expression of *POLR2K* in BLCA tissues compared normal tissues ($p ⊠ 0.01$). *POLR2K* ranked within the top 17% of transcriptome profile and within the top 9% of DNA CNVs (Figure 1), when the fold differences were adjusted to 2. In addition, correlation analysis of *POLR2K* with clinic pathological features of 408 bladder cancer patients from the Cancer Genome Atlas database was conducted by UALCAN website. And the results consistently suggested high transcription levels of *POLR2K* in BLCA. Compared with healthy people, the levels of transcription for *POLR2K* was much higher in bladder cancer patient based on the sub-group such as disease stages, gender, smoking condition, histological subtypes, and nodal metastasis status (Figure 2). Furthermore, Kaplan-Meier

survival analysis displayed high levels of *POLR2K* was positively related to poor progression-free survival(PFS) and overall survival(OS) in bladder cancer patients (Supplementary Figure 1).Therefore, *POLR2K* expression could become a potential diagnostic indicator in BLCA.

3.2 Alterations of *POLR2K* in bladder cancer

3.2.1 *Type and frequency of* POLR2K *alterations in bladder cancer*

In order to investigate the type and frequency of *POLR2K* changes within bladder cancer, we then used the ciBioPortal to analyze sequence data from BLCA patient issues in the TCGA database. Compared with normal healthy tissues, *POLR2K* levels was altered in 141 of 408 (35%) BLCA patient tissues (Figure 3a). These alterations includes mRNA up-regulation in 139 cases (34.5%), amplification in 59 cases (14.7%), multiple alterations in 55 cases (13.7%) mRNA and down-regulation in 2 case (0.49%) (Supplementary Fig 2). Therefore, amplification is identified in BLCA, as the most prevalent type of *POLR2K* CNV.

3.2.2 *Interaction network of* POLR2K *alterations in bladder cancer*

Biological interaction network of *POLR2K* in BLCA was next investigated. To achieve this, we used Common Pathways to obtain *POLR2K*−neighboring genes, and then used Cbioportal to inquire the alteration frequencies of these genes. (Figure 3b and Supplementary Fig 2). The neighboring genes of *POLR2K* makes KDMSA (35%), KMT2D (33%) together with INTS8 (32%) as one of the most frequent alterations in BLCA. With the help of Cytoscape, analysis of GO category annotations implied these genes encoding proteins localized primarily to, or functioned at, RNA polymerase complex, the spliceosomal complex, and MLL3/4 complex. These proteins are fundamentally related to posttranscriptional gene expression processes (i.e. snRNA and dsRNA processing, transcription-coupled nucleotide-excision repair, etc), but they were also enriched such as "U2-type spliceosomal complex" and "basal transcription machinery binding" (Figure 4a−4c). As expected, KEGG pathway analysis indicated that *POLR2K* is connected to Spliceosome, RNA polymerase, mRNA surveillance and Basal transcription factors (Figure 4d). As a result, the interaction network of *POLR2K* alterations is related to several RNA metabolic processes and the gene expression regulation.

3.3 Enrichment analysis based on *POLR2K* functional network in bladder cancer

3.3.1 *GO annotation and KEGG enrichment analysis for co-expression genes related to* POLR2K *in bladder cancer*

Using LinkedOmics online tool, we took advantage of TCGA database to explore mRNA sequencing data comprising 408 patients of BLCA. From the volcano plot depicted in Figure 5a, significant negative correlations with *POLR2K* was suggested among 2395 genes (dark green dots), while significant positive correlations was indicated among 2635genes (dark red dots) (P-value<0.01, false discovery rate FDR< 0.05). The heat map displayed 50 significant genes whose mRNA expression positively or negatively correlated with *POLR2K* in BLCA (Figure 5b, 5c). This result demonstrates that *POLR2K* have an extensive influence on the transcriptome. Supplementary Figure 3a−3c show

scatter plots for each gene. Significantly, *POLR2K* expression is positively related to the expression of ZNF706 (positive rank #1, Pearson correlation = 0.7887, p = 8.055e-88), YWHAZ (Pearson correlation = 0.6695, p = 2.224e-54), and COX6C(Pearson correlation = 0.6638, p = 3.552e-53), which mirrors alterations in mitochondrial composition as well as apoptosis, DNA repair, and transcriptional regulation. Further exploration by GSEA analysis demonstrated differentially expressed genes were primarily related to mitochondrial protein complex, spliceosome complex, replication fork and condensed chromosome, whereas those genes are involved mainly in mitochondrial gene expression, mRNA/ncRNA processing, DNA replication and cell cycle. These genes also serves as constituent of mitochondria and ribosomes (Figure6a–6c and Supplementary Tables 1–3).

Interestingly, GSEA analysis implied that these genes were also located in MHC protein complex, secretory granule membrane, cell leading edge and extracellular matrix, where they participated in the positive regulation of cell motility, antigen binding and immune response-regulating signaling pathway. They were also involved in response to type I interferon, I-kappaB kinase/NF-kappaB signaling, Toll-like / NOD-like receptor signaling pathway and JAK-STAT signaling pathway, all of which were associated to tumor immune escape. Also, GSEA analysis of KEGG pathway showed highly significant enrichment in cell cycle, DNA replication, ribosome and spliceosome, as well as in Natural killer cell mediated cytotoxicity and Cytokine-cytokine receptor interaction (Figure 6d, 6e and Supplementary Table 4). Taken together, *POLR2K* function networks were mainly responsible for gene expression, mRNA surveillance, cell cycle, DNA replication while were also involved in tumor immune response and survival, which demonstrated significant deregulation of cancer related pathways.

3.3.2 POLR2K *networks of transcription factor, miRNA and kinase targets and in bladder cancer*

Transcription factor, miRNA and Kinase target network of associated gene dataset created by GSEA were used to examine targets of *POLR2K* in BLCA. The significant kinase-target networks were associated mainly with mitogen-activated protein kinase 6 (Kinase_MAPK6), homeodomain interacting protein kinase 2（Kinase_HIPK2）and mitogen-activated protein kinase 7 (Kinase_MAPK7) (Table 1 and Supplementary Tables 5–7); the miRNA-target network was generally related to AACTGGA_MIR145, TTGCCAA_MIR202, ACTTTAT_MIR507, GTGCAAT_MIR25_MIR32_MIR92_MIR363_MIR367 and TATTATA_MIR374; the transcription factor-target network was involved fundamentally in the cell cycle regulation factors including V\$E2F1_Q6, V\$FOXO1_02 and V\$CDC5_01, whereas also involved in the IRF Transcription Factor (IRF) family including V\$IRF7_01, V\$IRF1_01, and V\$ISRE_01. Correlation among genes for the Kinase MAPK7, miR-145 and E2F1_Q6 respectively was uncovered by subsequent protein-protein interaction networks created by GeneMANIA. These genes enriched for transcription factor E2F1 has been linked primarily to modulating cell cycle checkpoint, DNA repair and DNA replication, while associated with MCM complex, G1/S transition of mitotic cell cycle and telomere maintenance (Figure 7). Furthermore, genes enriched for MAPK7 is related mainly to regulation of innate immune response, toll-like receptor family signaling and stress-activated protein kinase signaling cascade (Supplementary Figure 4). And genes enriched for miRNA-374 can be related primarily to modulation of Notch signaling pathway, activating transcription factor binding and modulation of transcription in other organism involved in symbiotic interaction (Supplementary Figure 5)

Table 1
The Kinase, miRNA and transcription factor-target networks of POLR2K in bladder urothelial carcinoma (LinkedOmics).

Enriched Category	Geneset	LeadingEdgeNum	FDR
Kinase Target	Kinase_HIPK2	6	0.0468
	Kinase_MAPK7	12	0.0025
	Kinase_MAPK6	8	0.0380
miRNA Target	AACTGGA_MIR145	73	0.0286
	TTGCCAA_MIR507	47	0.0097
	ATAGGAA_MIR202	34	0.0243
	GTGCAAT_MIR25_MIR32_MIR92_MIR363_MIR367	82	0.0184
	TATTATA_MIR374	106	0.0185
Transcription Factor Target	V$E2F1_Q6	74	0.0312
	V$CDC5_01	101	0.0095
	V$FOXO1_02	93	0.0134
	V$IRF_Q6	74	0.0016
	V$ISRE_01	92	0
	V$IRF7_01	74	0.0078
	V$IRF1_01	81	0.0069

Abbreviations: LeadingEdgeNum, the number of leading edge genes; FDR, false discovery rate from Benjamini and Hochberg from gene set enrichment analysis (GSEA). V$, the annotation found in Molecular Signatures Database (MSigDB) for transcription factors (TF).

4. Discussion

That DNA-directed RNA polymerases (Pol I, Pol II and Pol III) is differentially expressed and deregulated has been detailed in multiple cancers [20]. *POLR2K*, a subunit of this polymerases, participates in multiple steps of transcription [21].In this paper, we initially discovered that *POLR2K* was overexpressed in 408 BLCA samples from the TCGA and that high *POLR2K* expression may serve as an indicator of poor survival. Subsequently, further studies of the expression of *POLR2K* and its regulatory network will be necessary to obtain additional insight into the possible function of *POLR2K* in BLCA. Therefore, we conducted bioinformatic analyses of sequence data and hoped to motivate the future research of bladder cancer.

It has been a wearing problem for years since early detection approach of BLCA eluded many clinicians. Cystoscopy and urine cytology are currently employed to diagnose BLCA [22]. However, cystoscopy is an invasive method and also low sensitivity for bladder carcinoma in situ. Urine cytology is a non-invasive and has a higher specificity but lower sensitivity for low-grade urothelial tumors. In spite of the search for urinary biomarkers for the early and non-invasive detection of BLCA, no available biomarkers are currently employed in clinical practice. Thus, potential BLCA biomarkers are urgently required to increase the early diagnosis of BLCA. We also demonstrated that CNVs and mRNA expression of *POLR2K* are much higher in bladder cancer tissues compared with normal tissues by analyzing

transcriptional sequence data from clinical patient tissues. We discovered that *POLR2K* overexpression exists in many cases of bladder carcinogenesis. We think that POLR2K research could move the field of liquid biopsy biomarkers forward, which deserves additional clinical and research attention. And it needs to be verified whether POLR2K could be detectable in liquid biopsy.

It has been reported that CNVs could directly influence gene expression and have drastic phenotypic consequences due to altering gene dosage or disrupting coding sequences. [23, 24]. This work discovered that amplification was the main type of *POLR2K* change and *POLR2K* copy number was augmented, which was related to poor survival including overall survival (OS) and disease-free survival (DFS) (Supplementary Figure 1). In light of these findings, we inferred that alterations in chromosomal structure might be involved in the altered *POLR2K* expression as well as *POLR2K* dysfunction in BLCA. *POLRAK* genetic change could lead to alterations in numerous downstream signals that may ultimately result in carcinogenesis for *POLR2K* plays a critical role in multiple biological functions. In fact, neighboring gene networks in close proximity to *POLR2K* display amplification with varying strengths in BLCA. Meanwhile, associated functional network are found to be related to spliceosome signaling, mRNA surveillance and ribosome signaling. Therefore, the network of *POLR2K* alterations is related to posttranscriptional modulation, which is involved in protein translation as well as RNA splicing, in consistence with several other published results about the biological roles of *POLR2K* [25, 26]. Furthermore, related functional networks are also involved in positive regulation of cell motility, antigen binding and immune response-regulating signaling pathway. Therefore, the networks based on *POLR2K* genetic change is also associated with the tumor immune response, which is closely related to the mechanism by which tumor escapes from host immune system.

To reveal critical network of transcription factors, miRNAs and target kinases, GSEA analysis is performed. Our results imply that the functional network of *POLR2K* is involved generally in the ribosome, spliceosome, mRNA/ncRNA processing, cell cycle and DNA replication. Just the same as the mutation webwork, the functional association network that integrates the effects of *POLR2K* transcription alteration participates in RNA metabolic processes and gene expression regulation. As represented above, we conclude that *POLR2K* has a profound influence on the maintenance of short introns splicing [27, 28].

Oncogenic kinases play a critical role in coupling intracellular signaling pathways with extracellular signals, which promote cancer progression in all stages [29]. We discovered that *POLR2K* in BLCA is related to a network of kinases including Kinase_MAPK7, Kinase_MAPK6 and Kinase_HIPK2. All of these kinases regulate mitosis, gene expression and apoptosis. Indeed, HIPK2 has been recognized as a signaling molecule which acts in various signal transduction pathways and cellular processes such as cell proliferation and apoptosis, transcriptional regulation and antiviral responses[30, 31] while MAPK7 could facilitate tumor escape from immune surveillance [32, 33] . Deregulated activity of HIPK2 may affect the genome integrity, leading to cancer development [34]. In BLCA, *POLR2K* might modulate DNA replication, repair, and gene expression via HIP2 kinase.

"Continuous proliferation" has been proposed as top one of 10 hallmark features of tumors [35]. One crucial explanation is that cell cycle-associated proteins, if aberrantly expressed, could contribute to cell cycle disorder in tumor cells, which results in decreased differentiation, abnormal proliferation and rapid progression in cancer cells[36]. E2F1 is among key links in cell cycle modulation web-work. Abnormal E2F1 expression proactively is related to the tumor formation and progression of BLCA [37]. There is one study has published that elevated levels of E2F1 is involved in shorter survival of bladder cancer patients [38], and another study showed that the *POLR2K* network of transcription factor targets is related to E2F1[39]. Thus, these analyses suggest that E2F1 may be a critical target of *POLR2K* which modulates cell cycle and cancer progression in BLCA patients by acting through

E2F1 transcription factor. Interestingly, according to known transcription factor binding site motifs from the TRANSFAC, we found that E2F1 could directly bind the promoter region of *POLR2K* likely to modulate the expression of POLR2K (http://amp.pharm.mssm.edu/Harmonizome/gene_set/E2F1/TRANSFAC+Predicted+Transcription+Factor+Targets). More convincingly, *POLR2K* gene is verified, by ChIP-seq in MCF-7 cell line , with transcription factor binding evidence in the HA-E2F1_MCF-7_hg19_1 transcription factor binding site profile from the ENCODE Transcription Factor Binding Site Profiles dataset (http://amp.pharm.mssm.edu/Harmonizome/gene_set/HA-E2F1_MCF7_hg19_1/ENCODE+Transcription+Factor+Binding+Site+Profiles). Thus, this analyses indicate that E2F1 could interact with POLR2K to modulate the survival of cancer cells.

Cancer and chronic inflammation could be linked together by activating innate immune responses through NLR and TLR signals [40]. Chronic inflammation, mainly due to aberrant inflammasome or NF-κB activation, is closed coupled with cancer through TLRs-involved cytokine production[41]. The IFN-regulatory factor family proteins are transcription factors with varying biological functions, which might render a microenvironment for immune evasion and tumor progression [42]. Therefore, our analysis could represent that *POLR2K* might play a vital role in tumor microenvironment where *POLR2K* act as a chronic inflammation keeper and act as an immune surveillance regulator.

Our data mining also recognized some miRNAs that were related to *POLR2K*. These miRNAs are short non-coding RNA molecules that post-transcriptional regulate protein expression. Distinct miRNAs alterations could be used to characterize BLCA [43, 44]. The particular miRNAs in this paper are associated with cancer occurrence, metastasis and invasion. Indeed, miR-145, miR-202 and miR-374 has been proposed as diagnostic, prognostic or therapeutic marker of BLCA [45, 46] .miR-202 and miR-347 participate in invasion, metastasis and cancer progression[46, 47], while miR-145 modulates *suppressor of cytokine signaling 7* (*socs7*) to enhance IFN-β expression, thus contributing to BLCA apoptosis. We speculate that deregulation of these miRNAs would be in consistence with the phenotype of *POLR2K* overexpression in BLCA, which should be further verified by experiments.

Our study presents striking evidence for the significance of *POLR2K* in urinary bladder carcinogenesis and demonstrate its potential as an early indicator in BLCA. This study imply that *POLR2K* overexpression in BLCA has profound impacts on tumor immune surveillance and on multiple steps of gene expression and of the cell cycle, thus contributing to immune surveillance evasion ultimately. *POLR2K* is particularly associated with some tumor-related transcription factors such as E2F1 and IRF family, miRNAs such as miRNA-145 and kinases such as HIPK2. Our study deploys current online websites to conduct bioinformatics analysis of bladder cancer. This strategy has superiorities in terms of simplicity and large sample size, which enables us to perform more large-scale *POLR2K* genomics research and functional studies free of charge, compared with classic chip screening.

Meanwhile, by using the TCGA database, we also face some limitations. The first one is that the BLCA samples in TCGA database include three ethnic group (Caucasian, African-american and Asian). Different genetic background of BLCA patients can influence gene expression profiles. The second limitation is that the BLCA samples is relatively small in stage 1(e.g., the relatively low number of severe cases), which may have limited the statistical significance of some correlations. Thus, adequate inclusion of each ethnic group and BLCA stages should be included in further prospective studies to know more about the generalisability of this study. The third one is that transcriptional sequencing could directly present information about protein level or protein activity, which should be tackled in subsequent studies with experiments. The fourth limitation is that the infiltration of tumor-associated normal cells has an profound influence on the analysis of clinical tumor samples by genomic methods, such as copy number

data or gene expression profiles, and therefore, biological interpretation of the analysis results deserves considerable attention considering sample heterogeneity, which could be the reason why the 50 most frequently altered neighbor genes enriched in the CD56+NKCells and 721_B_lymphoblasts (Supplementary Table 8).

5. Conclusion

This work reveal novel function of POLR2K as a regulator of cancer proliferation and present a potential strategy of treatment for BLCA. *POLR2K* expression might function as a novel therapeutic target for bladder cancer.

Abbreviations

Bladder Carcinoma (BLCA)

Copy number variations (CNVs)

Disease-free Survival (DFS)

E2F Transcription Factor 1(E2F1)

Gene Expression Omnibus (GEO)

Gene Ontology (GO)

Gene Expression Profiling Interactive Analysis (GEPIA)

Gene Set Enrichment Analysis (GSEA)

Interferon-α/β receptor (IFN)

NOD-Like Receptors (NLRs)

Overall Survival (OS)

RNA polymerase II subunit K POLR2K

Suppressor of Cytokine Signaling 7 (socs7)

The Cancer Genome Atlas (TCGA)

Toll-like Receptors (TLRs)

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Authors ' contributions

Liliang Yang and Kaizhen Wang designed and performed the research, analyzed and interpreted data. Liliang Yang and Wenjing Guo drafted the manuscript. Kaizhen Wang and Xian Chen participated in data analysis and figure preparation. Yuxin Zhou, Qinglong Guo and Libin Wei reviewed the manuscript. Liliang Yang and Kaizhen Wang conceived the study and participated in research design and data interpretation. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Babjuk, M., et al., *European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ)-2019 Update.* European Urology, 2019. **76**(5): p. 639-657.
2. Nadal, R. and J. Bellmunt, *Management of metastatic bladder cancer.* Cancer Treat Rev, 2019. **76**: p. 10-21.
3. Loriot, Y., et al., *Erdafitinib in Locally Advanced or Metastatic Urothelial Carcinoma.* N Engl J Med, 2019. **381**(4): p. 338-348.
4. Hindy, J.R., et al., *Targeted therapies in urothelial bladder cancer: a disappointing past preceding a bright future?* Future Oncol, 2019. **15**(13): p. 1505-1524.
5. Ding, Z.Y., et al., *Prognostic role of cyclin D2/D3 in multiple human malignant neoplasms: A systematic review and meta-analysis.* Cancer Med, 2019. **8**(6): p. 2717-2729.
6. Michiels, S., et al., *Genetic polymorphisms in 85 DNA repair genes and bladder cancer risk.* Carcinogenesis, 2009. **30**(5): p. 763-8.

7. Natrajan, R., et al., *An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers.* Breast Cancer Research and Treatment, 2010. **121**(3): p. 575-589.

8. Heidenblad, M., et al., *Tiling resolution array CGH and high density expression profiling of urothelial carcinomas delineate genomic amplicons and candidate target genes specific for advanced tumors.* BMC Med Genomics, 2008. **1**: p. 3.

9. Jung, C.R., S. Choi, and D.S. Im, *The NS5A protein of hepatitis C virus represses gene expression of hRPB10alpha, a common subunit of host RNA polymerases, through interferon regulatory factor-1 binding site.* Virus Res, 2007. **129**(2): p. 155-65.

10. Giurato, G., et al., *Quantitative mapping of RNA-mediated nuclear estrogen receptor beta interactome in human breast cancer cells.* Sci Data, 2018. **5**: p. 180031.

11. Lin, Y., et al., *An in-depth map of polyadenylation sites in cancer.* Nucleic Acids Res, 2012. **40**(17): p. 8460-71.

12. Rhodes, D.R., et al., *ONCOMINE: A cancer microarray database and integrated data-mining platform.* Neoplasia, 2004. **6**(1): p. 1-6.

13. Cerami, E.G., et al., *Pathway Commons, a web resource for biological pathway data.* Nucleic Acids Research, 2011. **39**: p. D685-D690.

14. Chandrashekar, D.S., et al., *UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses.* Neoplasia, 2017. **19**(8): p. 649-658.

15. Cerami, E., et al., *The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data (vol 2, pg 401, 2012).* Cancer Discovery, 2012. **2**(10): p. 960-960.

16. Vasaikar, S., et al., *LinkedOmics: Analyzing multi-omics data within and across 32 cancer types.* Cancer Research, 2019. **79**(13).

17. Tang, Z.F., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* Nucleic Acids Research, 2017. **45**(W1): p. W98-W102.

18. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.* Nucleic Acids Research, 2010. **38**: p. W214-W220.

19. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Research, 2016. **44**(W1): p. W90-W97.

20. Wang, M.Q., et al., *POLR1D promotes colorectal cancer progression and predicts poor prognosis of patients.* Molecular Carcinogenesis, 2019. **58**(5): p. 735-748.

21. Cui, J.R., A. Gizzi, and J.T. Stivers, *Deoxyuridine in DNA has an inhibitory and promutagenic effect on RNA transcription by diverse RNA polymerases.* Nucleic Acids Research, 2019. **47**(8): p. 4153-4168.

22. Oeyen, E., et al., *Bladder Cancer Diagnosis and Follow-Up: The Current Status and Possible Role of Extracellular Vesicles.* International Journal of Molecular Sciences, 2019. **20**(4).

23. Kohsaka, S., et al., *Comprehensive assay for the molecular profiling of cancer by target enrichment from formalin-fixed paraffin-embedded specimens.* Cancer Science, 2019. **110**(4): p. 1464-1479.

24. Cheng, L.J., et al., *Integration of genomic copy number variations and chemotherapy-response biomarkers in pediatric sarcoma.* Bmc Medical Genomics, 2019. **12**.

25. Gutierrez-Arcelus, M., et al., *Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions.* Nature Communications, 2019. **10**.

26. Carles, C., et al., *Two additional common subunits, ABC10 alpha and ABC10 beta, are shared by yeast RNA polymerases.* J Biol Chem, 1991. **266**(35): p. 24092-6.

27. Knowles, M.R., et al., *Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms.* Am J Hum Genet, 2013. **93**(4): p. 711-20.

28. Reich, C., et al., *The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable in vivo and in vitro.* Molecular Microbiology, 2009. **71**(4): p. 989-1002.

29. Gross, S., et al., *Targeting cancer with kinase inhibitors.* Journal of Clinical Investigation, 2015. **125**(5): p. 1780-1789.

30. Ke, C.Y., et al., *IRF6 and TAK1 coordinately promote the activation of HIPK2 to stimulate apoptosis during palate fusion.* Science Signaling, 2019. **12**(593).

31. Cao, L.L., et al., *HIPK2 is necessary for type I interferon-mediated antiviral immunity.* Science Signaling, 2019. **12**(573).

32. Stecca, B. and E. Rovida, *Impact of ERK5 on the Hallmarks of Cancer.* International Journal of Molecular Sciences, 2019. **20**(6).

33. Thakur, A., H. Mikkelsen, and G. Jungersen, *Intracellular Pathogens: Host Immunity and Microbial Persistence Strategies.* Journal of Immunology Research, 2019.

34. Kuwano, Y., et al., *Homeodomain-Interacting Protein Kinase-2: A Critical Regulator of the DNA Damage Response and the Epigenome.* International Journal of Molecular Sciences, 2016. **17**(10).

35. Fouad, Y.A. and C. Aanei, *Revisiting the hallmarks of cancer.* American Journal of Cancer Research, 2017. **7**(5): p. 1016-1036.

36. Alenzi, F.Q.B., *Links between apoptosis, proliferation and the cell cycle.* British Journal of Biomedical Science, 2004. **61**(2): p. 99-102.

37. Richter, C., et al., *Rewiring E2F1 with classical NHEJ via APLF suppression promotes bladder cancer invasiveness.* Journal of Experimental & Clinical Cancer Research, 2019. **38**.

38. Lee, J.S., et al., *Expression Signature of E2F1 and Its Associated Genes Predict Superficial to Invasive Progression of Bladder Tumors.* Journal of Clinical Oncology, 2010. **28**(16): p. 2660-2667.

39. Jin, N., et al., *Screening biomarkers of bladder cancer using combined miRNA and mRNA microarray analysis.* Molecular Medicine Reports, 2015. **12**(2): p. 3170-3176.

40. Portou, M.J., et al., *The innate immune system, toll-like receptors and dermal wound healing: A review.* Vascular Pharmacology, 2015. **71**: p. 31-36.

41. Ben-Neriah, Y. and M. Karin, *Inflammation meets cancer, with NF-kappaB as the matchmaker.* Nat Immunol, 2011. **12**(8): p. 715-23.

42. Chow, M.T., A. Moller, and M.J. Smyth, *Inflammation and immune surveillance in cancer.* Seminars in Cancer Biology, 2012. **22**(1): p. 23-32.

43. Catto, J.W.F., et al., *MicroRNA in Prostate, Bladder, and Kidney Cancer: A Systematic Review.* European Urology, 2011. **59**(5): p. 671-681.

44. Piao, X.M., et al., *Urinary cell-free microRNA biomarker could discriminate bladder cancer from benign hematuria.* International Journal of Cancer, 2019. **144**(2): p. 380-388.

45. Zhang, L.Q., et al., *miR-202 Inhibits Cell Proliferation, Migration, and Invasion by Targeting Epidermal Growth Factor Receptor in Human Bladder Cancer.* Oncology Research, 2018. **26**(6): p. 949-957.

46. Chen, X.L., et al., *MicroRNA-374a Inhibits Aggressive Tumor Biological Behavior in Bladder Carcinoma by Suppressing Wnt/beta-Catenin Signaling.* Cellular Physiology and Biochemistry, 2018. **48**(2): p. 815-826.

47. Xu, F.F., H. Li, and C.J. Hu, *MiR-202 inhibits cell proliferation, invasion, and migration in breast cancer by targeting ROCK1 gene.* Journal of Cellular Biochemistry, 2019. **120**(9): p. 16008-16018.
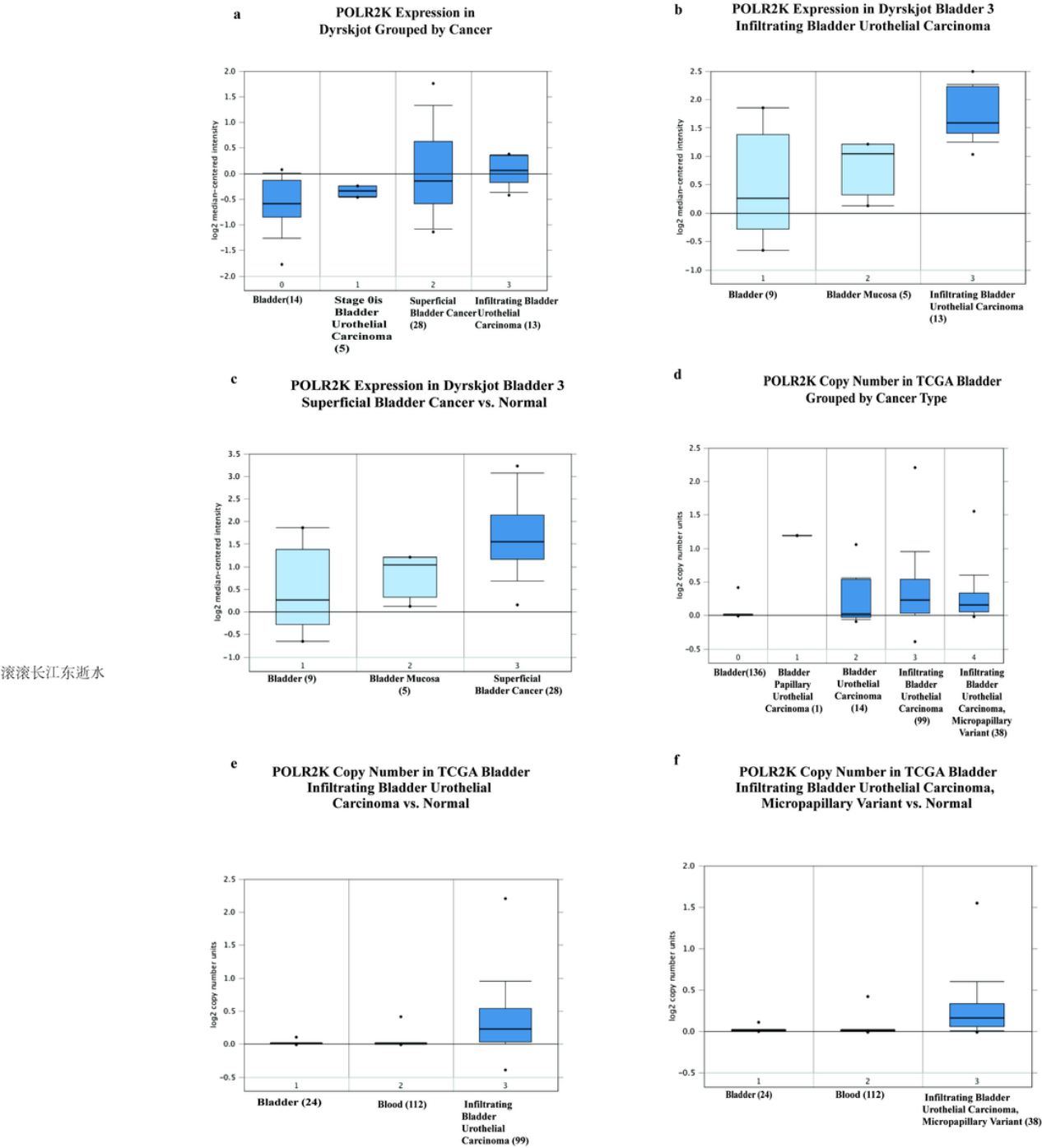
Figures

滚滚长江东逝水

Figure 1

POLR2K transcription in bladder cancer (Oncomine). Levels of POLR2K mRNA and DNA copy number were remarkably higher in bladder urothelial carcinoma than in normal tissues. (a) Box plot showing POLR2K mRNA

levels in Dyrskjot dataset. Threshold by: p-value 10e-4, fold change 2, gene rank top 10%. (b-c) Box plot showing POLR2K mRNA levels of either Infiltrating Bladder Urothelial Carcinoma or Superficial Bladder Cancer versus normal bladder tissue. For figure b: p-value 3.72e-5, fold change 2.211, gene rank top 8%; for figure c: p-value 6.43e-5, fold change 2.178, gene rank top 17%. (d) Box plot showing POLR2K copy number in The Cancer Genome Atlas (TCGA) Bladder datasets. Threshold by: p-value 10e-4, fold change 2, gene rank top 10%. (e-f) Box plot showing POLR2K copy number of either Infiltrating Bladder Urothelial Carcinoma or Infiltrating Bladder Urothelial Carcinoma, Micropapillary Variant versus normal bladder tissue. For figure e: p-value 5.11e-12, fold change 1.218, gene rank top 4%; for figure f: p-value 1.23e-5, fold change 1.179, gene rank top 9%.
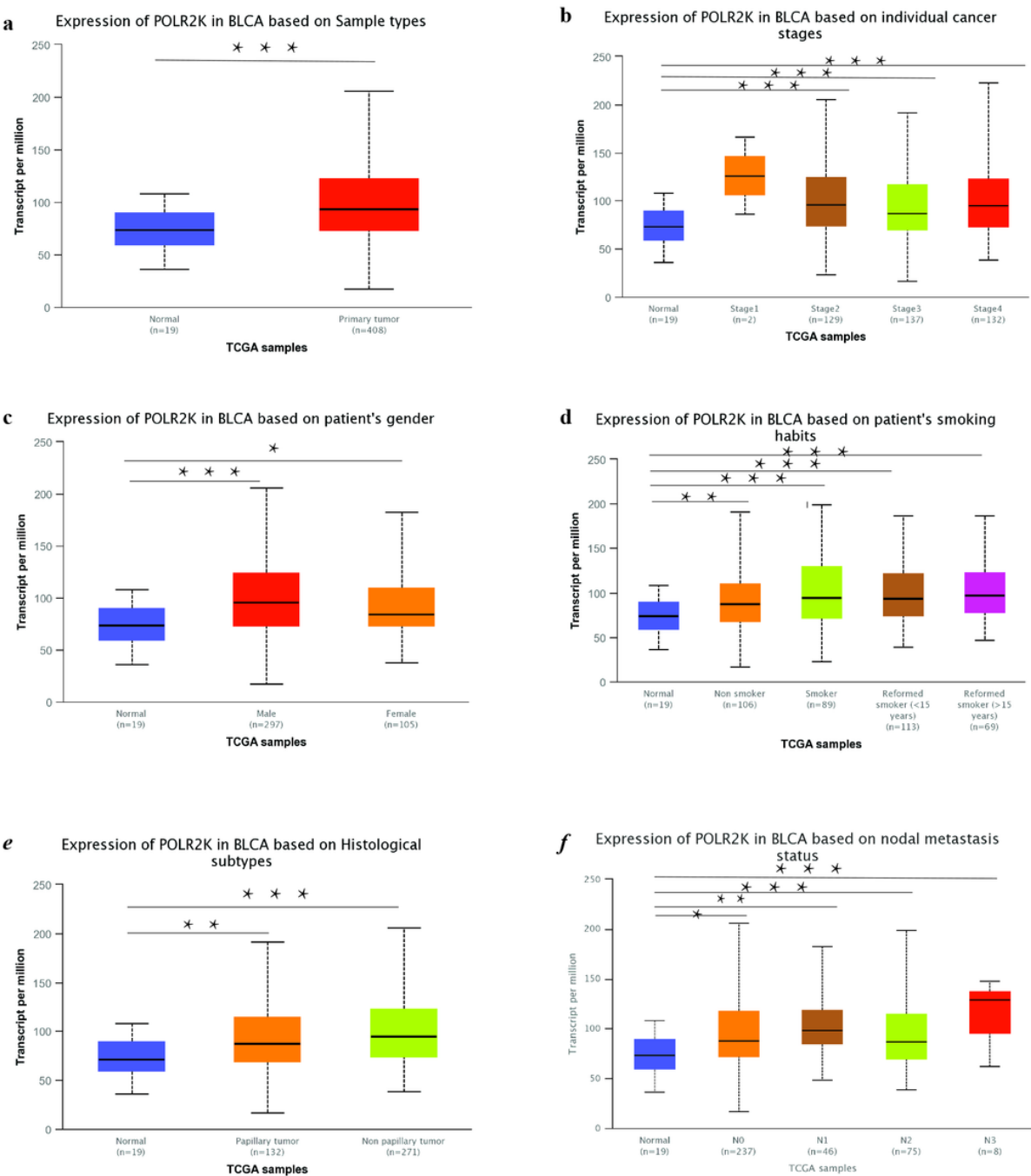

Figure 2

POLR2K transcription in subgroups of patients with bladder cancer, stratified based on disease stage, gender and other criteria (UALCAN). (a) Boxplot displaying relative level of POLR2K in normal and BLCA tissue. (b) Boxplot displaying relative level of POLR2K in normal individuals or in BLCA patients in stages 1, 2, 3 or 4. (c) Boxplot displaying relative level of POLR2K in normal individuals of either gender or male or female BLCA patients. (d) Boxplot displaying relative levle of POLR2K in normal individuals or in BLCA patients of Non-smoker, Smoker, Reformed smoker (<15 years) and Reformed smoker (>15 years). (e) Boxplot displaying relative levle of POLR2K in normal individuals or in BLCA patients of papillary and non-papillary tumor. (f) Boxplot displaying relative levle of POLR2K in normal individuals or BLCA patients with nodal metastasis N0, N1, N2 and N3 status. Data are mean ± SE. *, P < 0.05; **, P < 0.01; ***, P < 0.001.

Figure 3

Visual summary of POLR2K alterations and biological interaction webwork in BLCA. (a) OncoPrint of POLR2K alterations in BLCA. The OncoPrint overviews genomic alterations in POLR2K influencing individual samples (columns) in BLCA from the TCGA. The distinct categories of genetic alterations are featured in distinct colours. (b) Network view of the POLR2K neighborhood in BLCA. To visualize the neighborhood gene of POLR2K, Pathway Commons (https://www.pathwaycommons.org) and Cytoscape were used. POLR2K is used as a seed gene, and all other genes are automatically identified as altered in BLCA. Darker blue indicates increased frequency of alteration in BLCA.

Figure 4

Enrichment analysis of the genes altered in the POLR2K neighborhood BLCA. With the help of Cytoscape, the colorful pictures display the enrichment results of the top 50 genes altered in the POLR2K neighborhood in BLCA. (a) Cellular components. (b) Biological processes. (c) Molecular functions. (d) KEGG pathway analysis. Nodes represent ontology terms or KEGG pathways that were significantly enriched for the predicted target genes. Edges represent connections within the ontology tree and colors highlight terms correlated in meaning. The most significant terms of
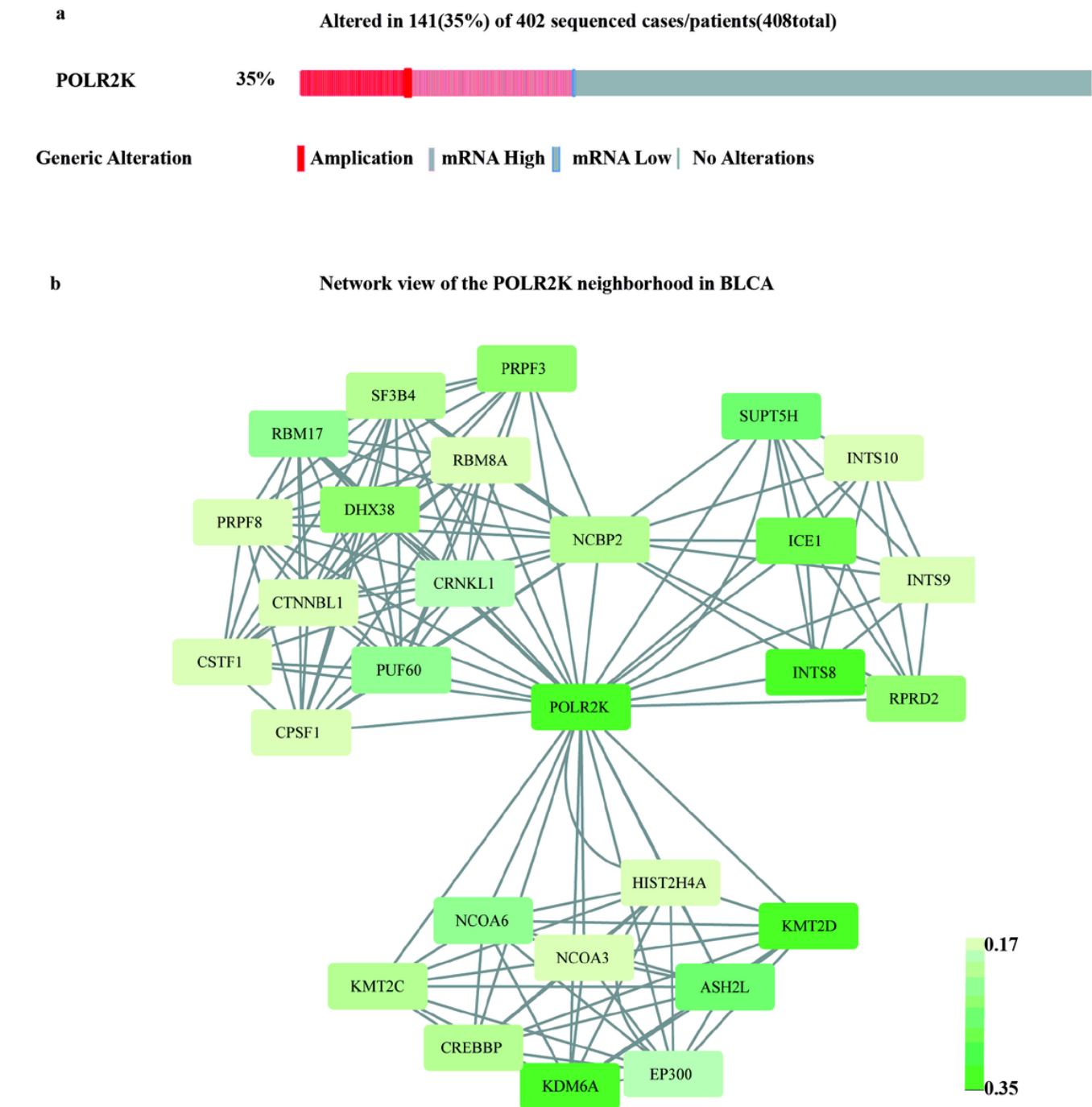
the group showing a biological meaning in association with the system under analysis are captioned. A p-value of < 0.01 was considered statistically significant.
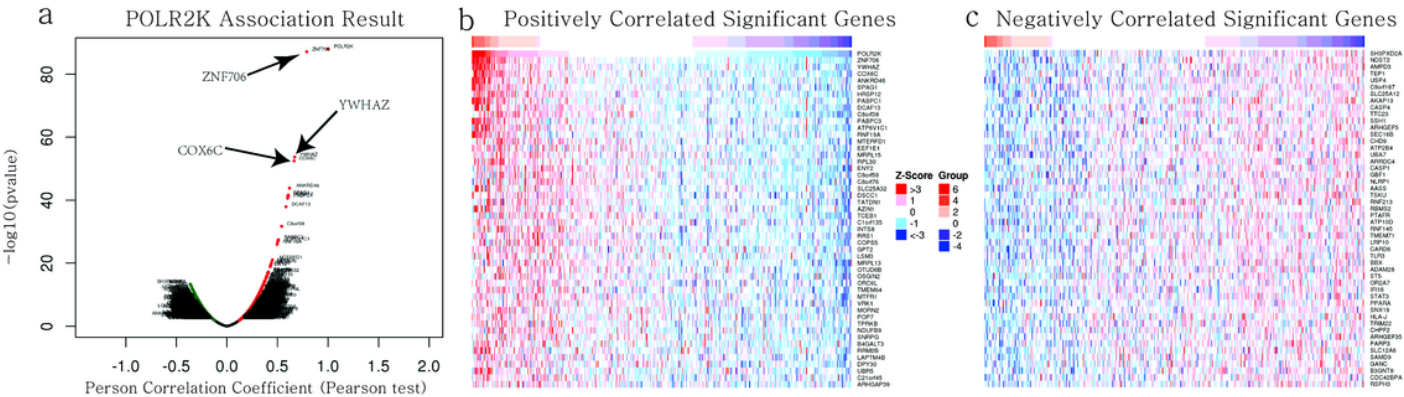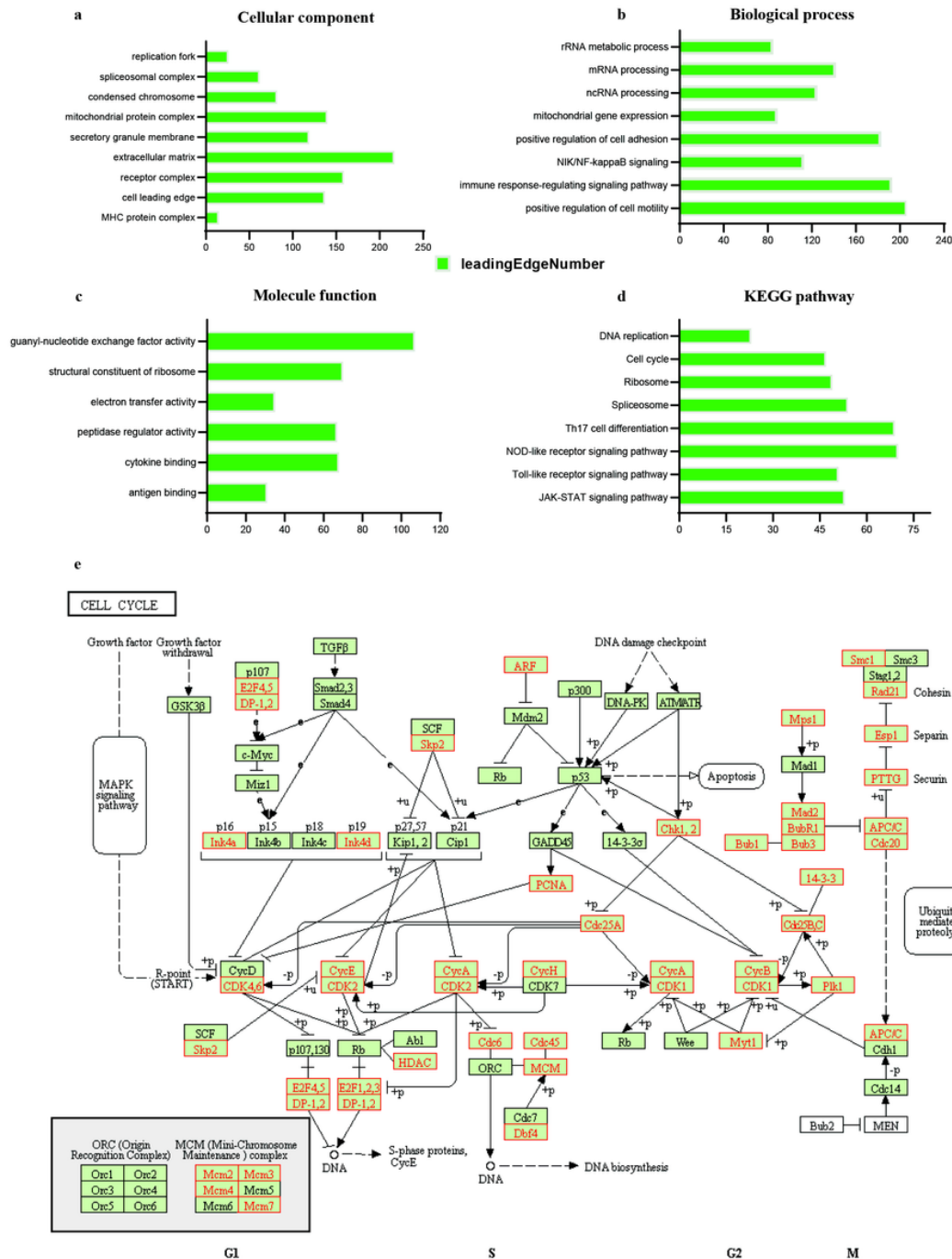

Figure 5

Genes differentially expressed in correlation with POLR2K in BLCA (http://linkedomics.org/). (a) A Pearson test was deployed to analyze relationships between POLR2K and genes differentially expressed in BLCA. (b–c) Heat maps dispalying genes positively and negatively correlated with POLR2K in BLACA (TOP 50). Red indicates positively correlated genes and green indicates negatively correlated genes.

Figure 6

Significantly enriched GO annotations and KEGG pathways of POLR2K in BLCA. The enriched GO annotations and KEGG pathways of POLR2K co-expression genes in BLCA were obtained by GSEA. (a) Cellular components. (b) Biological processes. (c) Molecular functions. (d) KEGG pathway analysis. The blue column represents the LeadingEdgeNum, and the FDR< 0.05. (e) KEGG pathway annotations of the cell cycle pathway obtained from https://www.genome.jp/kegg. Red marked nodes are associated with the LeadingEdgeGene.
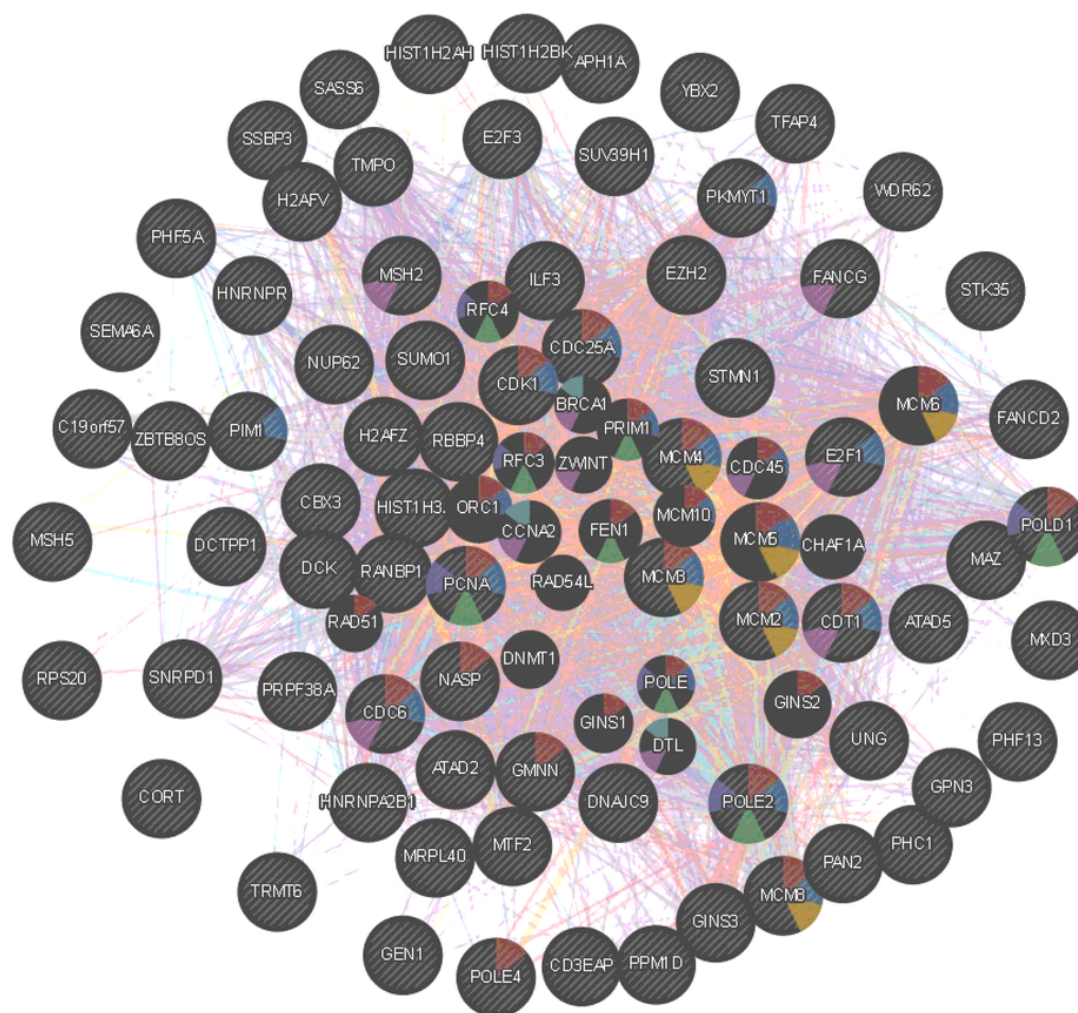
Figure 7

Protein-protein interaction network of transcription factor E2F1-target networks (GeneMANIA). Protein-protein interaction (PPI) network and functional analysis indicating the gene set that was enriched in the target network of transcription factor E2F1. Different colors of the network edge indicate the bioinformatics methods applied: co-expression, website prediction, pathway, physical interactions and co-localization. The different colors for the network nodes indicate the biological functions of the set of enrichment genes.

Supplementary Files

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

- SupplementaryMaterial.docx