Comprehensive Gene Expression Analysis in NMIBC Using RNA-seq Reveals New Therapy Strategies

Xiaoliang Chen[1†], Fuquan Jiang[1†], Chunshu Jia[2], Ming Liu[3], Yonghao Nan[4], Licheng Qu[2], Qingkuo Kong[2], Fangfang Hou[1], Wenshan Luo[1], Wanli Na[1], Xuefei Jin[1*] and Jiufeng Tan[1*]

[1] Department of Urology, China-Japan Union Hospital of Jilin University, Changchun, China, [2] Centre for Reproductive Medicine, Centre for Prenatal Diagnosis, First Hospital of Jilin University, Changchun, China, [3] Central People's Hospital of Siping City, Siping, China, [4] Department of Urology, the Frist Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Non-muscle invasive bladder cancer (NMIBC) patients often have fewer treatment options, and suffer the progression of disease due to mechanisms that are not clear, as well as due to its diversity. This study was designed to explore the molecular mechanism of bladder cancer through an RNA-seq. In addition to conventional analyses, we also simplified the network through modularization using the WGCNA algorithm, with the help of the topological overlapping matrix and hierarchical cluster tree, which are based on the PPI network of STRING. Furthermore, the hub genes were confirmed through survival analyses in the independent cohorts ($n = 431$). Among them, 15 genes were significantly associated with poor prognosis. Finally, we validated the results at mRNA and protein level using qRT-PCR, IHC and western blotting. Taken together, our research is important for the prediction, as well as the prospective clinical development of drug targets and biomarkers.

Keywords: NMIBC, modularization, WGCNA, PPI, biomarker

INTRODUCTION

Bladder cancer is a prevalent disease among the world, which is mainly attributed to smoking (1). Men are more likely to be affected than women, and morbidity increases with age (2). Bladder cancer grows through two distinct pathways: non-muscle invasive type and muscle invasive type. The majority of bladder patients are diagnosed through macroscopic hematuria, and diagnosis is confirmed after surgical resection, which is the primary stage of treatment (3). Although the 5 years survival rate as a result of current therapies is more than 80% for non-muscle invasive bladder cancer (NMIBC) patients, while a recurrence rate of nearly 70% results in patients being under lifelong surveillance and makes NMIBC the most expensive cancer from diagnosis to death (4, 5). Therefore, new therapeutic strategies are necessary to overcome these challenges.

Recent studies of bladder cancer based on gene expression profiles have gradually elucidated the molecular mechanism of the disease (6). Except for the transcriptional features have been discovered through earlier traditional microarrays, many molecular characteristics have been identified through integrative analyses (7, 8). Indeed, a lot of putative biomarkers and drug targets, including FGFR3, VEGF, CEBPA, and CCNE1, have been identified in different investigations (9–12). However, none of them have probed the regulation of their expression and of their associated genes, which could provide an extra perspective into the molecular mechanisms of disease progression.

In this study, we performed an RNA-seq on non-muscle bladder cancer patients who were subjected to surgical resection. After data processing, the reads were aligned with hg38 using STAR, and DESeq was used to filter the differentially expressed genes (DEGs). Then, we used these DEGs to execute functional annotation, including GO and KEGG enrichment analyses, which were further validated using GSEA. In order to simplify the network and identify functional clusters, modularization analysis was established through WGCNA and integrated with the PPI network from STRING, in order to model the dynamics of proteome changes. Afterwards, survival analysis was used to assess the clinical outcomes of hub genes, which are located as connections in each module. Finally, the results were confirmed through experiments, and it is hoped that they may be used as a reference for gene therapy for bladder cancer.

RESULTS

Identification of Differentially Expressed Genes (DEGs) and Functional Variation

DEGs were screened out using the DESeq package depending on read counts at transcription level, which were identified using an absolute $\log_2FC$ value >1 and adjusted p-value of $<10^5$ as the statistical conditions for filtering. We obtained 885 DEGs, including 54 significantly upregulated genes and 831 significantly downregulated genes between the bladder cancer tissue and adjacent tissue. A volcano plot was used to visualize the results, in which hub genes and significantly changed genes were indicated (**Figure 1A**).

In order to further analyze the DEGs, we explored functional variation between the two groups using the clusterProfiler package. 430 GO terms were identified with an adjusted p-value of <0.01. The GOSemSim package was used to remove similar terms by keeping only one representative term, which resulted in 142 unique GO terms ([13]). The top 20 are shown in **Figure 1C**. Even though the extracellular matrix and organization structure were the most statistically changed functions, many terms involved in the immune response and activation, such as the regulation of lymphocyte activation, regulation of humoral immune response, adaptive immune response, and T cell activation, were also found. KEGG analysis also revealed many DEGs related to the downstream pathways of immune activation (**Figure 1B**), such as the PI3K-Akt signaling pathway, MAPK signaling pathway, and NF-kappa B signaling pathway. Most DEGs were downregulated in these processes. Recently, an unsupervised clustering by cytogenetic analysis divided NMIBC into two subtypes, no cytogenetic changes subtype (genomic subtype1, GS1) and another subtype with loss of 9q in chromosome (GS2). GS2 often appear in high grade tumors, and loss some regulators of AKT/PI3K/mTOR pathway. This may be why the dysfunction of AKT/PI3K pathway in NMIBC ([14]). In order to further verify the relationship between phenotype and functionally changed genes, we performed GSEA on the whole genome at transcription level. The transcripts of bladder cancer were remarkably associated with downregulated genes related to T and B cell receptor signaling pathways and

their downstream pathways, which is in accordance with GO and KEGG enrichment analysis (**Figures 1D,E**).

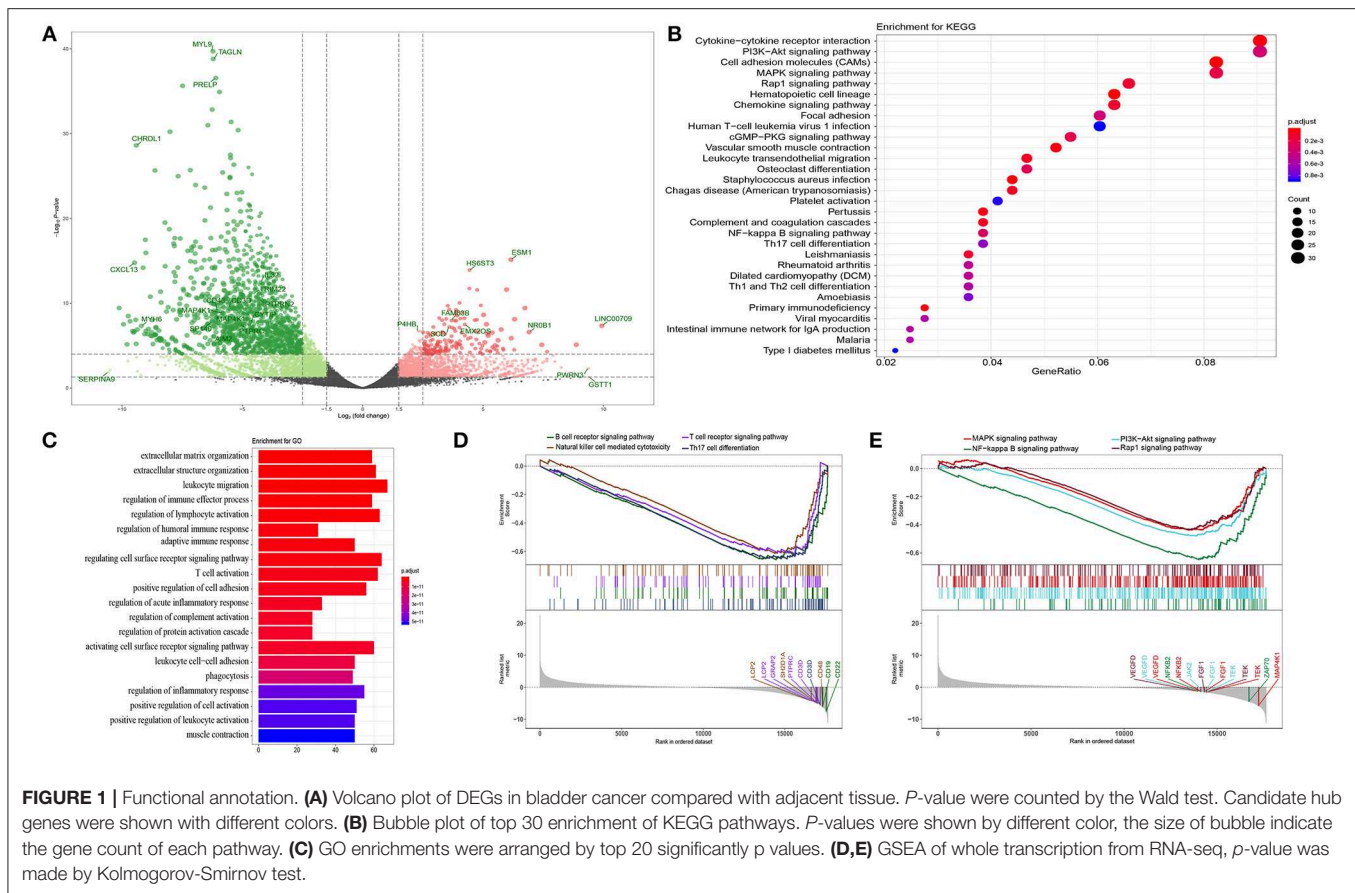### Integrative Network Analysis Reveals New Functional Modules

An integrative analysis method was used to model the dynamics of proteome changes upon cancer progression, as described previously ([15]). In brief, we applied WGCNA to all DEGs in order to cluster the correlative proteins that had similar molecular functions or biological processes ([16]). Later, these proteins were superimposed onto the PPI network in order to identify functional modules. As a result, we identified 132 modules, with the number of proteins ranging from 17 to 2 (**Figure 2B**), and 117 out of the 132 modules were highly interconnected through their members (**Figure 2A**). Each module was annotated by known functional terms or signaling pathways. For instance, the modules were remarkably enriched in the immune reaction system including the T cell-mediated immune response (module 1, 13, 26, and 34), B cell-mediated immune response (module 11, 28, 34, and 38), mast cell activation (Module 6, 10, and 51) and natural killer cell mediated immunity (Module 39, 41, 47, and 75). Furthermore, some of the modules involved in cell invasion and migration processes also contributed to the progression of tumorigenesis, as commonly known, through mechanisms such as extracellular matrix organization (Module 4, 72) and the integrin-mediated signaling pathway (Module 32, 51). In summary, progression of bladder cancer is through the rebalanced regulation and extensive reprogramming of mutually connected functional modules.

Survival Analysis of Hub Genes

Based on the expression profile and clinical data of 431 bladder cancer samples from TCGA database, the clinical outcomes of hub genes that are indicated in **Figures 1**, **2** were evaluated through survival analysis. 15 out of the 62 hub genes were significantly associated with poor prognosis, and were either positively or negatively correlated with a higher risk and were either upregulated or downregulated with bladder cancer (**Figure 3**). Among the hub genes, CD3D was the core factor of the network, which was involved in the T cell receptor signaling pathway and was connected to T cell and mast cell activation. We computed the Pearson correlation of CD3D using 26,483 transcripts of 431 bladder cancer patients. CD2, CD6, and UBASH3A were the most positively correlated genes, while CD3D, and SCAMP1, MARVELD2, and KDM5B were the most negatively correlated. These genes might also be involved in the regulation of bladder cancer progression, and might also be candidate biomarkers or drug targets for the disease.

Initial Validation of Transcriptome Results Using qRT-PCR and IHC

In order to confirm the DEGs found through the experiment, total RNA of 24 paired tumor tissues were isolated for qRT-PCR validation. Twenty six target DEGs were selected as shown in **Figure 4**. Moreover, IHC was also performed to further validate the five target genes of patients who underwent surgical resection (**Figure 5**). In brief, the DEGs were successfully validated and

FIGURE 1 | Functional annotation. **(A)** Volcano plot of DEGs in bladder cancer compared with adjacent tissue. *P*-value were counted by the Wald test. Candidate hub genes were shown with different colors. **(B)** Bubble plot of top 30 enrichment of KEGG pathways. *P*-values were shown by different color, the size of bubble indicate the gene count of each pathway. **(C)** GO enrichments were arranged by top 20 significantly p values. **(D,E)** GSEA of whole transcription from RNA-seq, *p*-value was made by Kolmogorov-Smirnov test.

showed good correspondence with the analysis of transcriptome, indicating that the RNA-seq results were precise and reliable.

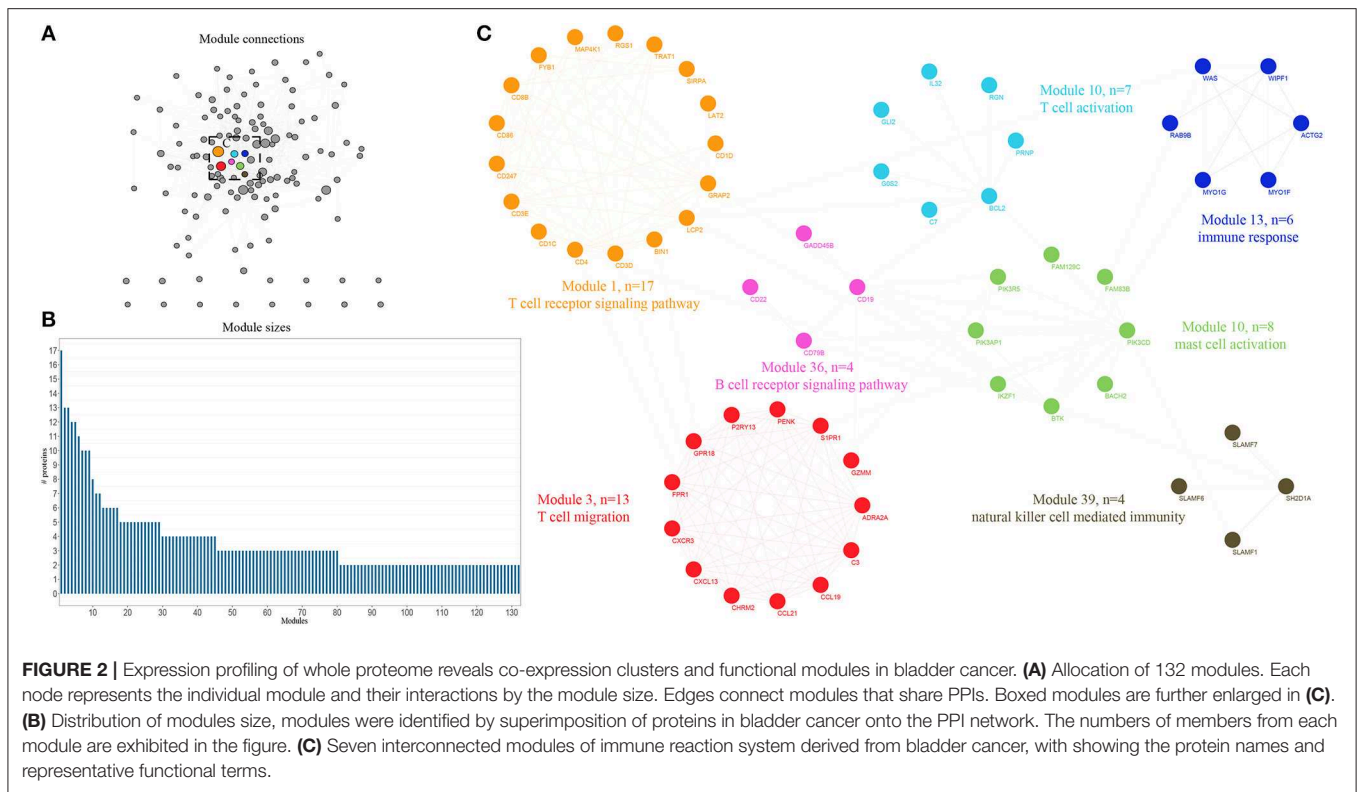## Signaling Pathway Validation Using Western Blotting

Finally, we wanted to confirm the signaling pathways at protein level. MAPKs are evolutionarily conserved kinases, ubiquitously expressed and regulate a wide range of biological processes, such as cell growth, differentiation and death (17, 18). In cancer, the MAPK signaling pathway can play a double role by either maintaining cell survival or impelling cell death, through different mechanisms (19). In this study, we found that Fibroblast growth factor receptor 1 (FGFR1), which is amplified in lung and breast cancer, was downregulated in bladder cancer samples compared with that of the controls (20, 21). FGFR1 genes are fused to TACC1 through interstitial deletions, which were also downregulated in our results (log$_2$FC = −0.91). The other three genes of the MAKP signaling pathway, PKCα, p21 Ras, and c-Fos, followed the same trend as that of FGFR1. More strikingly, protein phosphatase HePTP, which is a negative regulatory factor, also performed a similar action (**Figure 6**).

DISCUSSION

It is well-known that bladder cancer is the 11th most malignant tumor worldwide, and 70% of patients present with NMIBC.

However, the exact biological functional variation during the progression of bladder cancer is still obscure. In order to provide deeper insights into the molecular mechanism involved in this process, we performed an RNA-seq on three paired bladder cancer patients who underwent surgical resection at China-Japan Union Hospital of Jilin University, and made a comprehensive analysis of the results, together with data from TCGA database. We identified core DEGs, significant biological processes, pathways, and validated our results using qRT-PCR, IHC and western blotting. In general, our work revealed an interlaced network presented by central modules that are involved in bladder cancer development, in which hub genes may play an indispensable role.

We sought for the expression patterns of transcripts and functional variations between bladder cancer tissue and adjacent tissues using RNA-seq, which produced a massive amount of data. In order to extract useful information from the large amount of data to explain the molecular mechanism of bladder cancer, in our study, we focused on two concepts. First, the DEGs were annotated by GO and KEGG pathway analyses, and the results involved functions related with immunity, cell adhesion and cancer. Furthermore, GSEA provided a good method of validating the functional annotations of the whole genome at transcription level rather than the DEGs. We also deciphered the complex network through modularization using WGCNA superimposed onto the PPI database of STRING. Each

FIGURE 2 | Expression profiling of whole proteome reveals co-expression clusters and functional modules in bladder cancer. **(A)** Allocation of 132 modules. Each node represents the individual module and their interactions by the module size. Edges connect modules that share PPIs. Boxed modules are further enlarged in **(C)**. **(B)** Distribution of modules size, modules were identified by superimposition of proteins in bladder cancer onto the PPI network. The numbers of members from each module are exhibited in the figure. **(C)** Seven interconnected modules of immune reaction system derived from bladder cancer, with showing the protein names and representative functional terms.

module was facilitated through the hierarchical cluster tree and topological overlapping matrix, which echoed the annotated functions of GO and KEGG. Overall, the complicated network was simplified by modularization into modules, which made it easier for it to be learned by hub genes that were the connections among the modules. Second, the bladder cancer dataset obtained from TCGA was used to evaluate the clinical significance of the hub genes. Fifteen hub genes, including five upregulated and 10 downregulated, were associated with overall survival of patients, which indicates poor prognosis of bladder cancer. Among the hub genes, CD3D attracted our attention due to its location on the most important module. Pearson correlation was used to find the co-expression of CD3D and the expression pattern was assessed. Finally, partial hub genes were validated using qRT-PCR and IHC on specimens from the bladder cancer patients.

Along with the development and application of NGS technologies, a large number of sequencing data has been accumulated. However, we should be conscious of an analytical system that is so sophisticated that it is above our initial cognition. Fortunately, modern methodologies have provided us with a good way of simplifying complex networks, which include thousands of proteins that can be disassembled into several independent and correlated modules, and the hub genes of each module can be probed in detail. The active application of public databases promotes the elucidation of gene functions. As mentioned above, our study clearly presents the significant biological modules, pathways and hub genes involved in the progression of bladder cancer. However, some core genes might
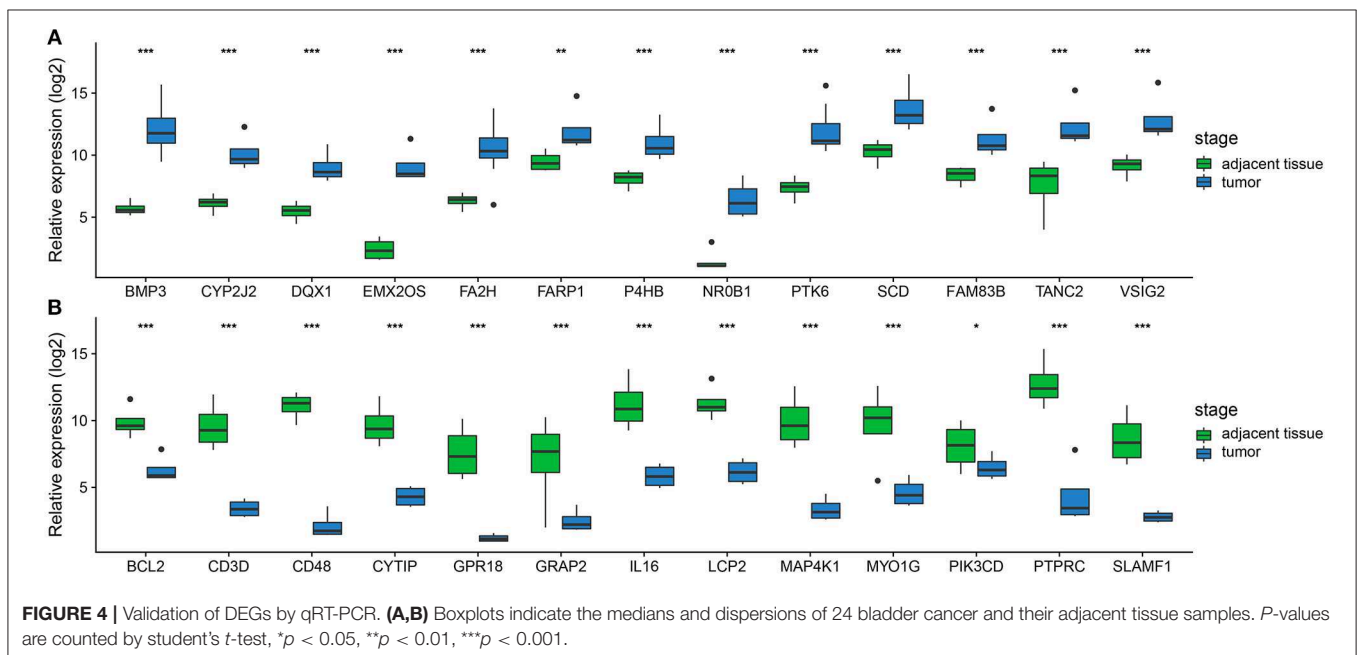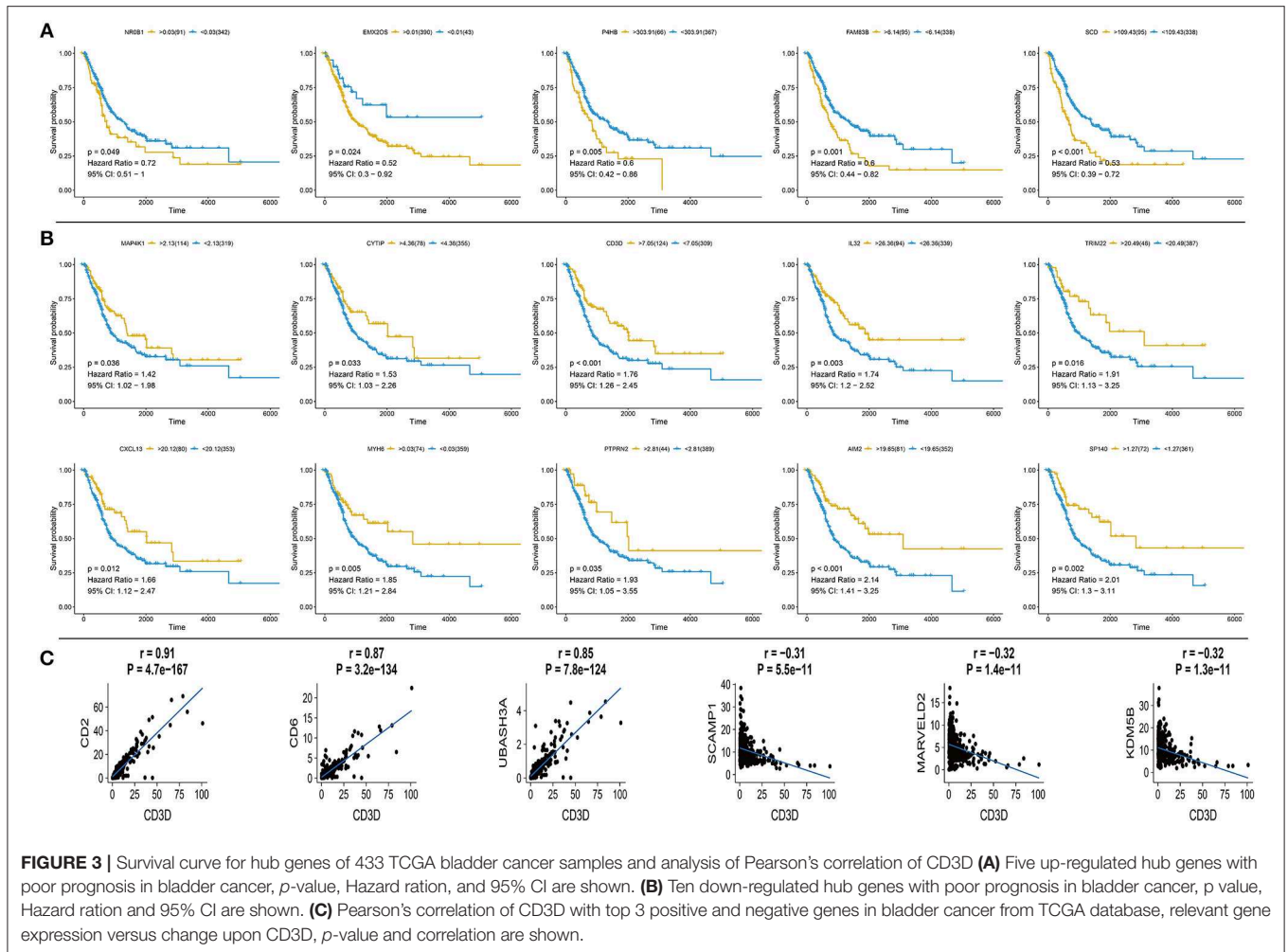
be shut out if they fail to be positioned in the modules, or have not been filtered out as DEGs, which may play an important role even though their expression does not greatly change during cancer progression. Therefore, we may have missed these genes in our analysis. Taken together, we systematically analyzed the molecular mechanism of functional variation in bladder cancer through biological modules and hub genes, which were confirmed using qRT-PCR, IHC, and western blotting. The revelation that they are involved in tumor progression could be used to design new strategies to treat aggressive carcinoma. For example, the downregulation of CD3D in bladder cancer samples and the T-cell receptors that are essential for the activation of T cell signaling, indicate a new therapeutic approach for bladder cancer. In addition, similarly, other hub genes may also prove to be useful drug targets and prognostic markers in gene therapies.
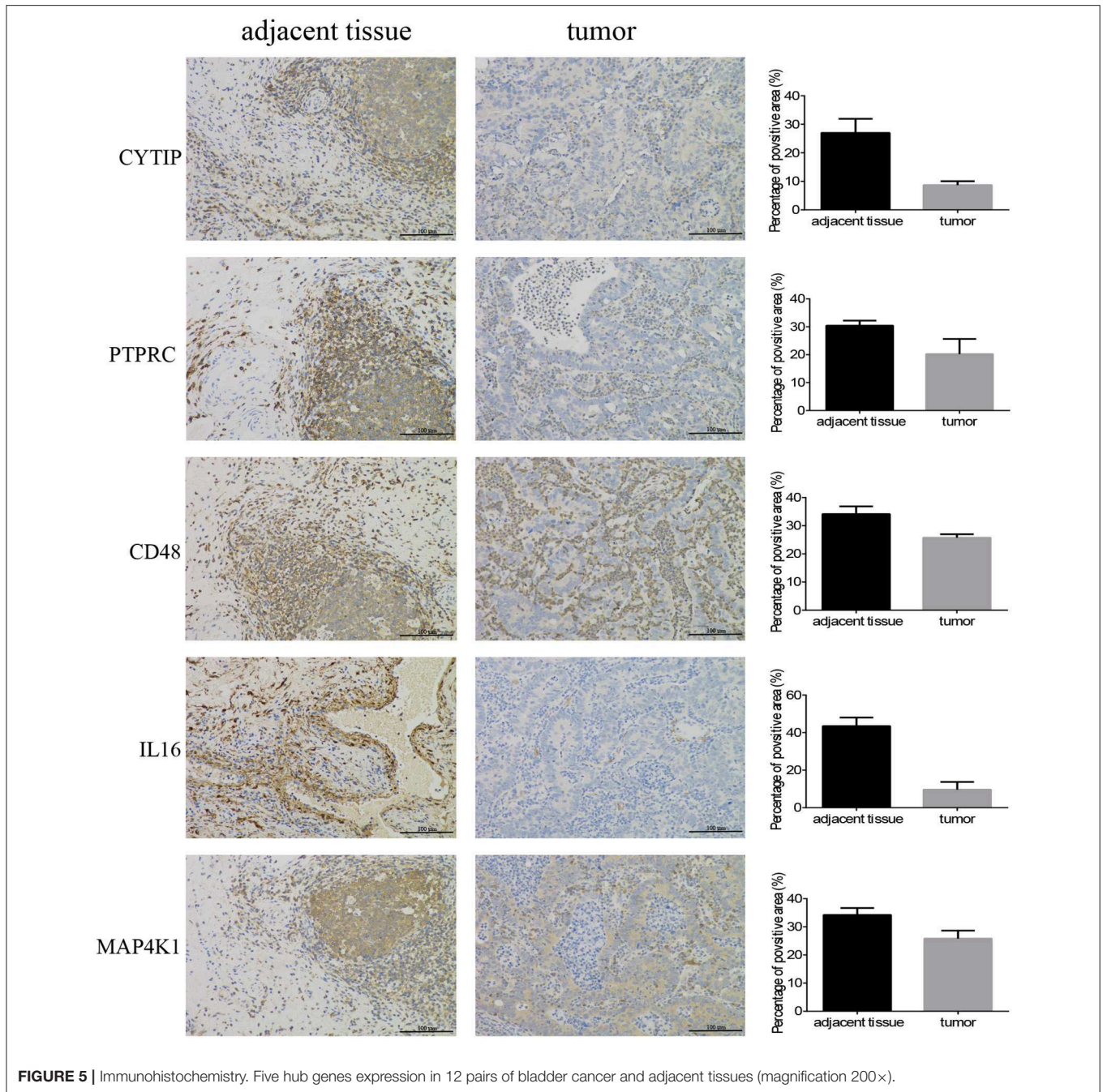
METHODS

Patients and Samples

All specimens were obtained from bladder cancer patients between April 2016 and December 2017 at China-Japan Union Hospital of Jilin University (Changchun, China), with the approval of the Ethics Committee. The samples were surgically resected followed by being treated with liquid nitrogen, and were then stored at −80°C. According to routine procedure, all samples were assessed using HE staining and diagnosis was made by three independent pathologists.

FIGURE 3 | Survival curve for hub genes of 433 TCGA bladder cancer samples and analysis of Pearson's correlation of CD3D **(A)** Five up-regulated hub genes with poor prognosis in bladder cancer, *p*-value, Hazard ration, and 95% CI are shown. **(B)** Ten down-regulated hub genes with poor prognosis in bladder cancer, p value, Hazard ration and 95% CI are shown. **(C)** Pearson's correlation of CD3D with top 3 positive and negative genes in bladder cancer from TCGA database, relevant gene expression versus change upon CD3D, *p*-value and correlation are shown.

FIGURE 4 | Validation of DEGs by qRT-PCR. **(A,B)** Boxplots indicate the medians and dispersions of 24 bladder cancer and their adjacent tissue samples. *P*-values are counted by student's *t*-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

FIGURE 5 | Immunohistochemistry. Five hub genes expression in 12 pairs of bladder cancer and adjacent tissues (magnification 200×).
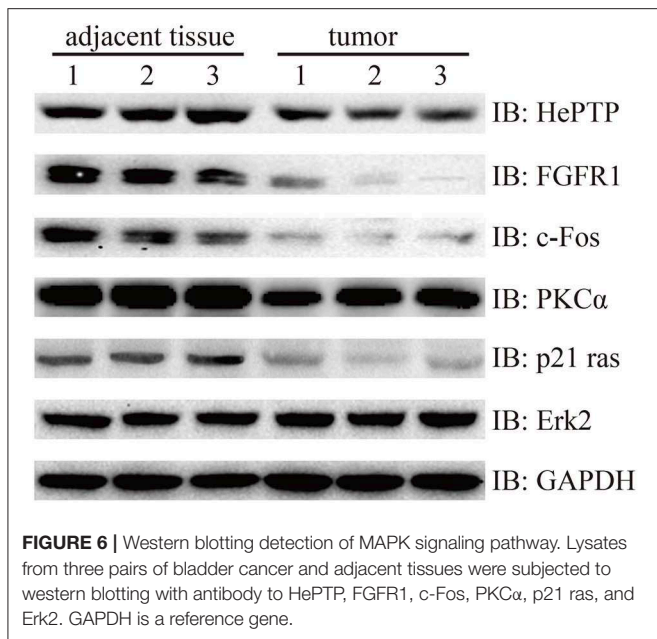
RNA-Seq and Data Processing

Total RNA of the three bladder cancer tissues and their paired adjacent tissues were isolated using Trizol reagent (Sangon Biotech, Shanghai, China), as described in other studies. The cDNA libraries were constructed using a custom protocol, which were sequenced using the Illumina Hiseq 2500 sequencer (Sangon Biotech, Shanghai, China). Raw data were uploaded to Sequence Read Archive (SRA) (PRJNA525544).

The adaptor sequences of raw reads were removed using cutadapt (22), then the clean reads were aligned to the human genome (hg38) using STAR (23). Prior to the next analysis,

the R package, DESeq (24), was used to remove bad counts and filter the differential expression according to the conditions of an absolute $\log_2$FC value of >1 and an adjusted p-value of $<10^5$.

Functional Analysis

The DEGs were used to perform GO and KEGG analyses using the clusterProfiler package (25), and an adjusted p-value of <0.01 was considered as a significant event. Moreover, in order to deeply analyze functional variations between the bladder cancer tissue and their adjacent tissue, GSEA was utilized to discover the

FIGURE 6 | Western blotting detection of MAPK signaling pathway. Lysates from three pairs of bladder cancer and adjacent tissues were subjected to western blotting with antibody to HePTP, FGFR1, c-Fos, PKCα, p21 ras, and Erk2. GAPDH is a reference gene.

molecular mechanism of the whole genome at transcription level, rather than the DEGs.

Network Analysis

All DEGs were used for the co-expression analysis using the WGCNA package (26) and were superimposed onto the PPI database of STRING (27). The co-expression analysis clusters were delineated using the dynamic tree cut package, with the minimum height for each module set to 0.2 (28). The overall trend of each module was based upon the eigengene, and the members of each module were collected through Pearson correlation from among DEGs and their interactors. Moreover, a topological overlapping matrix was also utilized to filter the PPI network (29). In the end, individual modules were annotated using clusterProfiler (25) and were visualized in Cytoscape (30).

Survival Analysis

The survival analysis was used to reveal the clinical outcomes of the hub genes in cancer prognosis. The expression profiles and clinical data of 431 bladder cancer patients were obtained from TCGA database using TCGAbiolinks (31). The 431 samples were split into a high expression group and a low expression group, according to the hub genes, using the survminer package for the best separation. A p-value of <0.05 was considered statistically significant and is shown in the results.

Quantitative Real-Time PCR (qRT-PCR)

qRT-PCR was used to verify the results of RNA-seq. The total RNA of 24 paired bladder tumors and their adjacent tissues were extracted using TRIzol. The genes of interest were then quantified through qRT-PCR using a One-Step qPCR Kit (Invitrogen, USA) and executed with a CFX ConnectTM Real-Time System

(BIO-RAD, USA), according to the manufacturer's instructions. The results were analyzed through the $2^{-\Delta\Delta CT}$ method (32), with GAPDH as a reference gene.

Immunohistochemistry (IHC)

The specimens from the bladder patients who underwent surgical resection were cut to 4 μm thick sections, were then formalin-fixed and paraffin-embedded for IHC, as described previously (33). The primary antibodies used are as follows: CD48 (No.133506, Abcam), MAP4K1 (No.33910, Abcam), IL16 (No.184161, Abcam), CYTIP (No.154847, Abcam), and PTPRC (No.40763, Abcam). Image Pro Plus 6.0 (Media Cybernetics, Bethesda, MD, United States) was employed to measure the positive area of hub genes for quantitative analysis.

Western Blotting

The tissue samples were stored at −80°C for 16 h and lysed with Tissue Extraction Reagent I (Invitrogen, USA) supplemented with protease and phosphatase inhibitors. The BCA assay kit (Thermo Scientific, USA) was then used to measure protein concentration. In brief, the lysate proteins were separated using SDS-PAGE, followed by being transferred into PVDF membranes (Invitrogen, USA), then subjected to the general process of western blotting, according to the instructions of the manufacturers of the antibodies, purchased from CST and Santa Cruz: PKCα (#59754, CST), FGFR1 (#9740, CST), c-Fos (#2250, CST), HePTP (sc-271245, Santa Cruz), p21 Ras (#3965, CST), Erk2 (#9108, CST) and GAPDH (#5174, CST).

Statistical Analysis

All experiments were performed in triplicate, at least. For the analysis between two groups, the student's t-test was leveraged for comparison between tumor tissue and its adjacent tissue. Data are presented as mean ± SDs, except when indicated otherwise. A p-value of <0.05 is considered to be statistically significant.

DATA AVAILABILITY

The datasets generated for this study can be found in Sequence Read Archive (SRA), PRJNA525544.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of CIOMS. The protocol was approved by the institutional review boards of the China-Japan Union Hospital of Jilin University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

XJ and JT conceived and designed the study. XC, FJ, CJ, and ML collected analyzed the data, XC and FJ wrote the manuscript. YN, LQ, QK, FH, WL, and WN collected the samples and revised the manuscript. All authors read and approved the manuscript.

REFERENCES

1. Sanli O, Dobruch J, Knowles MA, Burger M, Alemozaffar M, Nielsen ME, et al. Bladder cancer. *Nat Rev Dis Primers.* (2017) 3:17022. doi: 10.1038/nrdp.2017.22

2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* (2015) 136:E359–86. doi: 10.1002/ijc.29210

3. Elias K, Svatek RS, Gupta S, Ho R, Lotan Y. High-risk patients with hematuria are not evaluated according to guideline recommendations. *Cancer.* (2010) 116:2954–9. doi: 10.1002/cncr.25048

4. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer: a comprehensive review of the published literature. *Pharmacoeconomics.* (2003) 21:1315–30. doi: 10.1007/bf03262330

5. Kaufman DS, Shipley WU, Feldman AS. Bladder cancer. *Lancet.* (2009) 374:239–49. doi: 10.1016/s0140-6736(09)60491-8

6. Song BN, Kim SK, Chu IS. Bioinformatic identification of prognostic signature defined by copy number alteration and expression of CCNE1 in non-muscle invasive bladder cancer. *Exp Mol Med.* (2017) 49:e282. doi: 10.1038/emm.2016.120

7. Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet.* (2013) 45:1459–63. doi: 10.1038/ng.2798

8. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell.* (2014) 25:152–65. doi: 10.1016/j.ccr.2014.01.009

9. Zuiverloon TC, Beukers W, van der Keur KA, Nieuweboer AJ, Reinert T, Dyrskjot L, et al. Combinations of urinary biomarkers for surveillance of patients with incident nonmuscle invasive bladder cancer: the European FP7 UROMOL project. *J Urol.* (2013) 189:1945–51. doi: 10.1016/j.juro.2012.11.115

10. Rodriguez-Vida A, Saggese M, Hughes S, Rudman S, Chowdhury S, Smith NR, et al. Complexity of FGFR signalling in metastatic urothelial cancer. *J Hematol Oncol.* (2015) 8:119. doi: 10.1186/s13045-015-0221-6

11. Goyal L, Saha SK, Liu LY, Siravegna G, Leshchiner I, Ahronian LG, et al. Polyclonal secondary FGFR2 mutations drive acquired resistance to FGFR inhibition in patients with FGFR2 fusion-positive cholangiocarcinoma. *Cancer Discov.* (2017) 7:252–63. doi: 10.1158/2159-8290.cd-16-1000

12. Nogova L, Sequist LV, Perez Garcia JM, Andre F, Delord JP, Hidalgo M, et al. Evaluation of BGJ398, a fibroblast growth factor receptor 1-3 kinase inhibitor, in patients with advanced solid tumors harboring genetic alterations in fibroblast growth factor receptors: results of a global phase, I., dose-escalation and dose-expansion study. *J Clin Oncol.* (2017) 35:157–65. doi: 10.1200/jco.2016.67.2048

13. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics.* (2010) 26:976–8. doi: 10.1093/bioinformatics/btq064

14. Meeks JJ, Lerner SP. Molecular landscape of non-muscle invasive bladder cancer. *Cancer Cell.* (2017) 32:550–1. doi: 10.1016/j.ccell.2017.08.015

15. Tan H, Yang K, Li Y, Shaw TI, Wang Y, Blanco DB, et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity.* (2017) 46:488–503. doi: 10.1016/j.immuni.2017.02.010

16. Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* (2002) 12:37–46. doi: 10.1101/gr.205602

17. Chang L, Karin M. Mammalian MAP kinase signalling cascades. *Nature.* (2001) 410:37–40. doi: 10.1038/35065000

18. Kyriakis JM, Avruch J. Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. *Physiol Rev.* (2012) 92:689–737. doi: 10.1152/physrev.00028.2011

19. Grossi V, Peserico A, Tezil T, Simone C. p38alpha MAPK pathway: a key factor in colorectal cancer therapy and chemoresistance. *World J Gastroenterol.* (2014) 20:9744–58. doi: 10.3748/wjg.v20.i29.9744

20. Courjal F, Cuny M, Simony-Lafontaine J, Louason G, Speiser P, Zeillinger R, et al. Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.* (1997) 57:4360–7.

21. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med.* (2010) 2:62ra93. doi: 10.1126/scitranslmed.3001451

22. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* (2011) 17:10–2. doi: 10.14806/ej.17.1.200

23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635

24. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* (2010) 11:R106. doi: 10.1186/gb-2010-11-10-r106

25. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* (2008) 9:559. doi: 10.1186/1471-2105-9-559

27. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015) 43:D447–52. doi: 10.1093/nar/gku1003

28. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics.* (2008) 24:719–20. doi: 10.1093/bioinformatics/btm563

29. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science.* (2002) 297:1551–5. doi: 10.1126/science.1073374

30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

31. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2016) 44:e71. doi: 10.1093/nar/gkv1507

32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* (2001) 25:402–8. doi: 10.1006/meth.2001.1262

33. Chuntao G. SCF, regulated by HIF-1α, promotes pancreatic ductal adenocarcinoma cell progression. *PLoS ONE.* 3:e0121338. doi: 10.1371/journal.pone.0121338