

Comprehensive analysis of Downstream of Kinase (DOK) Genes in Pan-cancer

Xiaodong Wang

First Affiliated Hospital of Anhui Medical University

Lifeng Xu

First Affiliated Hospital of Anhui Medical University

Yaxian Li

First Affiliated Hospital of Anhui Medical University

Xin Xu

First Affiliated Hospital of Anhui Medical University

Yongxiang Li (✉ liyongxiang@ahmu.edu.cn)

First Affiliated Hospital of Anhui Medical University <https://orcid.org/0000-0001-7183-9485>

Research

Keywords: DOK1, DOK2, DOK3, DOK4, DOK5, DOK6, DOK7, pan-cancer

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-291409/v1>

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

[Read Full License](#)

Abstract

Background Human cells are often mutated in proto-oncogenes and tumor suppressor genes under the action of tumorigenic factors. When gene mutations accumulate gradually, cells will lose the normal regulation of growth and lead to abnormal proliferation, and then tumors occur. Therefore, understanding the role of these mutated genes in tumors may be the direction of future cancer therapy. Downstream of Kinase (DOK) proteins are a family of polygenic adapters, some of which are key negative regulators of immune cell signal transduction. Its expression level varies significantly in different types of tumors, which is closely related to tumor formation, tumor microenvironment, microsatellite instability, and tumor mutation load.

Methods We mainly use "R" software to process and analyze the data, use "Limma" package and "Wilcox" test to analyze the difference of gene expression, and use Cox proportional Hazards Regression, Kruskal Test, One-class Logistic Regression (OCLR) algorithm and "Corrplot" package and other analysis methods to further process the subsequent data. To get the result that we want to analyze.

Results We found that the expression of DOK family genes varies across multiple tumors and is associated with patient survival. Further analysis showed that DOK gene was significantly associated with tumor immunity, tumor microenvironment, tumor mutation load, etc. DOK2 was highly sensitive to a variety of drugs.

Conclusion DOK gene family includes seven genes, DOK1-DOK7, which are significantly differentially expressed in a variety of tumors, and are significantly correlated with tumor immunity, tumor mutation load, tumor microenvironment, stemness indices, etc., which may provide potential therapeutic targets for future clinical treatment.

Introduction

By 2021, the United States is expected to have 1,898,160 new cancer cases and 608,570 cancer deaths (1), The treatment of cancer remains a major global health problem. Downstream of tyrosine kinase (DOK) gene consists of DOK1, DOK2, DOK3, DOK4, DOK5, DOK6 and DOK7, is a family of genes involved in regulating cell growth, transformation, differentiation, movement, and death (2, 3). DOK proteins play a central role in the assembly of binding partners of different cell types, especially when triggered by receptor tyrosine kinases and immune receptors (3, 4).

They have very similar structural characteristics. At the amino terminus of these genes, they contain both a region that can bind to phosphotyrosine and a domain homologous to Pleckstrin. It is these special structures that provide the necessary conditions for their plasma membrane recruitment. It has been reported in the literature that the expression of DOK4 and DOK5 genes in human T cells (5), and subsequently confirmed that DOK4 is a negative regulator of T cell activation (6).

In addition, mouse models show an important role of these DOK proteins in the immune response (2). They are key negative regulators of immune cell signaling pathways. DOK1/2 is associated with a variety of hematopoietic malignancies, such as chronic myeloid leukemia, chronic lymphocytic leukemia, histiocytic sarcoma, and Burkitt's lymphoma (7). Another study reported that DOK7 expression was associated with survival and tumor recurrence in breast cancer patients (8).

Materials And Methods

1. Introduction to the abbreviation for tumor. The 33 types of TCGA tumors and abbreviations were as follows: adrenocortical carcinoma (ACC), Bladder Urothelial Carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), Cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), Kidney Chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), Acute Myeloid Leukemia (LAML), Brain Lower Grade Glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), Mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), Pheochromocytoma and Paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), Sarcoma (SARC), Skin Cutaneous Melanoma (SKCM), stomach adenocarcinoma (STAD), Testicular Germ Cell Tumors (TGCT), thyroid carcinoma (THCA), Thymoma (THYM), Uterine Corpus Endometrial Carcinoma (UCEC), Uterine Carcinosarcoma (UCS), and Uveal Melanoma (UVM).
2. Data downloading and preprocessing. We downloaded the latest HTSeq-FPKM, Phenotype, survival data, and mutation data for 33 GDC TCGA (The Cancer Genome Atlas) tumors from the UCSC Xena (<http://xena.ucsc.edu/>) website. At the same time, Perl (version 32, v5.32.0), R (version 3.6.3) and other tools were used to extract demographic information, tumor information and follow-up data of all patients from the TCGA database including a total of 11,057 patient data. Data related to the subsequent Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were also obtained from Gene Set Enrichment Analysis (GSEA).
3. Clinical correlation analysis. To analyze the gene expression differences and co-expression relationships of 7 genes in the DOK family in normal samples and tumor samples. The ggpubr "R" package and the "Wilcox" test were used to detect and count differences in the expression of DOK family genes in tumor and normal samples. Tumor data with less than 3 normal samples would not be counted. In Supplement, we used the "corrplot" R package to analyze the co-expression of 7 genes in the DOK family, to explore the potential co-expression association within the genes of the DOKs.
4. Clinical correlation and survival analysis. From the downloaded data, we first extracted the expression information of DOK family genes and analyzed its clinical relevance by combining the survival data of 33 tumors. We divided the gene expression into high expression group and low expression group according to the median and used R package to generate Kaplan-Meier map of DOK gene in pan-cancer. Besides, we adopt Cox proportional hazards regression analysis method, for

each tumor type of DOK1, DOK2, DOK3, DOK4, DOK5, DOK6, DOK7 genetic risk than for evaluation. Next, we specifically analyzed whether there were differences in the expression levels of DOK family in gastric cancer with different pathological stages. $P < 0.05$ was defined as a significant difference.

5. Immune subtype analysis. Tumor cells can promote the growth by changing and maintaining the conditions of their own survival through autocrine and paracrine. In recent years, great breakthroughs have been made in immunotherapy, which has better efficacy than conventional therapy in advanced patients. However, immunotherapy usually relies on the interaction of tumor microenvironment and internal immune regulation. Therefore, the analysis of gene differences in tumor microenvironment and immunity has a good guiding role for subsequent treatment. Thorsson et al, suggested that tumors have immunogenomic characteristics, including wound healing (C1), IFN- γ predominance (C2), inflammation (C3), lymphocytopenia (C4), immune calm (C5), and TGF- β predominance (C6) (9). Kruskal test was used to analyze the mRNA expression levels of DOK1, DOK2, DOK3, DOK4, DOK5, DOK6, DOK7 in 6 different immune subtypes of TCGA tumors.
6. Stemness indices and TME in pan-cancer. Tumor stem cells, a subset of cells that drive tumor growth, we provided a measure, Stemness Indices, to show how many cancer cells will resemble stem cells. Two independent stem cell similarity indices were established based on gene expression (RNAss) and DNA methylation signature (DNAss). The range of these indicators was set as 0 ~ 1, close to 0 indicated low similarity with stem cells, and close to 1 indicated high similarity with stem cells. We extracted DOKs gene expression data and stemness score for Spearman correlation analysis and analyzed TCGA tumor sample data using One-class Logistic Regression (OCLR) algorithm to obtain stemness indices.

Solid tumor tissue not only contains tumor cells, but also contains other cells such as stromal cells, immune cells and vascular cells, which together constitute the tumor microenvironment(TME) To better reflect the purity of tumor cells, we analyzed the proportion of tumor cells, stromal cells and immune cells in solid tumors. We used the “corrplot” package of R software combined with DOKs gene expression data to analyze the relationship between DOKs gene expression and tumor microenvironment, to guide treatment more effectively.

7. Drug sensitivity analysis in pan-carcinoma. CellMiner is a web-based tool compiled by the U.S. National Cancer Institute that provides genomic and pharmacological information for researchers to study transcriptional and drug responses using data from the NCI-60 cell line (10). We downloaded DOK gene expression level and drug sensitivity data from CellMiner and used “Impute” package to process the original data obtained. Pearson correlation analysis was used to analyze the potential relationship between DOK gene expression and compound sensitivity, and our analysis followed the method of Dong et al (11).

Results

1. Differential expression analysis and Co-expression Analysis of DOKs between Tumor and Normal Samples. (Fig. 1A) shows the expression levels of DOK gene family in tumors. It can be observed that the

overall expression levels of DOK1, DOK2, DOK3 and DOK4 are higher than those of DOK5, DOK6 and DOK7.

We used Wilcox assay to analyze the differential expression of seven DOK family genes between tumor samples and paracancerous samples. The DOK gene family is highly expressed in most tumors. However, DOK gene expression was basically low in patients with LUAD, LUSC and KICH, except DOK5 was highly expressed in LUAD. DOK6 is significantly low expressed in GBM, which is different from other DOK genes. In addition, DOK gene family was highly expressed in all gastric cancers (Fig. 1B).

Co-expression analysis showed the expression association between DOK gene families. It can be seen from the figure that DOK3 and DOK2 have the strongest synchronous co-expression (correlation coefficient = 0.66, $p < 0.001$). There are also significant co-expression correlations between DOK6 and DOK5, DOK1 and DOK2, and DOK1 and DOK3, with correlation coefficients of 0.3, 0.37 and 0.45, respectively. On the contrary, DOK2 has an opposite expression relationship with DOK5 and DOK6, the correlation coefficients are -0.13 and -0.18, and DOK6 and DOK7 also have an opposite expression relationship, the correlation coefficient is -0.19, $P < 0.001$ (Fig. 1C).

Almost all DOK family genes were significantly under-expressed in LUSC and KICH tumors. For CHOL tumors, all DOK genes were highly expressed, but only DOK1, DOK3, DOK4, and DOK7 showed statistically significant differences. Unlike DOK5 and DOK7, most DOK genes are significantly overexpressed in KIRC tumors. DOK6 is a special one in the DOK gene family. DOK6 expression is the lowest in the DOK family, and it is also the only gene with significantly low expression in GBM (figure S1-S2).

2. We performed KaplanMeier analysis on DOK1, DOK2, DOK3, DOK4, DOK5, DOK6 and DOK7 gene expression and overall survival time of 33 TCGA tumors (Fig. 2). First, we divided patients into high expression group and low expression group according to the limit of the median value of gene expression and compared whether different gene expression was associated with different survival time. There was a statistically significant difference between high DOK1 expression and poor prognosis in KIRC patients ($P = 0.001$). For LUAD patients, DOK1, DOK2, and DOK6 were all low expressed relative to adjacent or normal tissues, which was also consistent with the survival curve ($P = 0.032$, $P = 0.012$, and $P = 0.049$). The high expression of DOK2 and DOK3 is also a sign of poorer prognosis in GBM patients ($P = 0.014$, $P = 0.035$). DOK3 is highly expressed in KIRC patients, and higher expression means shorter survival time ($P = 0.01$), which is same as DOK5 in STAD patients ($P = 0.044$).

3. We mapped the forest figure to reflect the association between DOK family gene expression and prognosis of 33 TCGA tumor species (Fig. 3). Cox proportional hazard of regression method to detect DOK1, DOK2, DOK3, DOK4, DOK5, DOK6 and DOK7 prognostic role, and defines its hazard ratio (HR) > 1 was a significant prognostic factor. It can be concluded from the Fig. 3 that DOK1, DOK2 and DOK3 have

significance in most cancers. For STAD patients, DOK5, DOK6 and DOK7 are all correlated with their prognosis ($P = 0.04$, $P = 0.01$, $P = 0.01$).

4. We focused on analyzing the significance of DOK gene family in gastric cancer and exploring whether it is associated with different pathological stages. From the Fig. 4, we found that DOK2, DOK3 and DOK5 were significantly correlated with gastric cancer at different pathological stages. Interestingly, we found that the expression levels of stage I and IV genes were the lowest, while the expression levels of stage II and III genes were relatively high. The expression of these differences may be helpful in predicting the clinical development of tumors (Fig. 4).

5. Immune subtype analysis. More than 10,000 tumor samples from 33 types of TCGA were divided into six immune subtypes: C1-C6 was wound healing, IFN- γ -dominated, inflammatory, lymphodepletion, immunosilent, and TGF- β -dominated. We downloaded immune-related data from TCGA and used Kruskal test to analyze the mRNA expression of DOK family genes, to explore the relationship between seven DOK genes and various immune types. As can be seen from the (Fig. 5A), all DOK genes are significantly associated with the C1-C6 immune subtype $P < 0.001$. Further analysis showed that DOK4 had the highest overall expression level in C1-C6, while DOK6 had the lowest expression level. Interestingly, C5 has an abnormally high expression in DOK6, which is consistent with DOK5, whereas C5 has the lowest expression in DOK2 and DOK7. We continue analyzed the association between DOK family genes and immune subtypes in STAD patients. As shown in the (Fig. 5B), DOK2-DOK6 were significantly correlated with the immunity of gastric cancer $P < 0.001$, but DOK7 was not statistically significant with the six immune subtypes. The average expression of DOK4 was the highest among the six immune subtypes. The expression of C6 was roughly the same as the overall immune analysis, and it was in a high position among all DOK genes.

6. The stemness index (DNAss) based on NDA methylation and the stemness index (RNAss) based on mRNA expression were used to measure and analyze the correlation between DOK gene and tumor stem cells. In order to investigate the association between the stemness features of pan-cancer and DOK gene expression, we calculated the stemness indices of TCGA tumor samples by using a one-class logistic regression (OCLR) algorithm and performed Spearman correlation analysis based on gene expression and stemness scores.

We can see that the correlation between DOK family genes and DNAss is generally not high. DOK6 has a large negative correlation with OV and TGCT, with correlation coefficients of -0.68 and -0.78. DOK4 also had a significant negative correlation with TGCT, with a correlation coefficient of -0.73, suggesting that there were fewer tumor stem cells with high expression of these genes (Fig. 6A). Based on the analysis of RNAss, we found that most tumors were negatively correlated with DOK family genes, especially DOK5 and DOK6, which were significantly negatively correlated with most tumors. The special ones are LGG and THYM. DOK4 and DOK6 are positively correlated with LGG, with correlation coefficients of 0.59. In addition, DOK1 and DOK2 are also positively correlated with THYM, with correlation coefficients of 0.5 (Fig. 6B).

The occurrence, growth and metastasis of tumors are closely related to the surrounding environment. In addition to tumor cells, solid tumor tissues also include other normal cells, such as stromal cells and immune cells. Tumor cells could change this environment through autocrine or paracrine, and the body can also limit the occurrence and development of tumors by changing metabolism, secretion, immunity, structure, and other functions. All these constitute the tumor microenvironment (TME). The proportion of stromal cells and immune cells in solid tumors reflects the purity of the tumor, which has guiding significance for subsequent treatment.

We downloaded relevant data from TCGA database, ESTIMATE (using Expression data of Stromal cells and Immune cells in Malignant Tumors using Expression data) was used to calculate the scores of Stromal cells (figure S3) and Immune cells (figure S4) in tumor cells, and Spearman correlation analysis was used to describe the correlation between DOK family gene expression level and tumor purity. As can be seen from the (Fig. 6C), DOK2 and DOK3 are strongly correlated with stromal cells and immune cells of almost all tumors, which means that when DOK2 and DOK3 genes are highly expressed in patients, the purity of tumors will be reduced. The expression of DOK4 and DOK6 was positively correlated with the tumor purity of LGG, and the higher the gene expression, the lower the proportion of stromal cells and immune cells in the tumor, and the correlation coefficients were -0.5 and -0.6 , respectively.

7. We analyzed the correlation between DOK gene and DNASS, RNASS and TME in STAD tumors by using a scatter plot. Except DOK4 and DOK7, the other DOK family genes were negatively correlated with DNASS and RNASS. DOK2 and DOK3 were highly correlated with immune scores, with correlation coefficients of 0.91 and 0.82 , $P < 0.001$, respectively. DOK5 and DOK6 had a higher correlation with matrix score, the correlation coefficients were 0.79 and 0.76 , $P < 0.001$ (Fig. 7).

8. Drug sensitivity analysis in pan-cancer. We downloaded and processed the transcriptional expression of DOK family genes in NCI-60 cancer cell lines and the drug activity of 263 antitumor drugs from CellMiner database to analyze the potential influence of DOK family genes on drug response by Pearson correlation analysis. From the results of the analysis, we can conclude that DOK2 is a gene that is sensitive to a variety of cancer drugs. DOK2 had a significant positive correlation with the sensitivity of nelarabine and chelerythrine, with correlation coefficients of 0.725 and 0.706 , $P < 0.001$; DOK4 was negatively correlated with Okadaic acid, the correlation coefficient was -0.488 , $P < 0.001$. DOK6 was positively correlated with Estramustine, correlation coefficient was 0.547 , $P < 0.001$ (Fig. 8). This analysis of gene expression and drug sensitivity is expected to provide new ideas for clinical treatment and subsequent experimental basic research.

9. Tumor mutation load (TMB). With the rapid development of immunotherapy, the significance of detecting tumor mutation load is becoming more and more important. TMB refers to the number of somatic mutations in the tumor genome after the deletion of germ line mutations, that is, the deletion of innate mutations, only looking at the number of mutations specific to tumor cells. The higher the TMB, the more neoantigens the tumor produces, the more easily the tumor cells can be recognized by the

body's immune cells, and the more effective the efficacy of immunotherapy is likely to be. Through radar chart analysis, DOK4 has the highest TMB correlation with STAD (correlation coefficient 0.28, $P < 0.001$, Fig. 9A). However, DOK6 shows the opposite performance, with a correlation coefficient of -0.42, $P < 0.001$ (Fig. 9B). The correlation analysis of DOK family genes with TMB may provide reference for tumor immunotherapy.

Detection of microsatellite instability (MSI). Microsatellites are short tandem repeats throughout the human genome. Compared with normal cells, microsatellites in tumor cells change in length due to the insertion or deletion of repeat units, leading to the occurrence and development of tumors, which is called microsatellite instability. In the current clinical treatment, microsatellite instability is closely associated with colorectal cancer, and this phenomenon is present in about 15% of colorectal cancer, so we analyzed the correlation between DOK family genes and MSI and listed the radar map most related to COAD. DOK2 was positively correlated with MSI, the correlation coefficient was 0.22, $P < 0.001$, while DOK4 was the most significant gene negatively correlated with MSI, $P < 0.001$, (Fig. 9C-D). Further experiments are needed to prove whether patients with COAD can benefit from the expression of these two genes. Additional radar maps are shown in the attached picture (figure S5-S6).

10. Based on KEGG pathway analysis, we found that DOK family genes were enriched in multiple pathways related to STAD. DOK3, DOK6 and STAD are enriched in the autophagy pathway (Fig. 10).

Discussion

Downstream of kinase (DOK) proteins, which contain negative regulators of immune cell signaling, represent a multigene family of adapters (12). In this study, we investigated in detail the association between DOK gene transcriptional expression and 33 TCGA tumor characteristics. By multidimensional analysis, we first analyzed the differential expression of DOK gene in 10,327 tumor patients and 730 paracancerous patients, and further analyzed the association between DOK differential expression and survival in tumor patients. Then, we analyzed the association between DOK family genes and TME, MSI, TMB, immune subtypes, dryness and drug response from multiple perspectives through various databases, to more comprehensively analyze the function of DOK genes and provide new help in subsequent tumor studies.

Concordant with previous studies, DOK2 and DOK4 are low expressed in breast cancer, and lower gene expression means larger tumor size, later clinical stage, and more lymph node metastasis, which is consistent with the results of our statistical analysis (13, 14). DOK7 down-regulation has also been reported to inhibit the proliferation and invasion of breast cancer through the P13K/PTEN/AKT pathway (15). Similarly, DOK7 hypermethylation may also serve as a biomarker for the diagnosis of breast cancer (16). It is worth noting that DOK2, DOK3 and DOK5 showed statistically significant differences among gastric cancer patients with different stages. Among them, the high expression of DOK5 predicts a shorter survival time, which may be helpful for the gene research of gastric cancer. It has also been reported in

the literature that DOK6, as a connector interacting with a variety of molecules in signal transduction pathway, is an integrated biomarker for a variety of carcinogenic signals in gastric cancer (17). DOK2 is missing in about half of human lung adenocarcinoma, and DOK2 is associated with a co-deletion of DUSP4 that causes lung adenocarcinoma in mice (18). Another study found that DOK2 deletion is associated with mutations in EGFR in human lung adenocarcinoma, jointly promoting tumor development (19). In addition, DOK2 can be used as a poor prognostic marker for human glioblastoma multiforme and can inhibit tumor migration and invasion via the JAK2/STAT3 pathway (20, 21). In our study, in addition to the significant survival correlation between DOK2 and lung adenocarcinoma, DOK1 and DOK6 were also significantly associated with lung adenocarcinoma.

Then, tumor samples were classified according to different immune subtypes of C1-C6 (9), and the expression levels of DOK1-DOK7 RNA-seq were statistically analyzed, and the results showed that these genes were significantly correlated with the immunotyping. As previously reported, DOK3 is an important modulator of innate immune responses in macrophages and B cells and modulates downstream signaling pathways of various immune receptors (22, 23). DOK1 is a tumor suppressor that is often lost in malignant cells, but it still regulates the immunoreceptor activity of stromal cells in the tumor microenvironment and promotes the invasion of cancer cells (24, 25). The proportion of immune cells and stromal cells in solid tumors constitutes the tumor microenvironment TME. By calculating immune scores and stromal scores, we can infer the purity of tumors. Especially DOK2 and DOK3 are related to the immunity of most tumors, which affect tumor treatment methods to varying degrees (26).

It has been reported that tumors acquire stem cell-like properties (27) during their development, with self-renewal and dedifferentiated stem cell-like characteristics (28). In this study, we used OCLR method to calculate DNAss and RNAss scores in tumor samples and correlate them with DOK gene, to explore the role of DOK gene in tumor dryness. DOK gene was found to be negatively correlated with the dryness of most types of tumor cells, especially DOK5 and DOK6.

Studies have found that DOK1 has been identified as possibly related to cisplatin resistance (29), and DOK2 deficiency induces chemotherapy resistance by reducing the level of apoptosis in the treatment response (30). Therefore, we analyzed the correlation between DOK gene transcriptional expression level and various drug responses. As we analyzed, DOK2 was significantly associated with many drugs, among which Nelarabine had the strongest correlation with a correlation coefficient of 0.725, $P < 0.001$. Tumor mutation load and microenvironmental instability have recently become a hot topic of immunotherapy. We used radar map to analyze the microsatellite instability of colon cancer, which may provide reference for subsequent immunotherapy.

This study is a multi-dimensional study of DOK family genes in the pan-cancer, but there are still many obvious limitations. First, the samples in our study are all from public databases in the United States, so the sample size is insufficient. The model we built cannot represent the actual situation in other regions, such as Europe and Asia, and there is no other public external data to verify our model. Second, our research is based on biological information in a database and has not been validated at the molecular or

animal level. Thirdly, our research on DOK gene is based on data correlation, and we have not studied its specific molecular mechanism and action pathway in depth. Therefore, in future studies, we will further explore the specific molecular mechanism of these genes and hope to better explore the specific mechanism of DOK gene in tumor genesis and development combined with this study.

Conclusion

In this study, we conducted a multi-dimensional analysis of DOK1, DOK2, DOK3, DOK4, DOK5, DOK6 and DOK7. These included pan-cancer differential expression analysis, immune subtype analysis, clinical analysis, tumor purity analysis, dry correlation analysis, drug response, tumor mutation load and microenvironmental instability. DOK gene was expressed differently in different tumor types and different immune subtypes. This analysis reveals multiple expression patterns of the DOK family at the pan-cancer level and provides new clues for cancer treatment strategies.

Declarations

Acknowledgement

Thanks to Youliang Wu for guiding the format modification and submission of the magazine.

Statement of Ethics

All analyses were based on Public database; thus, no ethical approval and patient consent are required.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Sources

This work was supported by a grant from the National Natural Science Foundation of China (81874063).

Authors' Contributions

Xiaodong Wang collects all the article data and is responsible for writing the full text. Lifeng Xu, Yaxian Li participated in the writing of the article and the modification of the article format. Xin Xu was responsible for the editing of the pictures and participated in the writing of the full text. Yongxiang Li provided the ideas for the research and all the funding. All authors read and approved the final manuscript.

Availability of data and materials

The data used to support the findings of this study are included within the article.

Consent for publication

All authors agree to publish the paper.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics. 2021. CA: a cancer journal for clinicians. 2021;71(1):7–33.
2. Celis-Gutierrez J, Boyron M, Walzer T, Pandolfi PP, Jonjic S, Olive D, et al. Dok1 and Dok2 proteins regulate natural killer cell development and function. *EMBO J*. 2014;33(17):1928–40.
3. Mashima R, Hishida Y, Tezuka T, Yamanashi Y. The roles of Dok family adapters in immunoreceptor signaling. *Immunol Rev*. 2009;232(1):273–85.
4. Jordan MS, Singer AL, Koretzky GA. Adaptors as central mediators of signal transduction in immune cells. *Nat Immunol*. 2003;4(2):110–6.
5. Favre C, Gerard A, Clauzier E, Pontarotti P, Olive D, Nunes JA. DOK4 and DOK5: new Dok-related genes expressed in human T cells. *Genes Immun*. 2003;4(1):40–5.
6. Gerard A, Ghiotto M, Fos C, Guittard G, Compagno D, Galy A, et al. Dok-4 is a novel negative regulator of T cell activation. *J Immunol*. 2009;182(12):7681–9.
7. Coppin E, Gelsi-Boyer V, Morelli X, Cervera N, Murati A, Pandolfi PP, et al. Mutational analysis of the DOK2 haploinsufficient tumor suppressor gene in chronic myelomonocytic leukemia (CMML). *Leukemia*. 2015;29(2):500–2.
8. Heyn H, Carmona FJ, Gomez A, Ferreira HJ, Bell JT, Sayols S, et al. DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis*. 2013;34(1):102–8.
9. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity*. 2019;51(2):411–2.
10. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res*. 2012;72(14):3499–511.
11. Dong X, Huang D, Yi X, Zhang S, Wang Z, Yan B, et al. Diversity spectrum analysis identifies mutation-specific effects of cancer driver genes. *Commun Biol*. 2020;3(1):6.
12. Guittard G, Pontarotti P, Granjeaud S, Rodrigues M, Abi-Rached L, Nunes JA. Evolutionary and expression analyses reveal a pattern of ancient duplications and functional specializations in the diversification of the Downstream of Kinase (DOK) genes. *Dev Comp Immunol*. 2018;84:193–8.
13. Huang J, Peng X, Zhang K, Li C, Su B, Zhang Y, et al. Co-expression and significance of Dok2 and Ras p21 protein activator 1 in breast cancer. *Oncol Lett*. 2017;14(5):5386–92.
14. Heidarijadi A, Salimi M, Mozdarani H. Study of DOK4 gene expression and promoter methylation in sporadic breast cancer. *Neoplasma*. 2020;67(4):916–21.

15. Yue C, Bai Y, Piao Y, Liu H. DOK7 Inhibits Cell Proliferation, Migration, and Invasion of Breast Cancer via the PI3K/PTEN/AKT Pathway. *J Oncol*. 2021;2021:4035257.
16. Shirkavand A, Boroujeni ZN, Aleyasin SA. Examination of methylation changes of VIM, CXCR4, DOK7, and SPDEF genes in peripheral blood DNA in breast cancer patients. *Indian J Cancer*. 2018;55(4):366–71.
17. Leong SH, Lwin KM, Lee SS, Ng WH, Ng KM, Tan SY, et al. Chromosomal breaks at FRA18C: association with reduced DOK6 expression, altered oncogenic signaling and increased gastric cancer survival. *NPJ Precis Oncol*. 2017;1(1):9.
18. Chen M, Zhang J, Berger AH, DiIombi MS, Ng C, Fung J, et al. Compound haploinsufficiency of Dok2 and Dusp4 promotes lung tumorigenesis. *J Clin Invest*. 2019;129(1):215–22.
19. Berger AH, Chen M, Morotti A, Janas JA, Niki M, Bronson RT, et al. DOK2 inhibits EGFR-mutated lung adenocarcinoma. *PLoS One*. 2013;8(11):e79526.
20. Deshpande RP, Babu PP. pDok2, caspase 3 dependent glioma cell growth arrest by nitidine chloride. *Pharmacol Rep*. 2018;70(1):48–54.
21. Zhang XW, Wang L, Ding H. Long noncoding RNA AK089579 inhibits epithelial-to-mesenchymal transition of peritoneal mesothelial cells by competitively binding to microRNA-296-3p via DOK2 in peritoneal fibrosis. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. 2019;33(4):5112–25.
22. Gao WS, Qu YJ, Huai J, Wei H, Zhang Y, Yue SW. DOK3 is involved in microglial cell activation in neuropathic pain by interacting with GPR84. *Aging*. 2020;13(1):389–410.
23. Loh JT, Teo JKH, Lim HH, Lam KP. Emerging Roles of Downstream of Kinase 3 in Cell Signaling. *Front Immunol*. 2020;11:566192.
24. Li T, Li B, Sara A, Ay C, Leung WY, Zhang Y, et al. Docking protein-1 promotes inflammatory macrophage signaling in gastric cancer. *Oncoimmunology*. 2019;8(11):e1649961.
25. Chaki SP, Barhoumi R, Rivera GM. Nck adapter proteins promote podosome biogenesis facilitating extracellular matrix degradation and cancer invasion. *Cancer Med*. 2019;8(17):7385–98.
26. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501(7467):346–54.
27. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyski A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc Natl Acad Sci U S A*. 2019;116(18):9020–9.
28. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO Rep*. 2014;15(3):244–53.
29. Sarin N, Engel F, Rothweiler F, Cinatl J, Michaelis M, Frotschl R, et al. Key Players of Cisplatin Resistance: Towards a Systems Pharmacology Approach. *Int J Mol Sci*. 2018;19(3).
30. Lum E, Vigliotti M, Banerjee N, Cutter N, Wrzeszczynski KO, Khan S, et al. Loss of DOK2 induces carboplatin resistance in ovarian cancer via suppression of apoptosis. *Gynecol Oncol*.

Figures

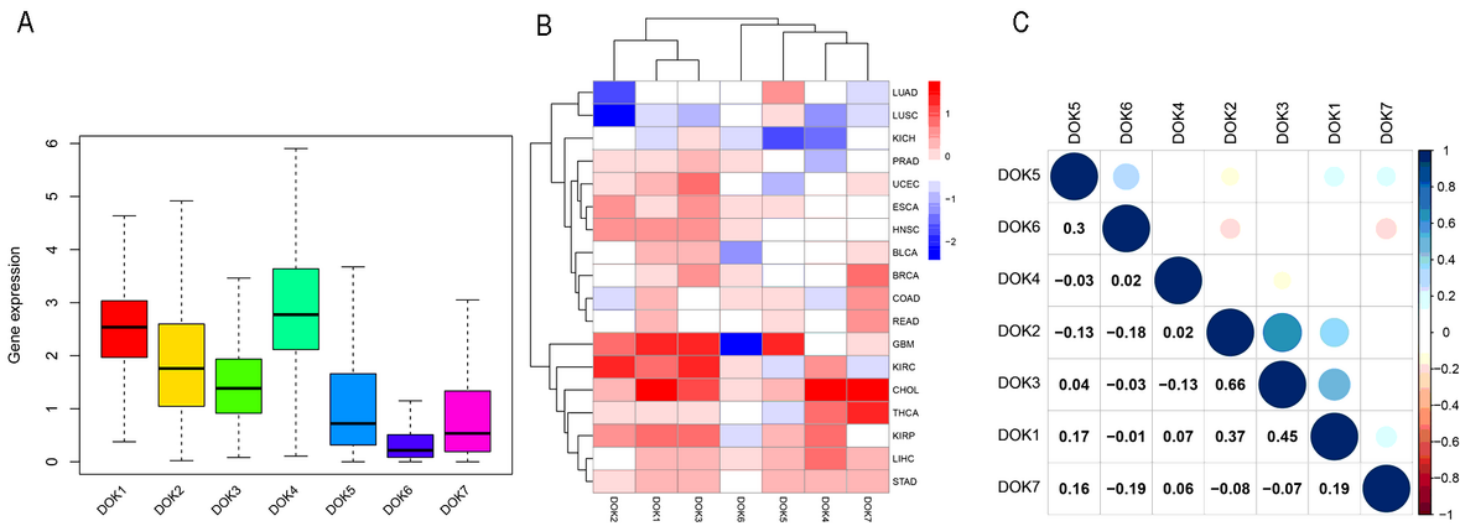


Figure 1

(A) The box plot showing the transcriptional expression levels of PPARs. (B) The heatmap showing the transcriptional level of DOKs in TCGA tumor types compared to normal tissues; the gradient colors represent the log Fold Change (logFC) value. (C) the correlation coefficients by co-expression analysis between every two genes are presented.

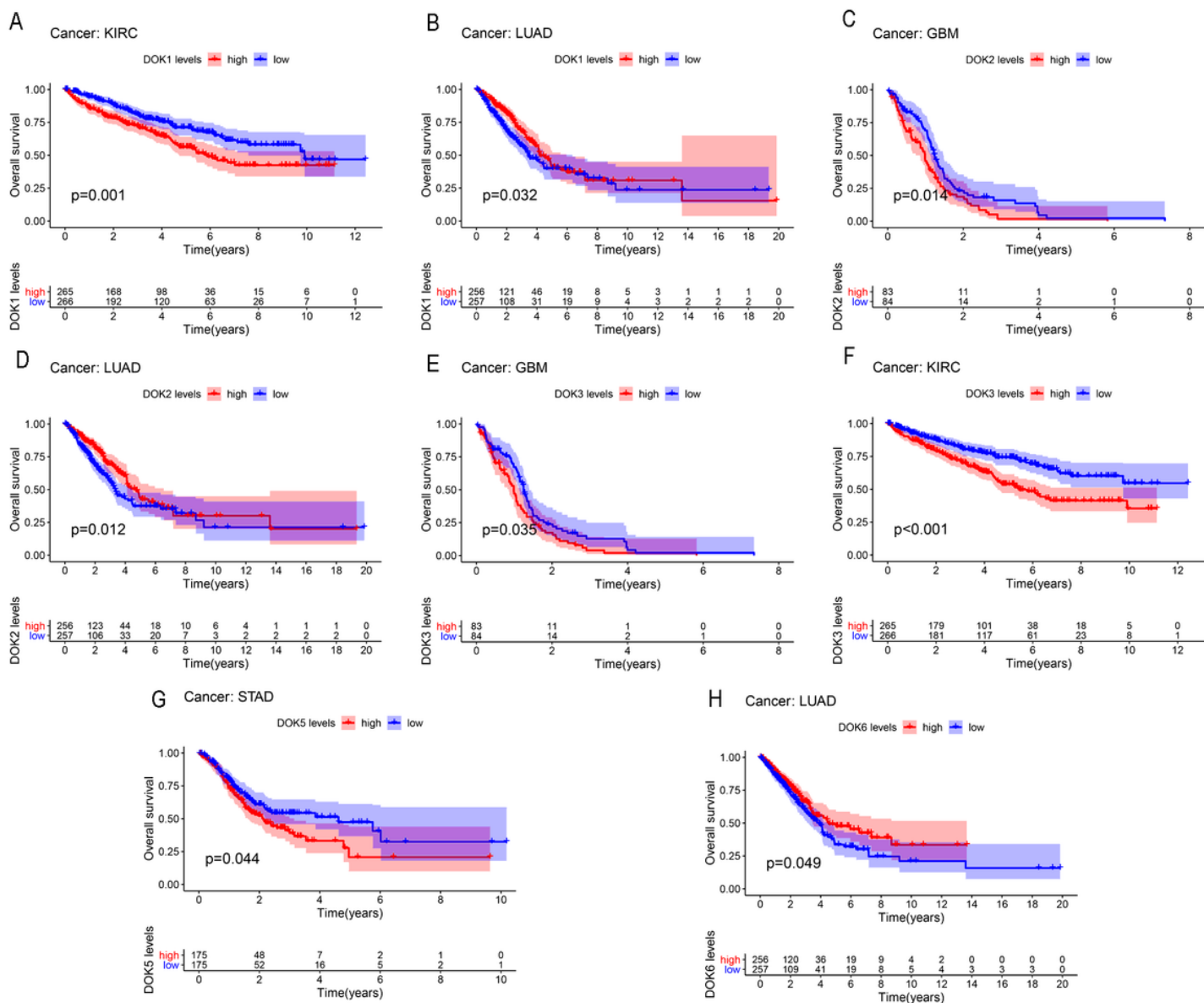


Figure 2

The results of survival analysis of DOKs in pan-cancer. (A–H) Kaplan-Meier plots of DOKs in pan-cancer showing the differential survival outcomes of high DOK and low DOK ($p < 0.05$)

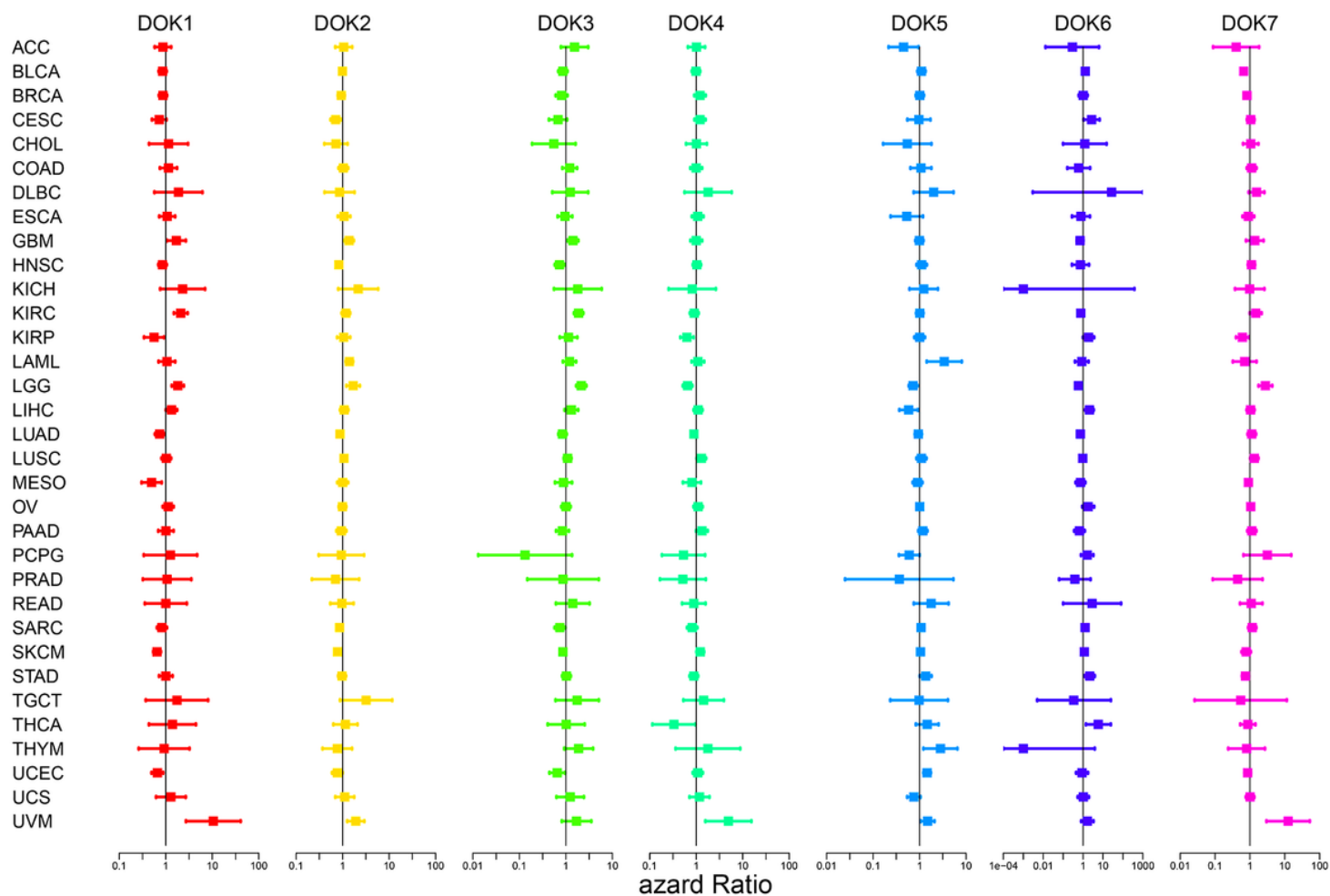


Figure 3

Cox proportional hazard analyses illustrating the hazard ratios (HRs) of DOKs in 33 TCGA tumors; those DOKs whose HR > 1 in certain types of cancer were regarded as danger factors of the very type of cancer, which were unfavorable for prognostic outcomes.

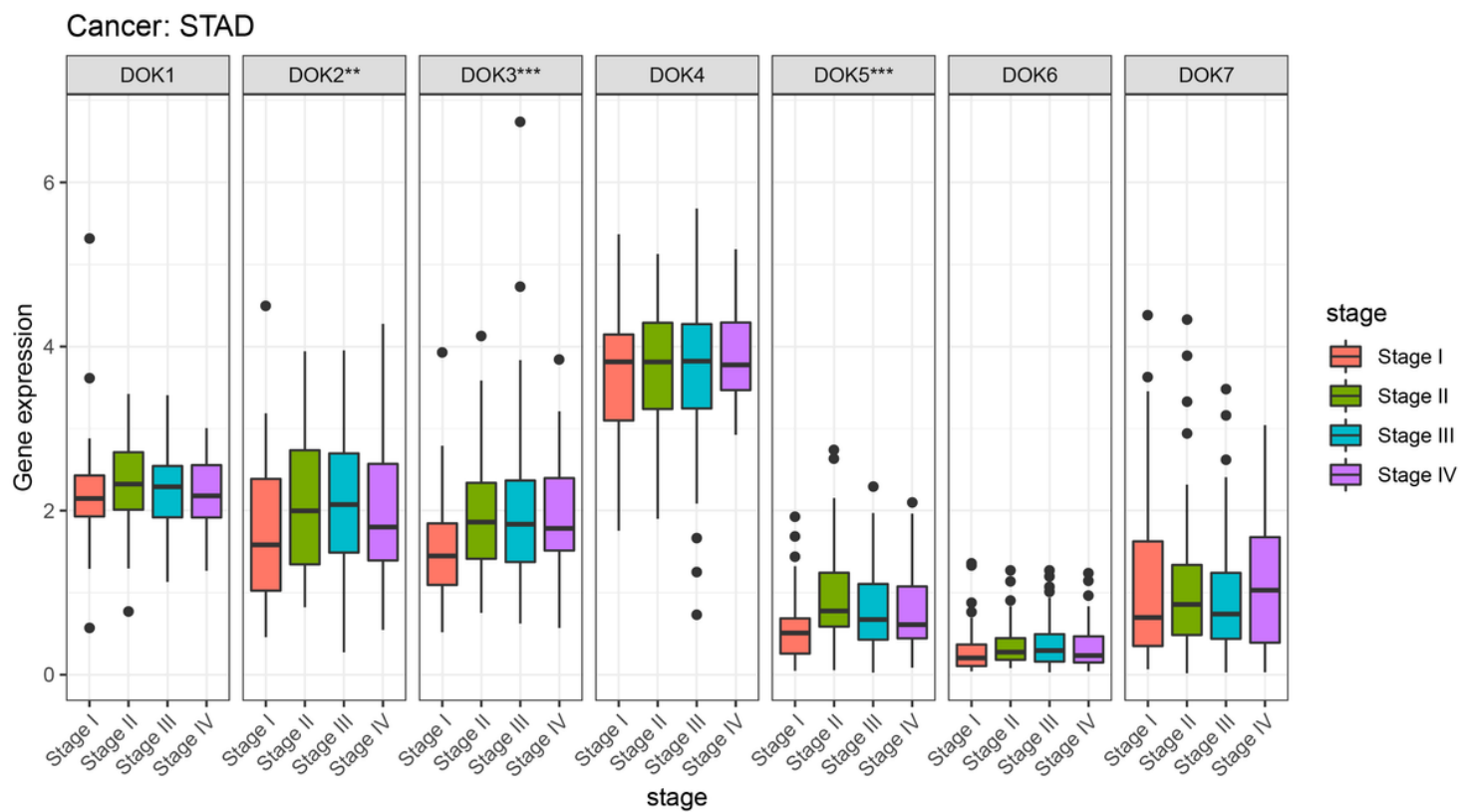
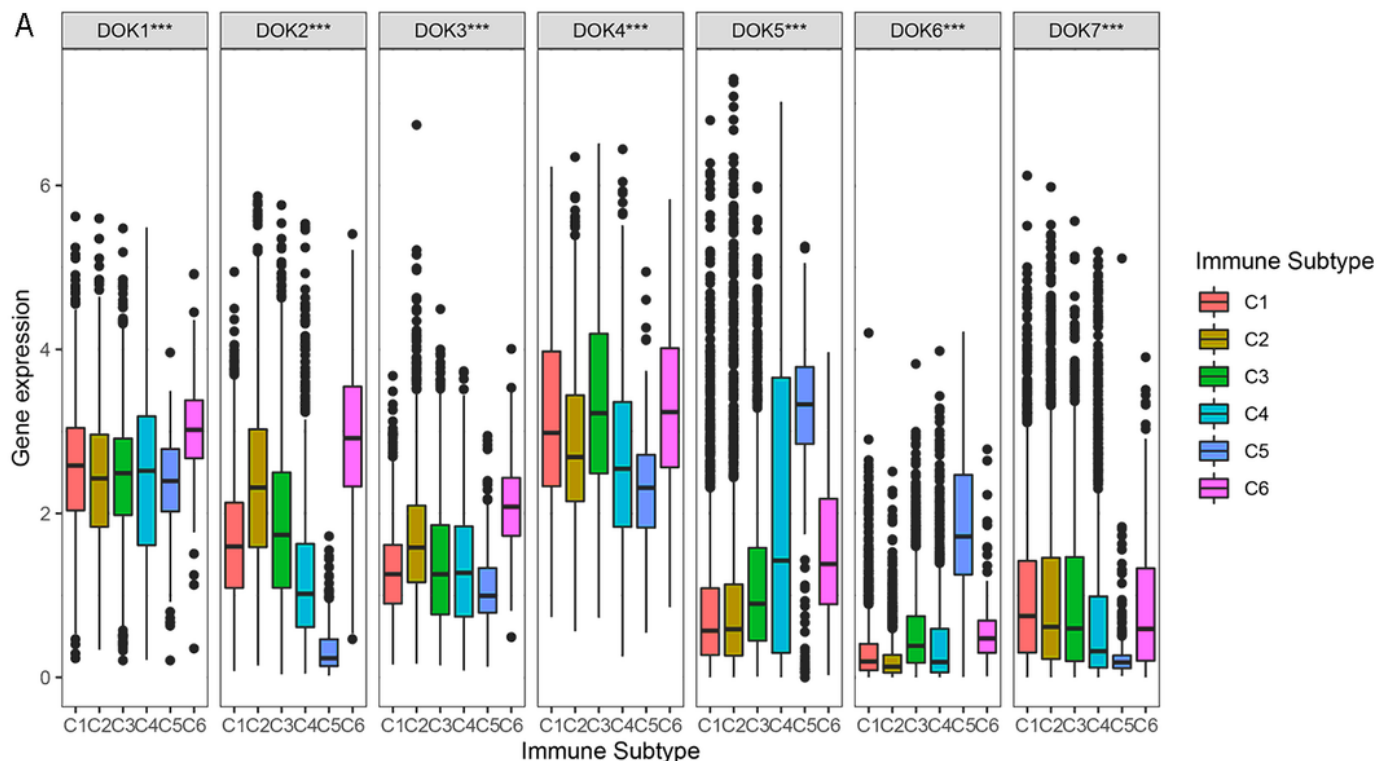


Figure 4

Differential gene expression level of DOKs in different tumor stages in STAD. (*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$)



Cancer: STAD

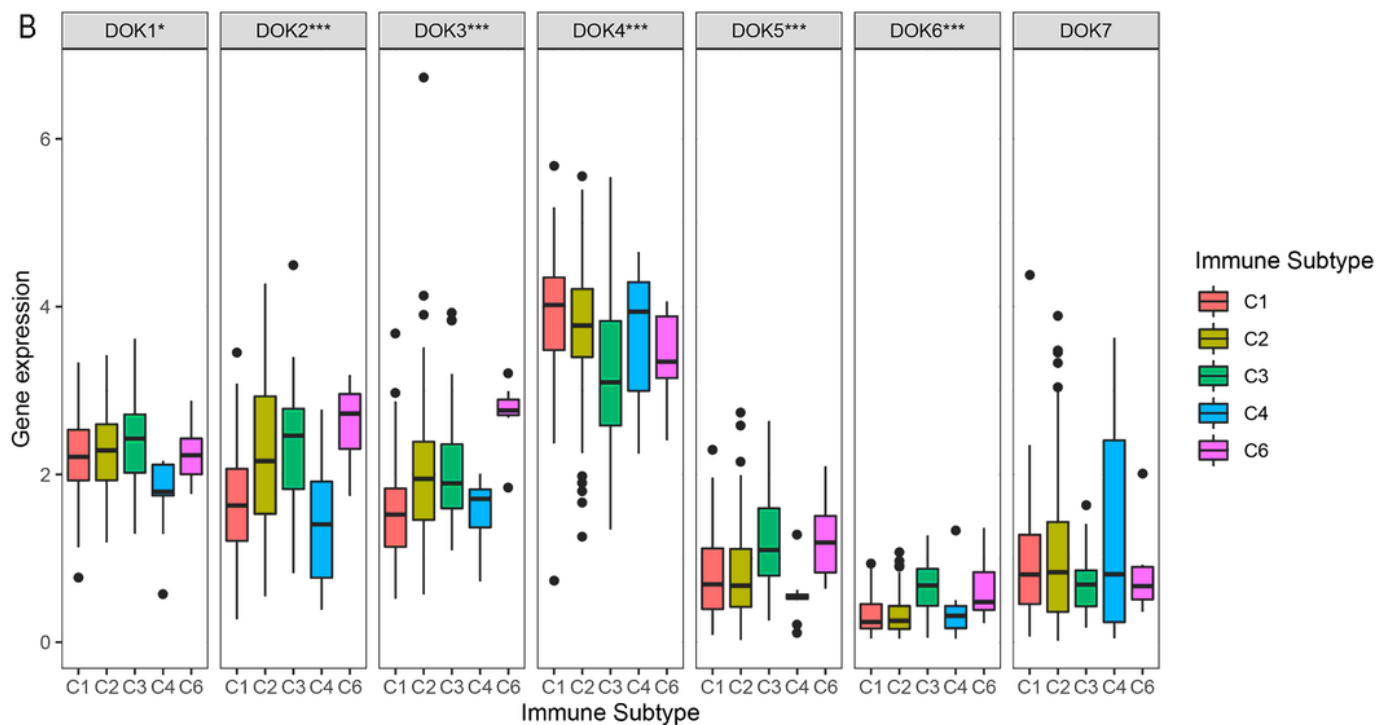


Figure 5

The results of correlation analysis between members of DOK and immune subtypes. (A) The transcriptional expression of DOKs in C1-C6 immune subtypes across TCGA cancers. (B) Box plots showing the expression level of DOK immune subtypes in STAD. (**p < 0:001; *p < 0:01; *p < 0:05)

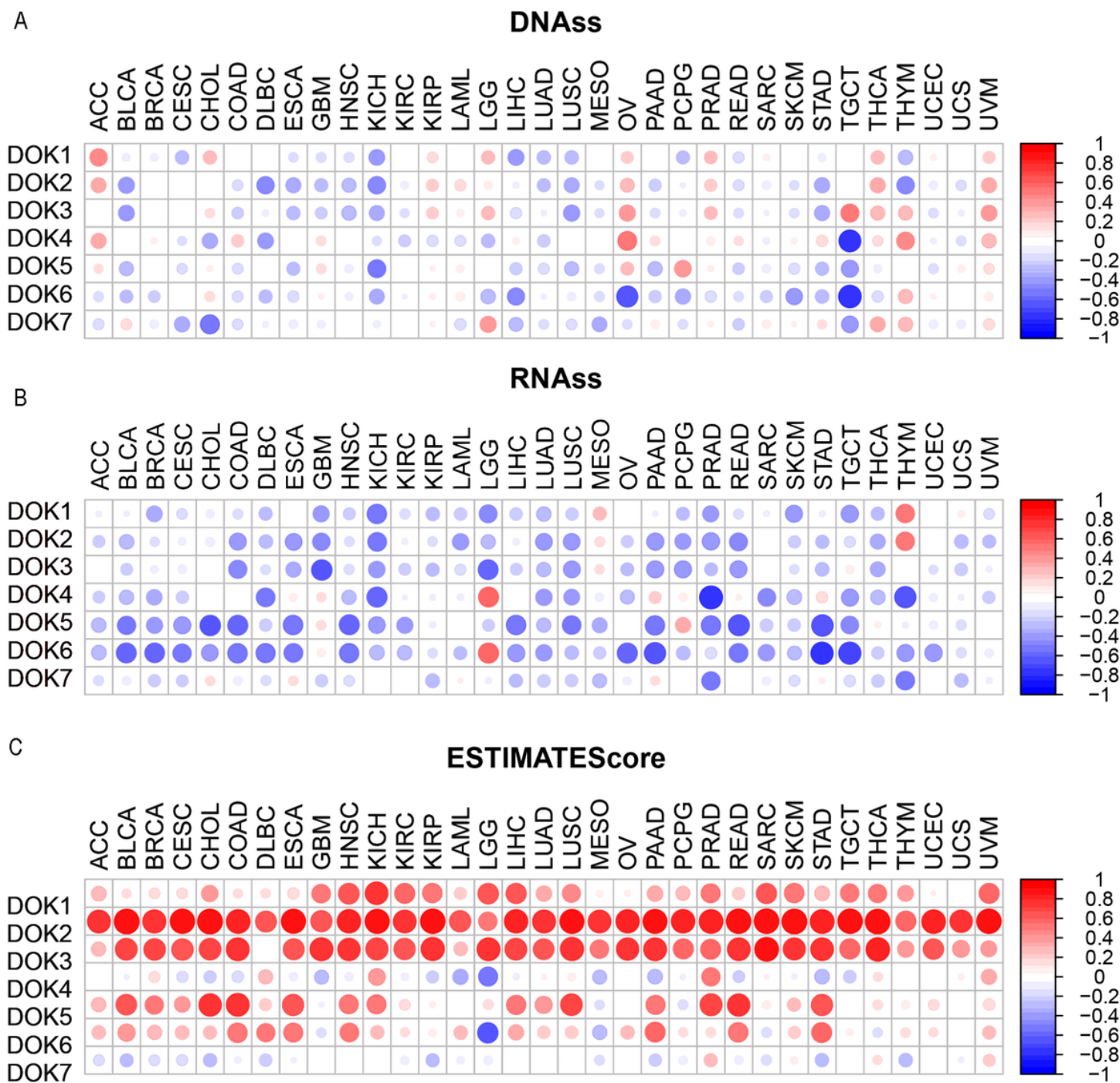


Figure 6

The results of correlation analysis between members of DOK and stemness indices and microenvironment scores. (A,B) The two heatmaps showing the correlation of the expression level of DOKs and stemness indices (DNAss and RNAss) in 33 TCGA cancer types. DNAss: DNA methylation-based stemness score; RNAss: RNA-based stemness score. (C) The heatmap showing the correlation between ESTIMATE scores and the mRNA expression of DOKs (red points represent a positive correlation while blue points represent a negative correlation).

Cancer: STAD

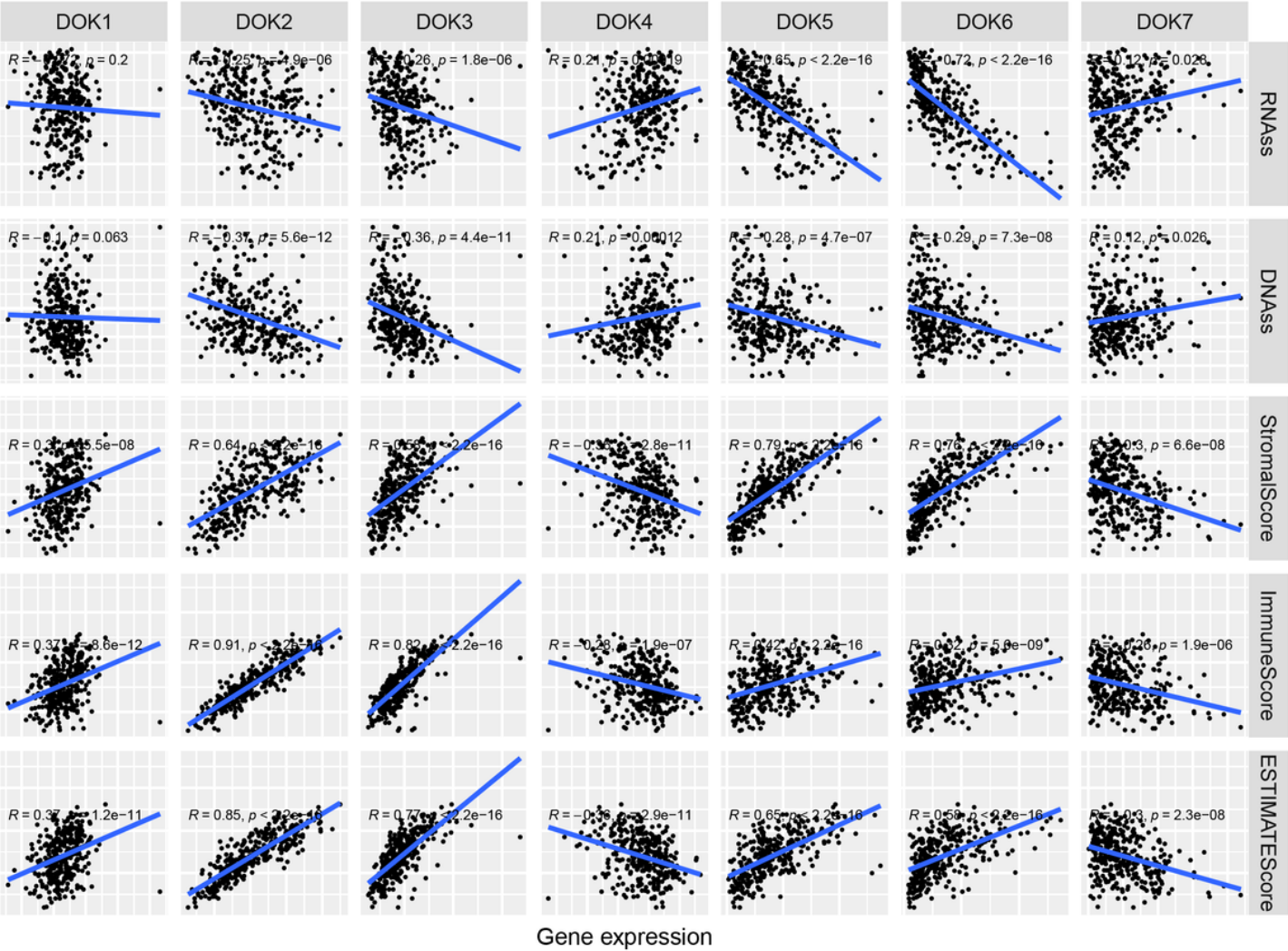


Figure 7

The correlation between DOKs and their coactivators and stemness scores (RNAss and DNAss), stromal scores, immune scores, and ESTIMATE scores in STAD.

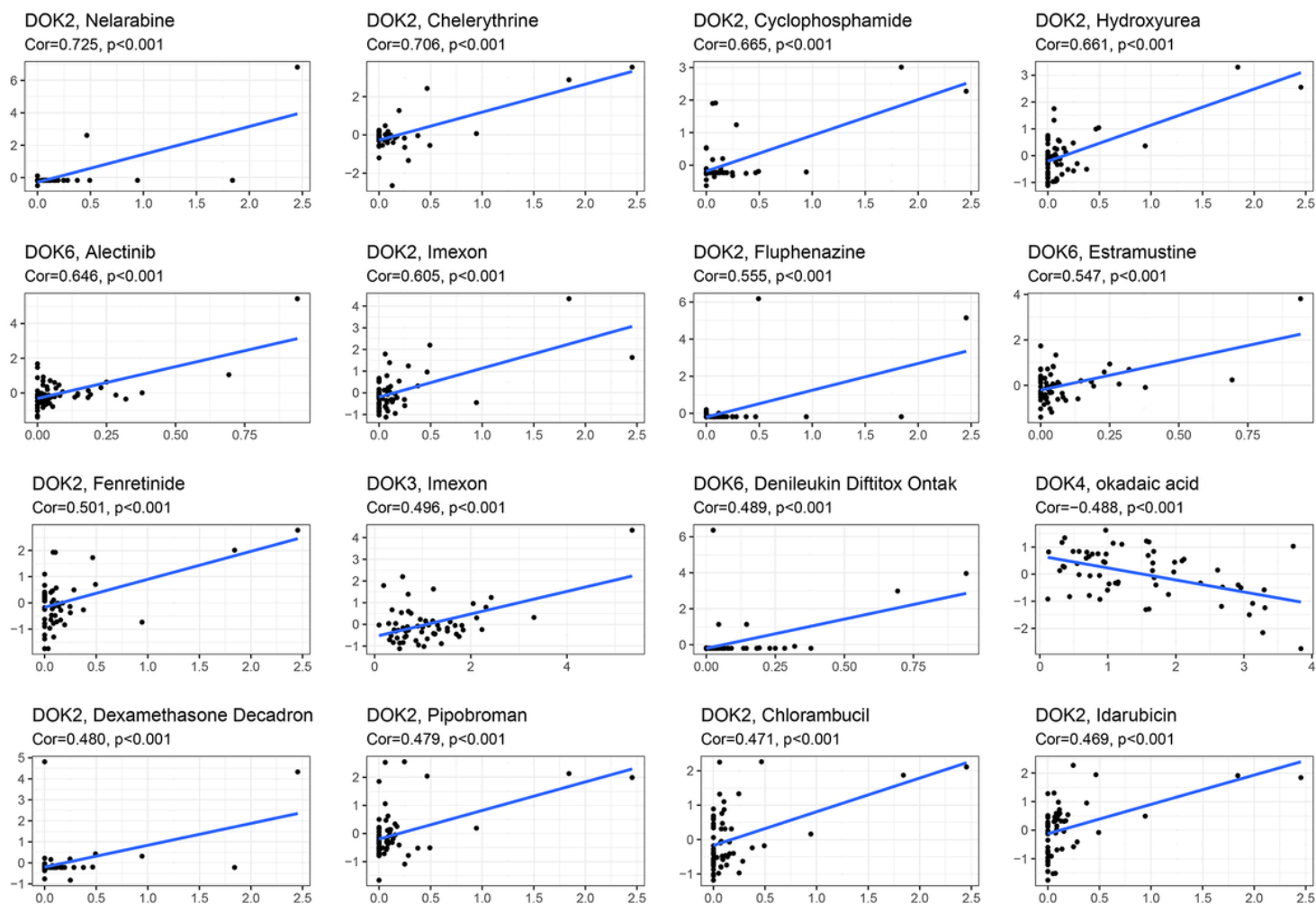


Figure 8

Drug response analysis. The correlation between drug sensitivity and DOKs across TCGA cancers. The scatter plots are ranked by p value.

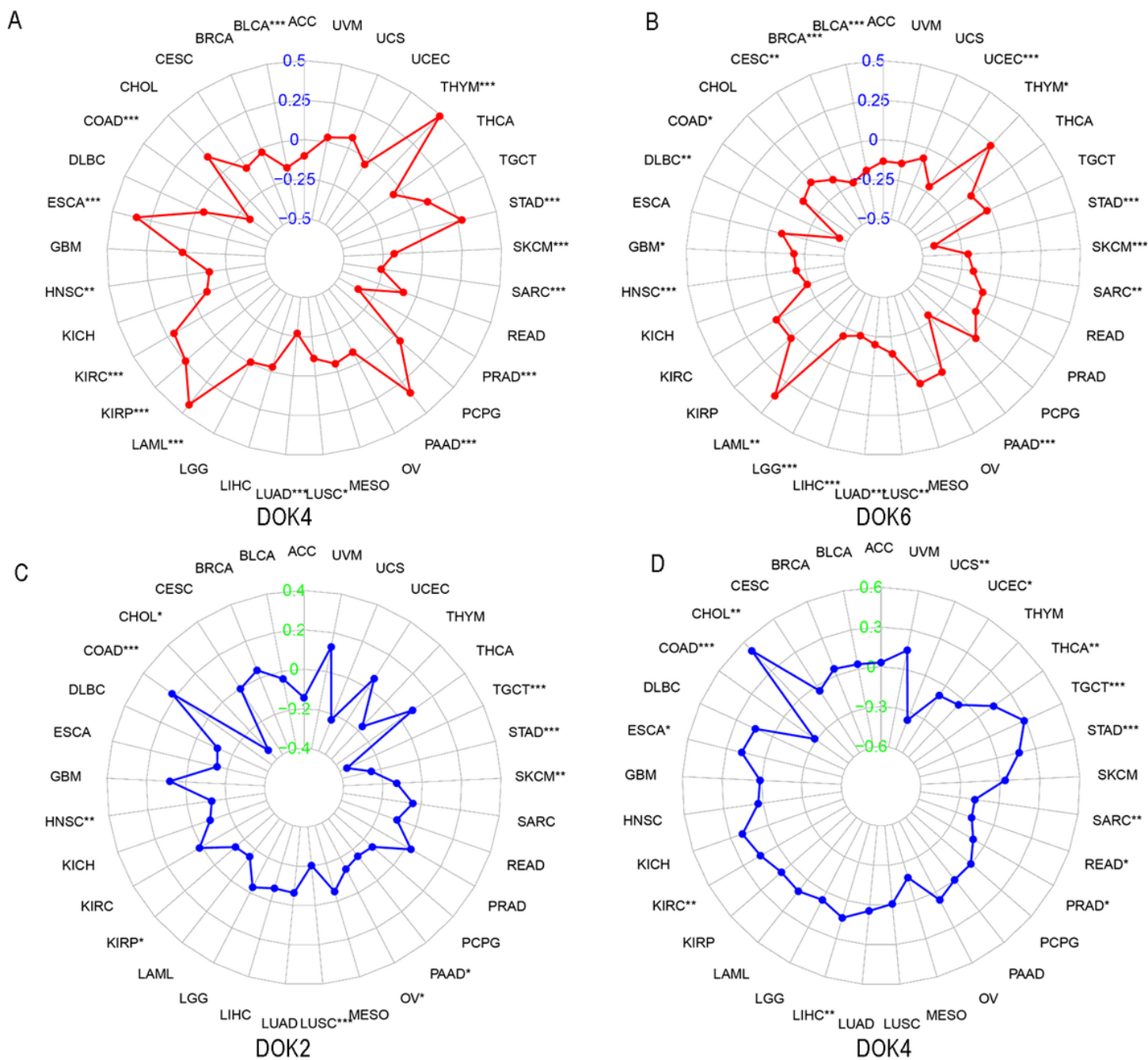


Figure 9

(A), (B) represent the tumor mutation load of DOK4 and DOK6. (C), (D) represent the microenvironment

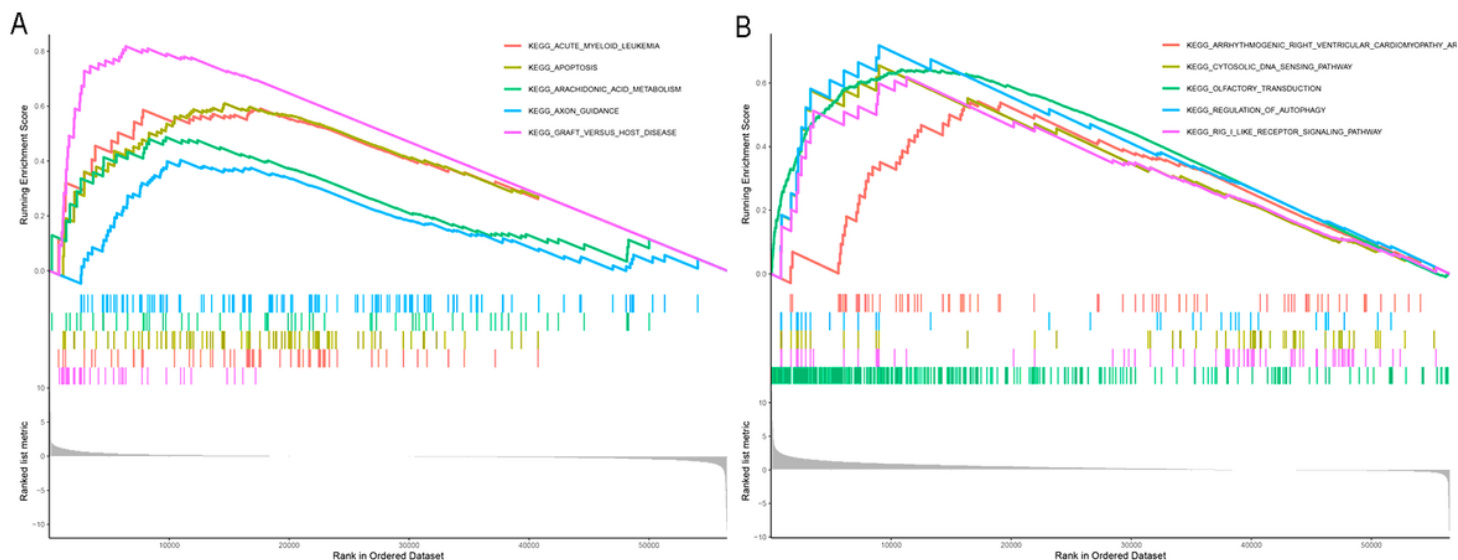


Figure 10

(A) and (B) respectively represent the pathway enrichment of DOK3 and DOK6 in STAD.

Supplementary Files

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

- [S1.tif](#)
- [S2.tif](#)
- [S3.tif](#)
- [S4.tif](#)
- [S5.tif](#)
- [S6.tif](#)