A pan-cancer analysis of the role of hexokinase II (HK2) in human tumors

Ruiqi Li（✉ 411439820180@email.ncu.edu.cn ）

The Second Affiliated Hospital of Nanchang University   https://orcid.org/0000-0002-4140-9553

Shuchong Mei

The Second Affiliated Hospital of Nanchang University

Qiang Ding

The Second Affiliated Hospital of Nanchang University

Qingming Wang

The Second Affiliated Hospital of Nanchang University

Li Yu

The Second Affiliated Hospital of Nanchang University

Fuming Zi

The Second Affiliated Hospital of Nanchang University

Abstract

Background:

In view of the complexity of tumorigenesis, it is very important to conduct a pan-cancer expression analysis and assess its correlation with clinical prognosis and potential molecular mechanisms of gene interested. More and more evidences show that HK2 is closely related to tumors. But no pan-cancer analysis is available. The aim of this paper was to explore the potential roles of HK2 across thirty-three tumors based on the datasets of TCGA (The Cancer Genome Atlas) and GEO (Gene Expression Omnibus).

Results:

HK2 is highly expressed in most tumors and related to the progression of some tumors. *HK2* expression were associated with the infiltration of T follicular helper cells (Tfh) for the TCGA tumors of UVM, BRCA, BRCA-LumA, HNSC, HNSC-HPV+, and cancer-associated fibroblasts for the tumors of LGG and STAD.

Conclusion:

Our first pan-cancer analyses of HK2 indicated statistical correlations of HK2 expression with clinical prognosis and immune cell infiltration across multiple tumors, which help to understand the role of HK2 in tumorigenesis from the perspective of clinical tumor samples.

1. Intruduction

In view of the complexity of tumorigenesis, it is very important to conduct a pan-cancer expression analysis and assess its correlation with clinical prognosis and potential molecular mechanisms of gene interested. The publicly funded TCGA project and the available GEO database contain functional genomics data sets of different tumors and thus allow us to conduct pan-cancer analysis[1, 2].

One hallmarks of tumors is metabolic reprogramming, i.e., the "Warburg effect" is their propensity to metabolize glucose to lactic acid even in the presence of oxygen[3]. The pivotal player in this frequent cancer phenotype is mitochondrial-bound hexokinase. Hexokinase 2 (HK2) is the major bound hexokinase isoform expressed in cancers that exhibit a "Warburg effect"[4]. It plays a central role in the cellular uptake and utilization of glucose and is highly expressed in a variety of cancer cells, but only in a few normal adult tissues (skeletal muscle, heart, adipose tissue)[4–6]. HK2 helps immortalize cancer cells, escapes chemotherapy inhibition. With the reemergence and acceptance of the "Warburg effect" as a prominent phenotype of clinical cancers, metabolic targeting as a therapeutic strategy is rational[3, 4, 6].

In this study, we first used the TCGA and GEO databases to conduct a pan-cancer analysis of HK2. We also included a group of factors, such as gene expression,

survival status, genetic alteration, immune infiltration, and relevant cellular pathway, to investigate the potential molecular mechanism of HK2 in the pathogenesis or clinical prognosis of different tumors.

2. Materials And Methods

2.1 Gene expression and analysis

We enter HK2 in the "Gene_DE" module of TIMER2 (tumor immune estimation resource, version 2) website (http://timer.cistrome.org/) and observed the expression of HK2 between tumor and normal tissues of the different tumors of the TCGA project. For tumors without normal control tissues, we used the "Expression analysis-Box Plot" module of the GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2) database (http://gepia2.cancer-pku.cn/#analysis) as a supplement[7] to obtain the expression of tumor tissues and corresponding normal tissues of the GTEx (Genotype-Tissue Expression) database, under the settings of P-value cutoff = 0.01, $\log_2$FC (fold change) cutoff = 1, and "Match TCGA normal and GTEx data". Furthermore, we got violin plots of the HK2 expression in different pathological stages (stage I, II, III, and IV) of all tumors via the "Expression analysis-Stage Plot" module of GEPIA2.

We used UALCAN portal (http://ualcan.path.uab.edu/analysis-prot.html) to conduct protein expression analysis with the CPTAC (Clinical proteomic tumor analysis consortium) module[8].

2.2 Survival prognosis analysis

The "Survival Analysis-Survival Map" module of GEPIA2 was used to obtain the overall survival (OS) and disease-free survival (DFS) data of HK2 in all TCGA tumors[7]. We used cutoff-high (50%) and cutoff-low (50%) values as the expression thresholds for defining the high-expression and low-expression cohorts. The log-rank test was used in the hypothesis test, and the survival plots were also obtained through the "Survival Analysis" module of GEPIA2.

2.3 Genetic alteration analysis

For the genetic alteration characteristics of HK2, we used the cBioPortal web (https://www.cbioportal.org/)[9] and chose the "TCGA PanCancer Atlas Studies-Query By Gene" in the "Quick select" section and entered "HK2". The alteration frequency, mutation type and CNA (Copy number alteration) of HK2 of all TCGA tumors were observed in the "Cancer Types Summary" module. The mutated site of HK2 were displayed in the schematic diagram of the protein structure or the 3D (Three-dimensional) structure via the "Mutations" module. The overall survival, disease-free survival, progression-free survival, and disease-free survival were obtained from the "Comparison" module of the TCGA cancer cases with or without HK2 genetic alteration. Kaplan-Meier plots with log-rank P-value were generated as well.

2.4 Immune infiltration analysis

The "Immune-Gene" module of the TIMER2 web was used to explore the association between HK2 expression and immune infiltration of all TCGA tumors. The immune cells of cancer-associated

fibroblasts (CAFs) and T follicular helper cells (Tfh) were selected. The EPIC, MCPCOUNTER, XCELL, TIDE, CIBERSORT, and CIBERSORT-ABS algorithms were applied for immune infiltration estimations. The P-values and partial correlation (cor) values were acquired via the purity-adjusted Spearman's rank correlation test. The data were demonstrated as a heatmap and a scatter plot.

2.5 HK-2-related gene enrichment analysis

Protein name ("HK2") and organism ("Homo sapiens") were entered in the STRING website (https://string-db.org/) for protein-protein interaction. Subsequently, we set the following main parameters: meaning of network edges ("evidence"), active interaction sources ("experiments"), minimum required interaction score ["Low confidence (0.150)"], and max number of interactors to show ("no more than 50 interactors" in 1st shell). Then, the determined HK2-binding proteins were available.

In order to obtain the top 100 HK2-correlated targeting genes based on the datasets of all TCGA tumors and normal tissues, the "Expression Analysis-Similar Gene Detection" module of GEPIA2 was used. We also used the "Expression Analysis-Correlation Analysis" module of GEPIA2 to perform a coupled gene Pearson correlation analysis of HK2 and selected genes. The log2 TPM was applied for the dot plot. Moreover, we used the "Exploration-Gene_Corr" module of TIMER2 to obtain the heatmap data of the selected genes.

We used Jvenn website (http://bioinformatics.psb.ugent.be/webtools/Venn/) to conduct an intersection analysis to compare the top 100 HK2-binding genes and interacted above ≤ 50 genes[10]. Then, the two parts of data were combined to perform KEGG (Kyoto encyclopedia of genes and genomes) pathway analysis. The above total gene list was upload to DAVID (Database for annotation, visualization, and integrated discovery) website (https://david.ncifcrf.gov/) with the settings of selected identifier ("OFFICIAL_GENE_SYMBOL") and species ("Homo sapiens"). Moreover, we analyzed above gene list with "Functional Annotation Chart" module of DAVID and chose only KEGG pathway. The file was download and data of P-value ≤ 0.05 were selected. The enriched pathways were finally visualized with the Bioinformatics website (http://www.bioinformatics.com.cn).

3. Results

3.1 Gene expression analysis data

In this study, we aimed to explore the role of human HK2 (HM_000189.5 for mRNA or NP_000180.2 for protein) in tumors. We first analyzed the expression of HK2 in different tumors and nontumor tissues. The TIMER2 database was used to analyze the expression of *HK2* in various cancers of TCGA. As shown in Fig. 1a, the expression of *HK2* in the tumor tissues of CHOL (Cholangiocarcinoma), HNSC (Head and neck squamous cell carcinoma), KIRC (Kidney renal clear cell carcinoma), KIRP (Kidney renal papillary cell carcinoma), LUSC (Lung squamous cell carcinoma), PRAD (Prostate adenocarcinoma), STAD (Stomach adenocarcinoma), THCA (Thyroid carcinoma), UCEC (Uterine corpus endometrial carcinoma) (*p < 0.001*), CESC (Cervical squamous cell carcinoma and endocervical adenocarcinoma), ESCA (Esophageal

carcinoma), GBM (Glioblastoma multiforme) ($p < 0.01$), and BLCA (Bladder urothelial carcinoma) ($p < 0.05$) is higher than the corresponding normal tissues.

We further evaluated the expression of HK2 between the normal tissues and tumor tissues of ACC (Adrenocortical carcinoma), LGG (Brain lower grade glioma), THYM (Thymoma), and UCS (Uterine carcinosarcoma) (Fig. 1b, $p < 0.01$) with GTEx dataset as a complement of TIMER2 database. However, there were no significant difference for other tumors, such as BRCA (Breast invasive carcinoma), LIHC (Liver hepatocellular carcinoma), PAAD (Pancreatic adenocarcinoma), PCPG (Pheochromocytoma and paraganglioma) as shown in Fig. 1a.

The results of the CPTAC datasets showed higher expression of HK2 total protein in the primary tissue of breast cancer, KIRC, colon cancer, LUAD (Lung adenocarcinoma), ovarian cancer and UCEC (Fig. 1c, $p < 0.001$) than in normal tissues.

The "Pathological Stage Plot" module of GEPIA2 was used to analyze the correlation between *HK2* expression and the pathological stages of cancers, including CESC, KICH (Kidney Chromophobe), LIHC, ovarian cancer and PAAD (Fig. 1d, all $p < 0.05$) but not others.

3.2 Survival analysis data

According to the expression levels of *HK2*, we divided the cancer cases into high-expression and lower-expression groups and investigated the correlation of *HK2* expression with the prognosis of patients with different tumors, mainly using the datasets of TCGA and GEO, respectively. As shown in Fig. 2a and 2b, highly expressed *HK2* was linked to poor prognosis of overall survival (OS) for cancers of CESC ($p = 0.00069$), KIRP ($p = 0.043$), LGG ($p = 0.0000061$), LIHC ($p = 0.027$), LUAD ($p = 0.016$), SARC (Sarcoma) ($p = 0.025$) and disease-free survival (DFS) for cancers of KICH ($p = 0.048$), LGG ($p = 0.0000058$) within the TCGA project. Additionally, low expression of the *HK2* gene was related to poor OS for ACC ($p = 0.022$) as shown in Fig. 2a.

3.3 Genetic alteration analysis data

The genetic alteration status of HK2 in different tumor samples of the TCGA cohorts was analyzed. As shown in Fig. 3a, the highest alteration frequency of HK2 (> 5%) appears for patients with UCEC and SKCM with "mutation" as the primary type. The "amplification" type and "deep deletion" type of CNA was the primary type in the DLBC, which show an alteration frequency of ~ 2%. Significantly, all PCPG cases with genetic alteration (~ 1% frequency) had "amplification" of HK2 (Fig. 3a). The types, sites and case number of the HK2 genetic alteration are further showed in Fig. 3b. As shown in Fig. 3b, missense mutation of HK2 was the main type of genetic alteration. We can also observe that V431Cfs*26, L795Rfs*10 and A901Rfs*74 alteration were detected in cases of STAD, LUSC and UCEC, respectively, which were able to induce a frame shift mutation of the HK2 gene, translation from V (Valine) to C (Cystine) at the 431 site, L (Leucine) to R (Arginine) at the 795 site, and A (Alanine) to R (Arginine) at the 901 site of HK2 protein. The 3D structure of HK2 protein were demonstrated in Fig. 3b of the V431 site,

L795, and A901 site. Additionally, we explored the association between genetic alteration of HK2 and the survival prognosis of different types of cancers. However, there are no statistical differences between HK2 mutation and survival of various tumors. The data of Fig. 3c indicate that UCEC cases with altered HK2 seemed to have a survival advantage in overall survival ($p = 0.0999$), disease-specific survival ($p = 0.145$), disease-free survival ($p = 0.372$) and progression-free survival ($p = 0.0785$), compared with cases without HK2 alteration.

3.4 Immune infiltration analysis data

As important components of tumor microenvironment, tumor-infiltrating immune cells are closely related to tumor initiation, progression and metastasis[11, 12]. Cancer-associated fibroblasts (CAFs) of the tumor microenvironment were reported to participate in modulating the cancer invasion, tumor growth, and cross talk with other immune cells[12]. Therefore, we used the TIMER, CIBERSORT, CIBERSORT-ABS, XCELL, MCPCOUNTER, TIDE and EPIC algorithms to investigate the potential relationship between the infiltration of different immune cells and HK2 gene expression in diverse cancer types of TCGA. As demonstrated in Fig. 4a and 4b, we observed a statistical positive correlation of HK2 expression and the estimated infiltration of CAFs for the TCGA tumors of LGG and a negative correlation of STAD. T follicular helper cell (Tfh) is another important tumor-infiltrating immune cells of tumor microenvironment[13–15]. As demonstrated in Fig. 4c and 4d, we observed a statistical positive correlation of HK2 expression and the estimated infiltration of Tfh for the TCGA tumors of UVM (Uveal melanoma) and a negative correlation of BRCA, BRCA-LumA, HNSC, HNSC-HPV+.

3.5 Enrichment analysis of HK-2-related partners

In order to investigate the molecular mechanism of the HK2 gene in tumorigenesis, the targeting HK2-binding proteins and the HK2 expression-correlated genes for pathway enrichment analyses were screened. On the basis of the STRING tool, we obtained a total of 50 HK2-binding proteins. The interaction network of these proteins was shown in Fig. 5a. We used the GEPIA2 tool to combine all tumor expression data of TCGA and obtained the top 100 genes that correlated with HK2 expression. As shown in Fig. 5b, the HK2 expression level was positively correlated with that of ACTR3 (Actin related protein 3) (R = 0.45), BZW1 (Basic leucine zipper and W2 domains 1) (R = 0.41), CPSF2 (Cleavage and polyadenylation specific factor 2) (R = 0.44), GSK3B (Glycogen synthase kinase 3 beta) (R = 0.42), GSPT1 (G1 to S phase transition 1) (R = 0.39), KCMF1 (Potassium channel modulatory factor 1) (R = 0.49), MAPK6 (Mitogen-activated protein kinase 6) (R = 0.47), NAA50 (N-alpha-acetyltransferase 50) (R = 0.44), PGK1 (Phosphoglycerate kinase 1) (R = 0.43), and PSMD12 (Proteasome 26S subunit, non-ATPase 12) (R = 0.41) genes (all $p < 0.001$). The corresponding heatmap data was showed in Fig. 5c demonstrating a positive correlation between HK2 and the above ten genes in the majority of detailed cancer types. An intersection analysis of the above two groups showed no common member. (Fig. 5d).

We combined the two datasets to perform KEGG enrichment analyses. The KEGG data of Fig. 5e suggest that "RNA transport" might be involved in the effect of HK2 on tumor pathogenesis.

4. Discussion

It has been reported that HK2 plays an important role in various biological processes of tumor cells[16, 17]. Whether HK2 play a role in the pathogenesis of different tumors through common molecular mechanisms is unclear. A pan-cancer analysis of HK2 from the perspective of overall tumors was failed to retrieve any publication. Therefore, we examined HK2 genes in a total of 33 different tumors based on TCGA, CPTAC and GEO database data, as well as molecular characteristics of gene expression, survival prognosis, genetic alteration, or immune infiltration.

Gene expression of HK2 was highly expressed in most tumors except BRCA, LIHC, PCPG. However, HK2 was lower expressed in tumor tissues than normal tissues of COAD (Colon adenocarcinoma), KICH, READ (Rectum adenocarcinoma) and THYM. Similarly, the survival prognosis analysis data of HK2 gene suggested distinct conclusions for different tumors.

For breast cancer, the expression of HK2 total protein was higher in tumor tissue than normal tissue. However, elevated HK2 expression did not affect the overall survival of breast cancer patients. In MDA-MB-453 xenografts, researchers have demonstrated that HK2 expression is decreased in tumors responding to trastuzumab[18]. However, *in vitro* study showed that knocking down HK2 did not enhance doxorubicin-induced apoptosis[19]. Retrospective study has revealed that patients with high HK2 expression have poor survival in patients with brain metastases of breast cancer[19]. So, the role of HK2 in breast cancer deserves to be further investigated.

For liver cancer, according to the bioinformatics analysis, there were no differences in gene and protein expression of HK2 in tumor tissue than normal tissue as demonstrated in Fig. 1a. However, previous studies have demonstrated that compared with non-dysplastic cirrhosis, the expression level of HK2 is higher in liver cell change/dysplasia in cirrhosis and hepatocellular carcinoma[20, 21]. The expression of HK2 is also associated with overall survival of LIHC as shown in Fig. 2 and other research[22]. Previous study and the bioinformatics analysis have revealed that differences in tumor HK2 expression was associated with cancer stage (p = 0.001)[22].

For lung adenocarcinoma, high HK2 protein expression is associated with poor overall survival as shown in Fig. 1 and Fig. 2. Hypoxia can induce HK2 gene expression which protects human epithelial-like A549 cells against oxidative injury[23, 24]. *In vitro* studies have shown that HK2 inhibition can suppress lung tumor growth[25, 26]. Depletion of HK2 confers sensitization to chemotherapeutic drugs[27]. Recent study has revealed that HK2 identifies a novel circulating tumor cell population associated with poor prognosis in lung cancer patients[28].

Most of studies have revealed that lower expression of HK2 is associated with better overall survival such as CESC, KIRP, LGG, LIHC, LUAD, SARC[22, 28–31]. But for adrenocortical carcinoma, low expression of the *HK2* gene was related to poor OS for ACC (*p* = 0.022) as shown in Fig. 2a. Duan *et al.* reported that cytoplasmic expression of HK1 was significantly higher in ACC than in normal adrenal cortical tissue samples[32]. In cervical carcinoma cells, the HK1 but not HK2 knockdown induced a phenotypic change

characteristic of epithelial-mesenchymal transition, which accelerated tumor growth and metastasis[33]. As suggested in recent studies, HK1 is also involved in the carcinogenesis of some solid tumors[34–36].

In this study, we first presented the association between genetic alteration of HK2 and the survival prognosis of different types of tumors. However, there are no statistical differences between HK2 mutation and survival of various tumors. Furthermore, we integrated the information on HK2-binding components and HK2 expression-related genes across all tumors for a series of enrichment analyses and identified the potential impact of "RNA transport" might be involved in the effect of HK2 on tumor pathogenesis. We applied multiple immune deconvolution methods to discuss a statistical negative correlation between *HK2* expression and the immune infiltration level of T follicular helper cell in the tumors of BRCA, BRCA-LumA, HNSC, HNSC-HPV[+]. In addition, our findings first suggested the association of *HK2* expression had a statistical positive correlation with infiltration of CAFs in LGG and a negative correlation in STAD.

5. Conclusion

Taken together, our first pan-cancer analyses of HK2 indicated statistical correlations of HK2 expression with clinical prognosis and immune cell infiltration across multiple tumors, which help to understand the role of HK2 in tumorigenesis from the perspective of clinical tumor samples.

Abbreviations

TCGA: The cancer genome atlas; GEO: Gene expression omnibus; KEGG: Kyoto encyclopedia of genes and genomes; DAVID: Database for annotation, visualization, and integrated discovery; CHOL: Cholangiocarcinoma; HNSC: Head and neck squamous cell carcinoma; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; LUSC: Lung squamous cell carcinoma; PRAD: Prostate adenocarcinoma; STAD: Stomach adenocarcinoma; THCA: Thyroid carcinoma; UCEC: Uterine corpus endometrial carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; BLCA: Bladder urothelial carcinoma; ACC: Adrenocortical carcinoma; LGG: Brain lower grade glioma; BRCA: Breast invasive carcinoma.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing Interest

The authors declare that they have no competing interests.

Authors' contributions

Conceptualization: F.M.Z; Methodology: Q.M.W, L.Y & S.C.M; Writing-Original Draft: R.Q.L & Q.D; Final Writing & Editing: R.Q.L,Q.D & F.M.Z. All authors have read and approved the fnal manuscript.

Availability of data and materials

Not applicable.

Figures

Figure 1

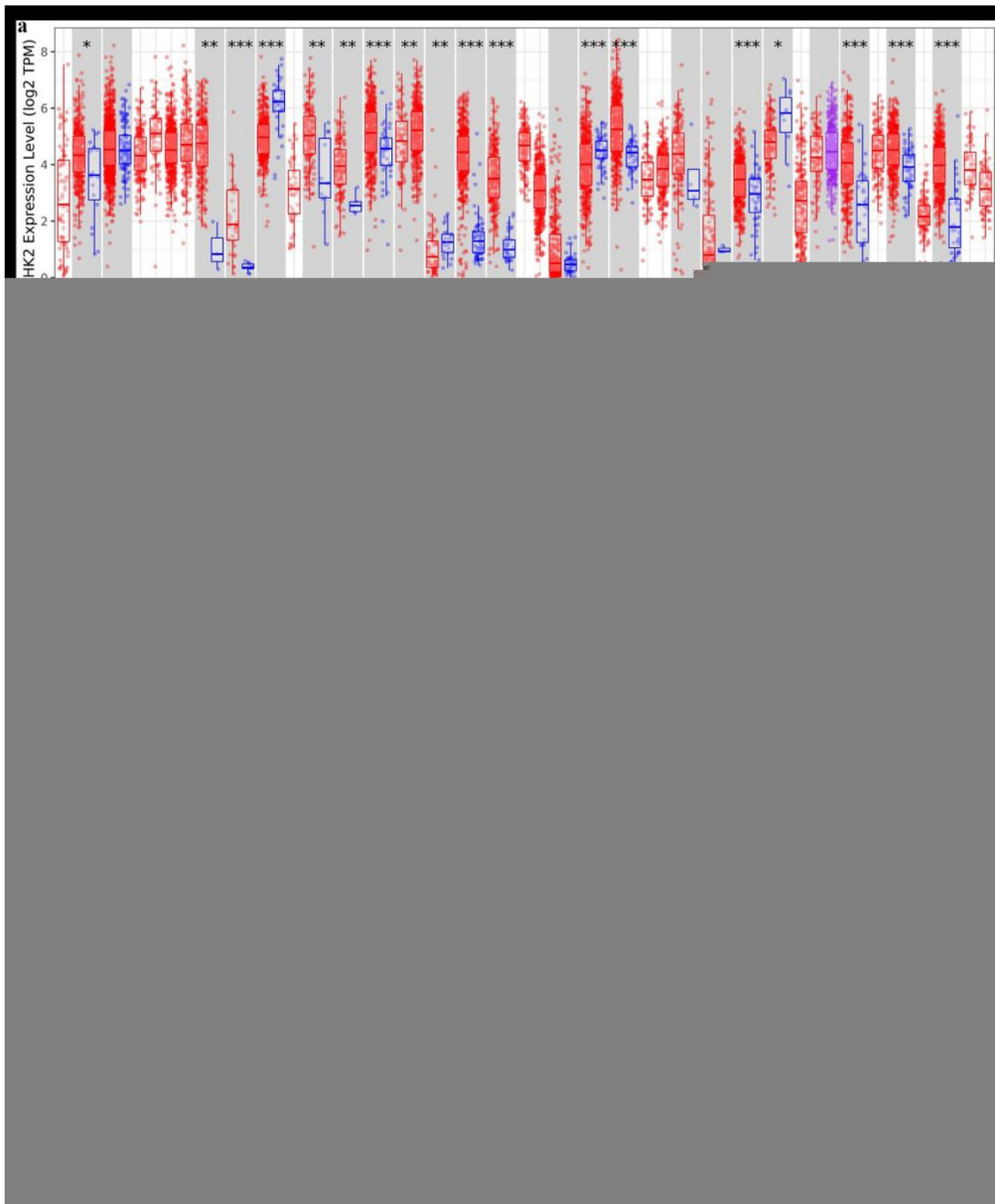Expression level of *HK2* gene in different tumors and pathological stages. (a) The expression status of the *HK2* gene in different cancers or specific cancer subtypes was analyzed through TIMER2. * *P*<0.05; ** *P*<0.01; *** *P*<0.001. (b) For the type of ACC, LGG, THYM and UCS in the TCGA project, the corresponding normal tissues of the GTEx database were included as controls. The box plot data were supplied. ** *P*<0.01. (c) The expression level of HK2 total protein based on the CPTAC dataset between normal tissue

and primary tissue of breast cancer, KIRC, colon cancer, LUAD, ovarian cancer and UCEC. *** *P*<0.001. (d) Based on the TCGA data, the expression levels of the *HK2* gene were analyzed by the main pathological stages (stage I, stage II, stage III, and stage IV) of CESC, KICH, LIHC, ovarian cancer and PAAD. Log2 (TPM+1) was applied for log-scale.
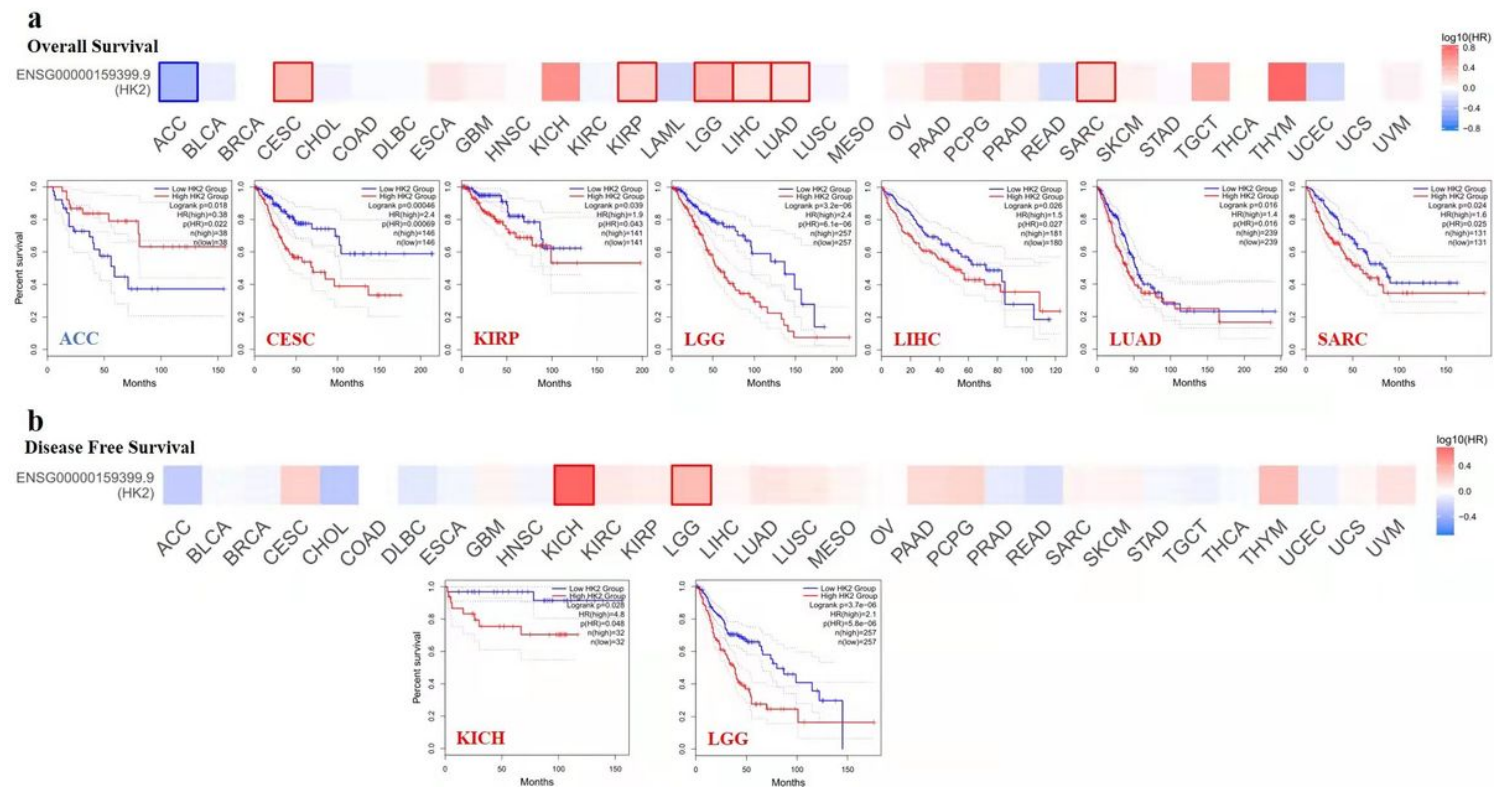

Figure 2

Correlation between *HK2* gene expression and survival prognosis of cancers in TCGA. The GEPIA2 was used to perform overall survival (a) and disease-free survival (b) analyses of different tumors in TCGA by *HK2* gene expression. The survival map and Kaplan-Meier curves with positive results are given.
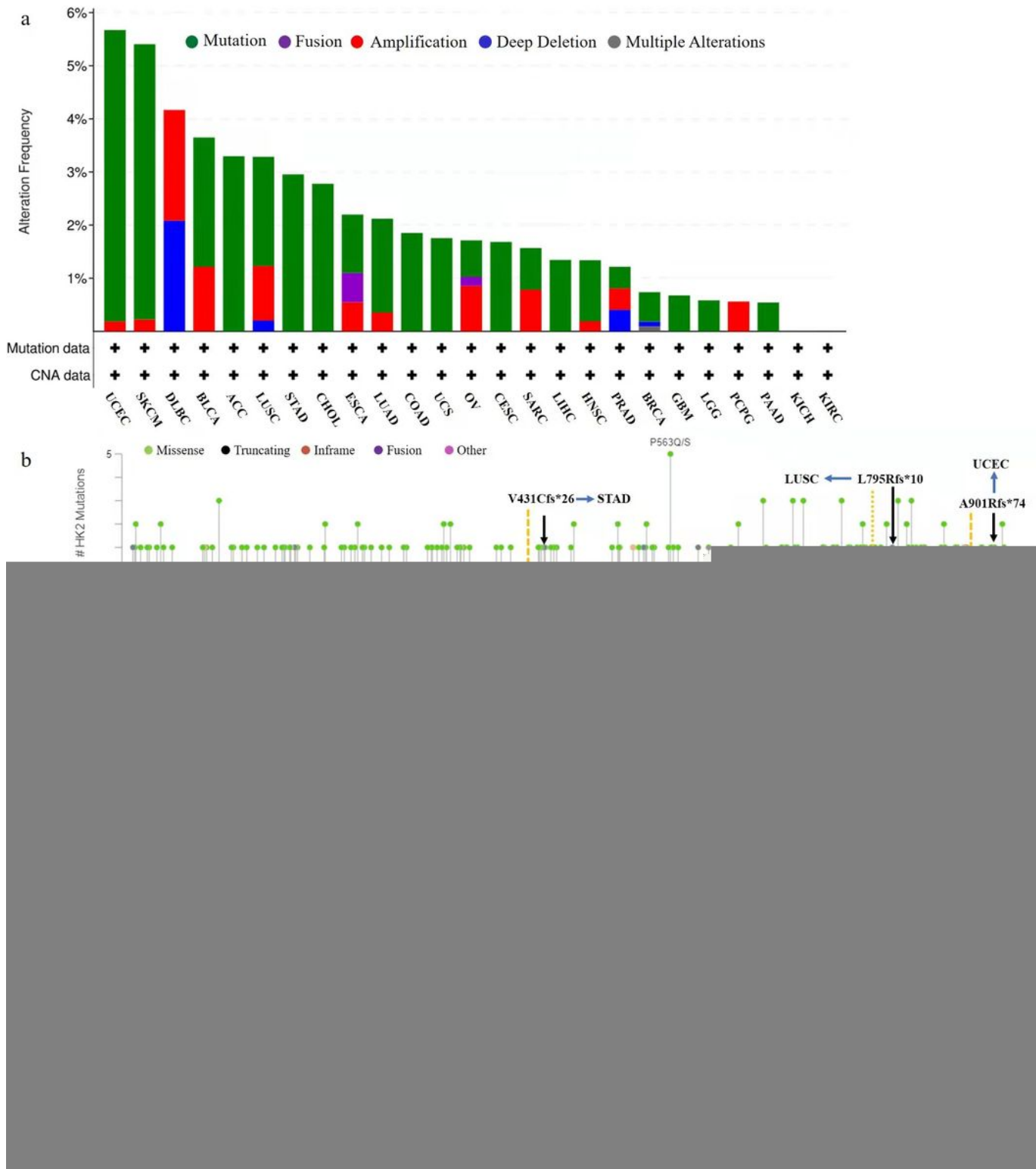
Figure 3

Mutation feature of HK2 in different tumors of TCGA. The mutation features of HK2 for the TCGA tumors were analyzed using the cBioPortal tool. The alteration frequency with mutation type (a) and mutation site (b) are displayed. The mutation site (V431Cfs*26, L795Rfs*10 and A901Rfs*74) in the 3D structure of HK2 were also displayed (b). The potential correlation between mutation status and overall, disease-

specific, disease-free and progression-free survival of UCEC (c) were also analyzed using the cBioPortal tool.

Figure 4

Correlation analysis between *HK2* expression and immune infiltration of cancer-associated fibroblasts and T follicular helper cells. Different algorithms were used to explore the potential correlation between the expression level of the HK2 gene and the infiltration level of cancer-associated fibroblasts and T follicular helper cells across all types of cancer to TCGA.
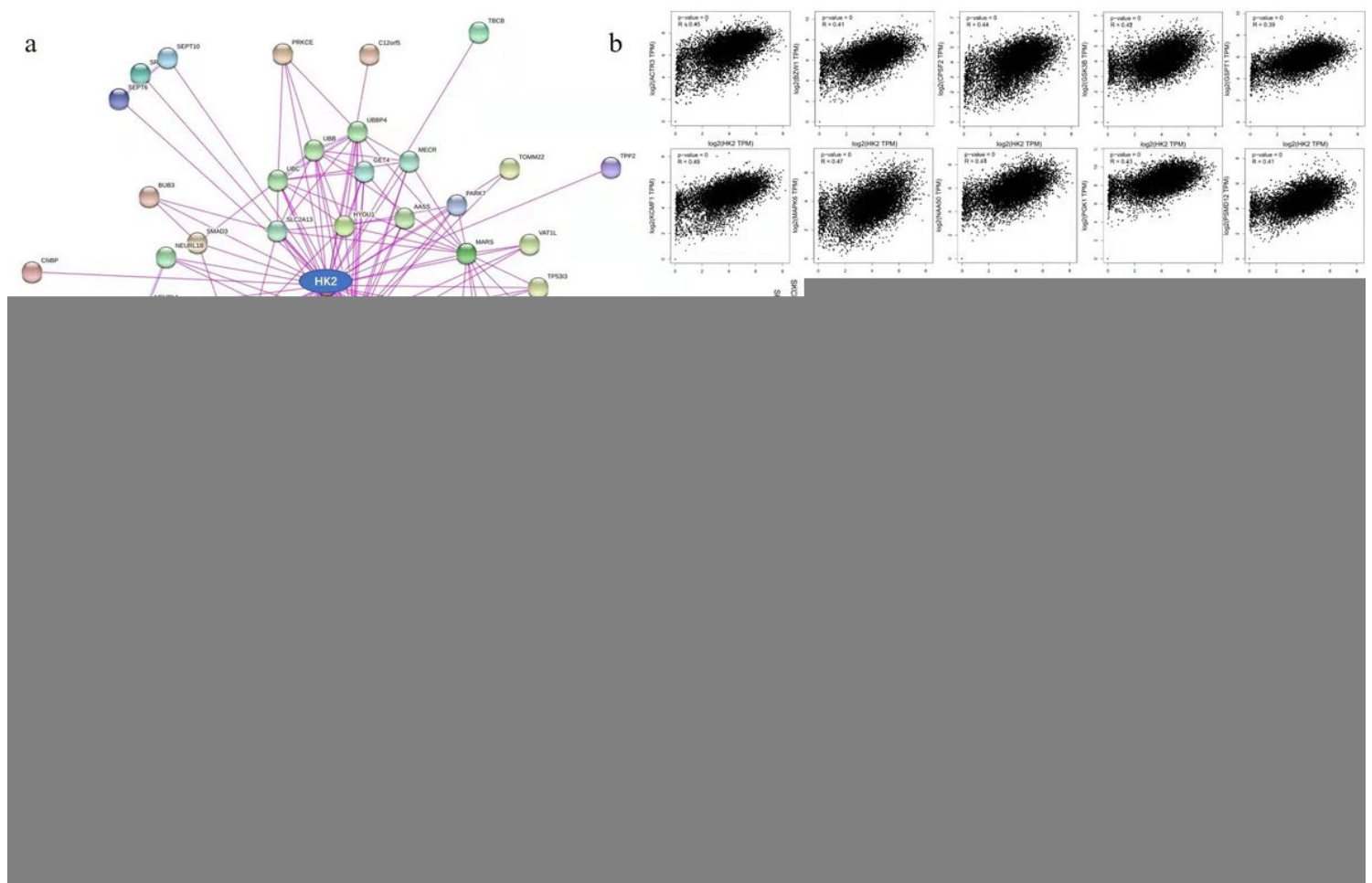

Figure 5

HK2-related gene enrichment analysis. (a) The STRING was used to obtain the available experimentally determined HK2-binding proteins. (b) Using the GEPIA2 approach, we obtained the top 100 HK2-correlated genes in TCGA projects and analyzed the expression correlation between HK2 and selected targeting genes, including ACTR3, BZW1, CPSF2, GSK3B, GSPT1, KCMF1, MAPK6, NAA50, PGK1 and PSMD12. (c) The corresponding heatmap data in the detailed cancer types are displayed. (d) An intersection analysis

of the HK2-binding and correlated genes was conducted. (e) Based on the HK2-binding and interacted genes, KEGG pathway analysis was performed.