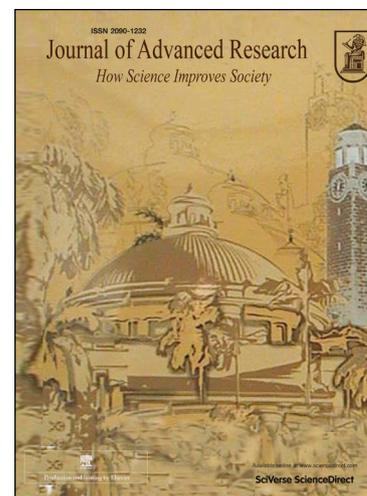


## Journal Pre-proofs



Comprehensive Analysis of DOK Family Genes Expression, Immune Characteristics, and Drug Sensitivity in Human Tumors

Yongjun Guan, Man Li, Zhendong Qiu, Junhui Xu, Yichao Zhang, Ning Hu, Xuanzhe Zhang, Wenyi Guo, Jingping Yuan, Qiao Shi, Weixing Wang

PII: S2090-1232(21)00104-1  
DOI: <https://doi.org/10.1016/j.jare.2021.06.008>  
Reference: JARE 1017

To appear in: *Journal of Advanced Research*

Received Date: 8 February 2021  
Revised Date: 31 May 2021  
Accepted Date: 9 June 2021

Please cite this article as: Guan, Y., Li, M., Qiu, Z., Xu, J., Zhang, Y., Hu, N., Zhang, X., Guo, W., Yuan, J., Shi, Q., Wang, W., Comprehensive Analysis of DOK Family Genes Expression, Immune Characteristics, and Drug Sensitivity in Human Tumors, *Journal of Advanced Research* (2021), doi: <https://doi.org/10.1016/j.jare.2021.06.008>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Cairo University

# Comprehensive Analysis of DOK Family Genes Expression, Immune Characteristics, and Drug Sensitivity in Human Tumors

Yongjun Guan<sup>a#</sup>, Man Li<sup>a#</sup>, Zhendong Qiu<sup>a</sup>, Junhui Xu<sup>a</sup>, Yichao Zhang<sup>a</sup>, Ning Hu<sup>a</sup>,  
Xuanzhe Zhang<sup>a</sup>, Wenyi Guo<sup>a</sup>, Jingping Yuan<sup>b</sup>, Qiao Shi<sup>c\*</sup>, Weixing Wang<sup>a\*</sup>

<sup>a</sup> Department of General Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China.

<sup>#</sup>Yongjun Guan and Man Li contributed equally.

<sup>b</sup> Department of Pathology, Renmin Hospital of Wuhan University, Wuhan 430060, China.

<sup>c</sup> Department of Pancreatic Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China.

## Corresponding Author:

\*Weixing Wang

Department of General Surgery, Renmin Hospital of Wuhan University, No. 238, Jiefang Road, Wuhan 430060, China. Electronic address: [sate.llite@163.com](mailto:sate.llite@163.com).

\*Qiao Shi

Department of Pancreatic Surgery, Renmin Hospital of Wuhan University, Wuhan 430060, China. Electronic address: [shiqiao614@163.com](mailto:shiqiao614@163.com).

## Data availability

The following information was supplied regarding data availability: TCGA raw data is available at UCSC xena browser (<https://xenabrowser.net/datapages/>). NCI-60 cell line data is available at CellMiner (<https://discover.nci.nih.gov/cellminer/home.do>).

## Ethics approval and consent to participate

The clinical sample collection was approved by the local ethics committee of the Renmin Hospital of Wuhan University (WDRY2019-K092), and written informed consent was obtained from each patient.

### **Author contributions**

GYJ, LM, WWX, and SQ came up with the design and conception. Material preparation, data collection, and analysis were performed by GYJ, QZD, XJH, ZYC, LM, HN, ZXZ, GWY, WWX, and SQ. Immunohistochemistry was performed by LM, SQ, and YJP completed. The first draft of the manuscript was written by GYJ, WWX, and SQ. GYJ and LM contributed equally to this study and shared co-first authors. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

### **Funding**

This study was funded by the National Natural Science Foundation of China (No. 81800574; No. 81870442) and the Natural Science Foundation of Hubei Province (No. 2018CFB649). Funding sources for this study had no role in the study design; data collection, analyses, or interpretation; or writing of the manuscript.

### **Abstract**

**Introduction.** *DOK* is a new type of regulatory protein family that participates in the regulation of tumor cell growth. However, most of the studies are conducted in cell lines, and systematic studies have not been conducted in human tumors.

**Objectives.** We conducted a comprehensive analysis of *DOK* based on its expression profile and its relationship with patient survival, immune infiltration, tumor microenvironment, and drug sensitivity.

**Methods.** We used the TCGA database to analyze the correlation between *DOK* family gene expression and prognosis and clinical stage. The protein expression of *DOK* in

tumor tissues was analyzed by immunohistochemistry. Use the cBioPortal database to analyze the alteration frequency in *DOK* family genes in human tumors. In addition, we used ESTIMATE algorithm and TIMER website to analyze the correlation between *DOK* family genes and tumor immunity. Finally, we further analyzed the relationship between *DOK* family genes and tumor stemness and the sensitivity of cancer cells to chemotherapy.

**Results.** We conducted a comprehensive analysis of *DOK* family genes based on its expression profile and its relationship with patient survival. We also confirmed this conclusion by immunohistochemistry. The expression of *DOK* family genes is related to OS, clinical stage, tumor mutation, methylation, CNV, and SNV. *DOK* family genes are significantly associated with poor prognosis of UVM. *DOK1-DOK3* has obvious correlation with tumor immunity. *DOK2* can increase the sensitivity of chemotherapy drugs, while *DOK4* reduces the sensitivity of multiple chemotherapy drugs. In addition, the expression level of *DOK* family genes is significantly correlated with the activity of cancer marker-related pathways.

**Conclusions.** *DOK* plays a role of tumor suppressor gene or tumor-promoting gene in different tumors. However, *DOK* family genes play a role in promoting cancer in UVM. *DOK* family genes are significantly associated with drug sensitivity.

**Keywords.** *DOK*; Tumor microenvironment; Immune infiltration; Drug sensitivity.

## Introduction

*DOK* (downstream of tyrosine kinase/Docking protein) is a tyrosine residue phosphorylated protein, it belongs to a new type of regulatory protein family, and plays an important role in the receptor tyrosine kinase signaling pathway[1]. So far, it has been found that there are 7 members of the *DOK* family (*DOK1-DOK7*), which have similar structural characteristics, that is, they all have an N-terminal PH (pleckstrin homology) domain, a central PTB (phosphotyrosine binding) domain and a C-terminal SH2 (src homology2) target motif[2-4]. According to gene structure and expression pattern, *DOK* family can be divided into three categories: *DOK1-DOK3*, *DOK4-DOK6* and *DOK7*. Among them, *DOK1-DOK3* is preferentially expressed in hematopoietic

and immune cells, and *DOK1* is highly expressed in myeloid and lymphocytes, *DOK2* is relatively highly expressed in T cells, and *DOK3* is relatively highly expressed in B cells[5, 6]. *DOK4-DOK6* are expressed in non-hematopoietic cells, *DOK4* is expressed in heart, skeletal muscle and lung tissues, and *DOK5* and *DOK6* are highly expressed in neurocytes[5, 6]. Studies have found that *DOK4*, *DOK5* and *DOK6* are positive regulators of the *MAPK* pathway. *DOK5* can promote the differentiation of PC12 cells by mediating RET, TrkB and other receptor signal transduction, thereby promoting neurite growth[7]. *DOK5* can also positively regulate the PTK pathway and does not bind to p120 rasGAP[8]. *DOK7* is preferentially expressed in skeletal muscle and heart, especially in the post-synaptic neuromuscular junction area, where it activates muscle-specific kinases to promote the accumulation of acetylcholine receptors on the postsynaptic membrane. Current studies have shown that the function of *DOK7* is less related to other members of the family[5, 6].

Recent studies have shown that abnormal *DOK* gene expression is closely related to leukemia[9], lung cancer[10], gastric cancer[11], colorectal cancer[11, 12], and breast cancer[13, 14]. However, most of the studies are conducted in cell lines and/or animal models, and systematic studies have not been conducted in human tumors. In this study, TCGA cancer data were used to study the relationship between the expression of this family genes in 33 cancers and overall survival, and its expression was analyzed in combination with tumor microenvironment and pharmacology to determine its potential function and unique prognostic value.

## **Materials & Methods**

### ***Human samples***

Pathological sections were collected from Department of Pathology, Renmin Hospital of Wuhan University (Wuhan City, Hubei Province) from January 2019 to December 2020. Remove adjacent normal tissue from the area > 2 cm from the primary tumor.

### ***Ethics statement***

All experiments involving human were conducted according to the ethical policies and procedures approved by the ethics committee of the Renmin Hospital of Wuhan

University, China (Approval no. WDRY2019-K092). All patients obtained written informed consent.

***Evaluating the expression and prognosis of DOK gene variants in different cancer***

Data of different types of cancers, including gene expression RNA-Seq (HTSeq-FPKM), clinical data, survival data, stemness scores based on mRNA (RNAss), and DNA-methylation (DNAss) were downloaded (March 2020) from xena browser(<https://xenabrowser.net/datapages/>). The cancer data of 33 primary tumors were described in supplementary table 1. The datasets include adrenocortical carcinoma(ACC), bladder urothelial carcinoma(BLCA), invasive breast 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), uveal melanoma(UVM). In total 14319 samples were available for this study, and each dataset contains normal samples and tumor samples (Table 1). Among them, 15 cancer types had none or less than 5 associated normal tissue samples, so only the rest of the 18 cancer types were used to investigate whether there was altered gene expression in tumors compared to adjacent normal tissues with linear mixed effects models (Table 1). The Wilcoxon test was used to study the changes in gene expression in adjacent normal tissues for these 18 types of cancers.

### ***Survival and clinical correlation analysis***

Cox regression analysis was used to evaluate the correlation between *DOK* expression in 33 types of cancers and the overall survival (OS). After dividing the patients into high and low *DOK* groups, the Kaplan-Meier method was used to construct a survival curve for each cancer type. The PrognScan database ([www.prognoscan.org](http://www.prognoscan.org)) combines several clinically annotated cancer microarray datasets from the GEO database and allows researchers to evaluate the expression of specific genes in cancer patients and their relationship with prognosis. We verified the relationship between *DOK* expression and prognosis in various types of cancer using the PrognScan database, including OS, recurrence-free survival (RFS), and DSS, with  $P < 0.05$  considered significant.

### ***Immunohistochemistry***

The incidence of lung adenocarcinoma, liver cancer, and colorectal cancer is relatively high, and *DOK1*, *DOK2*, *DOK4*, and *DOK6* are differentially expressed in these three cancers. Therefore, Immunohistochemical method was used to detect the expression of *DOK1*, *DOK2*, *DOK4*, and *DOK6* in lung adenocarcinoma, colorectal cancer, liver cancer patient tissues. The sections (4 $\mu$ m) were deparaffinized and dehydrated in xylene and a series of gradient ethanol solutions, and then pretreated in 10 mM citrate buffer (pH 6.0) at 121°C for 4 minutes. Hydrogen peroxide (0.3%) suspends the activity of endogenous peroxidase for 15 minutes at room temperature. Incubate the sections with *DOK1* (1:200, cat. no. PA87543HU, Abebio, Wuhan, China), *DOK2* (1:200, cat. no. PA84257HU Abebio, Wuhan, China), *DOK4* (1:200, cat. no. PA85947HU, Abebio, Wuhan, China) and *DOK6* (1:200, cat. no. PA85948HU, Abebio, Wuhan, China) primary antibodies at 4°C overnight. Finally, the sections were counterstained with hematoxylin. Image Pro-Plus (version 6.0; Media Cybernetics, Inc. Rockville, Maryland, USA) was used to analyze IHC staining. Select the positive staining area as the target area (AOI). The brown reaction product represents the positive expression of *DOK1*, *DOK2*, *DOK4*, and *DOK6*. Randomly select the area in the microscope field of view in each area, and take photos with 200x magnification in each group. Analyze each image and use Image-Pro Plus 6.0 software to generate the mean of integral optical density (MOD) value. The average MOD value of all images taken in each group is

used and finally the average expression index of each repeated sample is statistically analyzed.

#### ***Mutation and Methylation analysis***

We used the cBioPortal database (<https://www.cbioportal.org>) and the Cancer Cell Line Encyclopedia (CCLE) (<https://portals.broadinstitute.org/ccle>) to analyze the mutational changes in *DOK* family genes in a variety of cancers [15]. The cBioPortal data were obtained from the TCGA database and included data from 10,953 patients (10,967 samples). GSCALite (<http://bioinfo.life.hust.edu.cn/web/GSCALite/>) is a web-based analysis platform for genomic cancer analysis. We used GSCALite to analyze mRNA expression, Copy Number Variation (CNV), Single Nucleotide Variation (SNV), methylation, and pathway activity.

#### ***Tumor microenvironment and immune infiltration analysis***

ESTIMATE is an algorithm that uses gene expression signatures to infer the fraction of stromal and immune cells in tumor samples. The estimate score from this program was used to describe tumor purity. More information can be acquired online at <https://bioinformatics.mdanderson.org/public-software/estimate/>. Spearman correlation was used to analyze the correlation between *DOK* expression and estimate score, immune score, stromal score, tumor purity. The tumor stem features extracted from the transcriptome and epigenetics of TCGA tumor samples were used to measure the stem cell-like features of tumor cells. The correlation between cancer stemness and *DOK* expression was investigated by Spearman correlation test. The TIMER database (<https://cistrome.shinyapps.io/timer/>) contains 10,897 samples across 32 cancer types from TCGA to allow the evaluation of the abundance of immune infiltration[16]. We analyzed *DOK* expression in different types of cancers and the correlation of *DOK* expression with the abundance of immune infiltrates, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells. The relationship between the expression level of *DOK* and tumor purity was also determined.

#### ***Drug sensitivity analysis***

The National Cancer Institute (NCI) established a screening platform for 60 kinds of

whole-cell anticancer drugs (NCI-60 DTP Human Tumor Cell Line Drug Screen), which is a classic tool for in vitro drug screening. CellMiner provides the "NCI-60 Analysis Tool", which can quickly retrieve the transcripts of 22,379 genes and 360 microRNAs, as well as activity reports of 20,503 compounds. Therefore, we can calculate the Pearson correlation coefficient (PCC) to analyze the relationship between the mRNA expression and the 50% growth inhibitory concentrations of the drugs. The correlation analyses used 262 FDA-approved drug reactions or drugs in clinical trials.

### ***GEPIA dataset***

GEPIA is a newly developed interactive web server that uses standard processing pipelines to analyze the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the Cancer Genome Atlas and Genotype Tissue Expression (GTEx) project. We used GEPIA to analyze the correlation between *DOK* family genes.

### ***Statistical analyses***

We used t-test for normal distribution data. Non-normal distribution data sets were analyzed using Wilcoxon test. The Wilcoxon test was used to compare gene expression in normal tissues and in 18 types of cancers. All cancers had more than five adjacent normal samples and boxplots were used to describe gene expression data in each of the various cancers discussed in this study. Univariate Cox proportional hazards regression models were used to the association between gene expression and OS. Spearman or Pearson correlations were used to analyze the correlation between gene expression and Stemness score, stromal score, immune score, estimate score, immune checkpoints, and drug sensitivity.

## **Results**

### ***DOK expression in human tumors***

To understand the expression of the *DOK* in tumors, we analyzed 31 types of cancers in the TCGA database, and 18 of them have at least 5 pairs of normal control samples. The results showed that the expression levels of all genes in the *DOK* family members were significantly different in these 18 tumors (Fig. 1A-C). For example, the expression

level of *DOK4* exhibited the most apparent differences in various tumors, with the lowest expression level in BRCA, KICH, PRAD, LUSC, and the highest expression level in CHOL, KIRC, KIRP (Fig. 1G). Other members of the *DOK* family also showed different degrees of differences (Fig. 1D-J). Compared with other *DOK* family members (such as *DOK1*, *DOK2*, and *DOK3*), the average expression levels of *DOK5*, *DOK6* and *DOK7* in various cancers were lower. These analyses proved that there are apparent differences in the expression of *DOK* family members among different tumor types. However, the expression of each gene in the *DOK* family changes in different directions in various cancers (Fig. 1D-J). It was mainly up-regulated in GBM, KIRC, CHOL, THCA, KIRP, LIHC, and STAD, while down-regulated in LUSC and LUAD, but there were also some exceptions. In order to study the expression of *DOK* in different cancer, the expression of *DOK1*, *DOK2*, *DOK4*, and *DOK6* in human liver cancer, lung adenocarcinoma and colorectal cancer was examined by immunohistochemical methods. We collected 15 samples of lung adenocarcinoma, 15 samples of liver cancer, and 10 samples of colorectal cancer. The expression of *DOK1*, *DOK2*, *DOK4*, and *DOK6* was evaluated by IHC. The results showed that *DOK1*, *DOK2*, *DOK4*, and *DOK6* were lowly expressed in lung adenocarcinoma (Fig. 2A). Consistently, the quantitative score of staining intensity showed that the staining of *DOK1*, *DOK2*, *DOK4*, and *DOK6* in lung adenocarcinoma was significantly lower than the staining of *DOK1*, *DOK2*, *DOK4*, and *DOK6* in adjacent normal tissues (Fig. 2D). These results indicate that *DOK1*, *DOK2*, *DOK4*, and *DOK6* are lowly expressed in human lung adenocarcinoma tissues. The same method was used to analyze the expression of *DOK1*, *DOK2*, and *DOK6* in liver cancer and Colorectal cancer. The results show that *DOK1* is highly expressed in liver cancer and colon cancer (), while *DOK2* and *DOK6* are low in liver cancer and colorectal cancer (Fig. 2B, 2C, 2E, 2F). Subsequently, we analyzed the survival of these samples and found that *DOK1*, *DOK2*, and *DOK6* were related to the OS of patients with lung adenocarcinoma, and patients with low expression had a poorer prognosis (Fig 2G-2I). *DOK2* is related to the OS of liver cancer patients, and patients with low expression have a poorer prognosis (Fig 2J).

We analyzed the mRNA expression patterns of the *DOK* genes at different stages of

these diseases and found that changes in the expression of various *DOK* genes are related to the clinical stage of various cancers (Supplementary 1). We found that *DOK1* is not only linked to OS in LUAD and KIRC but can also indicate clinical stage in both of these diseases. *DOK3* is a known prognostic marker for KIRC but is also linked to changes in the clinical stage of these cancers. *DOK4* expression is linked to OS and cancer staging in UVM and KIRP. *DOK5* is linked to OS and cancer staging in UVM. *DOK6* is a prognostic marker for OS and cancer stage in BLCA and TGCT. *DOK7* is a marker for OS and cancer stage in BLCA, KIRC, and UVM. In addition, *DOK1*, *DOK2*, and *DOK3* are all markers for UCEC staging. *DOK1*, *DOK2*, and *DOK6* are all linked to the clinical stage of LUAD. *DOK1*, *DOK3*, *DOK4*, and *DOK6* expression are all markers for LGG staging. *DOK4*, *DOK5*, and *DOK7* are all linked to the clinical stage in UVM. *DOK3* and *DOK5* transcription were also linked to SKCM stage and *DOK1*, *DOK5*, and *DOK6* were determined to be markers of ESCA staging. The results of *DOK1* and LUAD, READ staging, *DOK4* and UVM staging, *DOK6* and TGCT staging, are close to the significance threshold. Therefore, it is also worthy of our consideration.

#### ***The relationship between DOK family genes expression and patient survival rate***

In order to determine the roles of genes in the *DOK* family genes in cancer progress and prognosis, the relationship between the expression of *DOK* family genes and the overall survival of patients in 33 cancers were analyzed. Using a univariate cox proportional hazard regression model, we claim that a  $P < 0.05$  indicated a significant correlation, and it could be made consistent with the forest plot in Figure 1K without multiple comparison adjustments (Supplementary 2). Our research found that the expression of *DOK* family genes was related to the overall survival rate of patients, but even the same gene showed different prognostic correlations with different tumors (Fig.1K and Supplementary 3). Among them, *DOK1* was related to poor prognosis of KIRC, LGG, UVM, while associated with higher survival rates in LUAD, MESO, SKCM, and UCEC. *DOK2* was related to poor prognosis of GBM, LAML, LGG, UVM, while associated with higher survival rates in HNSC, LUAD, SARC, SKCM, and UCEC. *DOK3* was associated with poor prognoses of GBM, KIRC, LGG, and THYM, while the prognoses were better in CESC, HNSC, PCPG, SARC, and UCEC. *DOK4*

was associated with poor prognosis of SKCM, UVM, and good prognosis of KIRP, LGG. *DOK5* was associated with poor prognosis of LAML, THYM, STAD, UCEC, UVM and good prognosis of ACC, LGG, LIHC, PCPG. Although *DOK6* predicted a good prognosis in LGG, it had a poor prognosis in other tumors. Except for *DOK3*, the other *DOK* family genes were all related to the poor prognosis of UVM. In addition, *DOK4*, *DOK5* and *DOK7* expression is linked to OS and cancer staging in UVM. Therefore, *DOK* family genes are significantly associated with poor prognosis of UVM. We used the PrognoScan Database to analyze the relationship between *DOK* family gene expression and survival in different cancers. The results showed that *DOK* family genes are related to the prognosis of many tumors. and the results are summarized in Table 2 and Supplementary 4.

#### ***DOK family genes was related to genomic alteration***

We evaluated the impact of the various *DOK* family genes alteration frequency using the cBio-Portal database. The alteration frequencies in *DOK2*, *DOK5*, *DOK6*, and *DOK7* are relatively high, while the alteration frequency in *DOK4* was shown to be the lowest. *DOK2* has the highest alteration frequency in prostate cancer and bladder cancer, *DOK3* has the highest alteration frequency in ccRCC, and *DOK5* has the highest alteration frequency in colorectal cancer (Fig. 3A-C and Supplementary 5). Then we used CCLE to analyze mutation data in 967 cell lines of 23 cancers. These data revealed that *DOK* family genes have relatively high mutation frequencies in many types of cancers (Fig. 3D). Finally, the frequency of changes in CNV and SNV in *DOK* family genes was analyzed, and the results showed that CNV and SNV had a higher frequency of changes (Fig. 3E-F). Using methylation level and expression profile data, we further analyzed the effect of promoter methylation on *DOK* family gene changes in 33 types of cancer (Figure 3G-H). Our results indicate that *DOK* family genes are methylated in a variety of tumors. Specifically, *DOK5* and *DOK6* are hypermethylated in most tumors. In addition, *DOK* family genes are mostly hypomethylated in KIRC and KIRP, but hypermethylated in HNSC and COAD. *DOK* family gene promoters are negatively correlated with their gene expression. These results reveal the highly heterogeneous

inheritance and expression changes of *DOK* family genes in different types of cancer.

***DOK family genes was related to tumor microenvironment and immune response***

Studies have shown that the expression of *DOK4* and *DOK5* mRNA could be detected in human T cells, which implies that they are involved in immune regulation[8]. In addition, it has also been reported that *DOK4* acts as a negative regulator of T cell activation[17]. To understand whether all *DOK* family members were related to tumor immunity, we analyzed the correlation between *DOK* and tumor immunity. *DOK1-DOK3* has the most significant correlation with immune score in various cancers (Fig.4A and Supplementary 6). Among them, *DOK2* had the strongest correlation ( $r=0.94$ ,  $P<0.001$ ), followed by *DOK3* ( $r= 0.86$ ,  $P<0.001$ ) and *DOK1* ( $r= 0.70$ ,  $P<0.001$ ) (Supplementary 6). *DOK4-DOK6* had both positive and negative correlations with immune score in various cancers. *DOK7* was less correlated with immune score in different cancers. We further found that all members of the *DOK* family were correlated with immune scores in LGG, STAD, PAAD, THCA, and TGCT. In addition, we also analyzed the estimate score and stromal score in *DOK* and tumors, and the results were similar to that of immune scores (Fig.4B-C and Supplementary 6). The correlation between *DOK* expression and immune infiltration levels in 32 cancers of TIMER was also analyzed (Fig.4D-F and Supplementary 7). The results showed that *DOK1-DOK7* were significantly correlated with infiltration of B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells in different tumors, among which *DOK1-DOK3* were the most significantly and positively correlated genes. *DOK7* showed the least correlation with immune cell infiltration, and was only related to immune infiltration in THCA, THYM, and DLBC. In addition, *DOK2* and *DOK3* had a significant and negative correlation with tumor purity in most tumors, while *DOK7* showed a small correlation with tumor purity. Finally, we analyze the relationship between *DOK* family genes expression and immune checkpoint gene expression (Fig.5 and Supplementary 8). We found that there is a high correlation between *DOK* family genes and immune checkpoints ( $P<0.05$ ). Among them, the correlation between *DOK1-DOK3* and immune checkpoint is the strongest and mostly positive correlation. The

correlation between *DOK6* and *DOK7* and immune checkpoint is the weakest.

***DOK family genes was related to tumor stemness and sensitivity of cancer cells to chemotherapy***

Tumors are a heterogeneous cell population. A small number of stem-like cells are similar to embryonic stem cells and have unlimited self-renewal and division capabilities, which facilitate the occurrence and development of tumors. In recent years, many studies have suggested that the existence of cancer stem cells is the leading cause of tumor recurrence and metastasis, and is closely related to chemotherapy resistance. Tumor stemness can be measured by RNA stemness score based on mRNA expression (RNAss) and DNA stemness based on DNA methylation pattern (DNAss). *DOK* family members showed various levels of association with RNAss and DNAss in different cancers (Fig.6A-B and Supplementary 9). We found that *DOK5* and *DOK6* were negatively correlated with RNAss and DNAss ( $P<0.05$ ), and *DOK6* has the most apparent correlation with RNAss ( $r=-0.73$ ). *DOK2*, *DOK3*, and *DOK4* were negatively correlated with RNAs, but were either positively or negatively correlated with DNAss in different tumors. *DOK1* and *DOK7* had relatively weak correlation with RNAs and DNAss ( $P<0.05$ ). Since *DOK* is related to stem cell-like characteristics, we next analyzed the expression of *DOK* in human cancer cell lines (NCI-60) as well as the sensitivity to chemotherapy drugs. Our study found that the increased expression of *DOK* was related to the increased sensitivity of a variety of chemotherapeutic drugs ( $r>0.4$  and  $P<0.001$ ), especially *DOK2*, *DOK4*, and *DOK6* (Fig.6C and Table 3). The increase in *DOK2* expression was related to the increased sensitivity of cancer cells to Chelerythrine, Hydroxyurea, Fenretinide, Idarubicin, Uracil mustard, and Chlorambucil. In addition, increased expression of *DOK6* also increased the sensitivity of cancer cells to a variety of chemotherapy drugs, including Ethinyl estradiol, Estramustine, Etoposide, Carmustine, and 7-Hydroxystaurosporine. However, the increase in *DOK4* expression was associated with increased cell resistance to drugs, such as Carfilzomib, Depsipeptide, Pipamperone, Eribulin mesilate, Vinblastine, Actinomycin D.

### ***Cancer-related pathways regulated by DOK family genes***

In order to further understand the molecular mechanism of *DOK* family genes involved in cancer, we analyzed the correlation between the expression of *DOK* family genes and 10 cancer-related pathways (Fig.6D). As a result, the expression of *DOK* family genes is related to the activation or inhibition of multiple oncogenic pathways. The expression of *DOK* family genes is related to the activation of EMT and hormone ER pathways, and is related to the inhibition of DNA damage response, cell cycle, and hormone AR pathways. In addition, genes cannot play a role alone. Interestingly, we found that there are highly correlated interaction patterns between *DOK* family genes. Among them, *DOK1* and *DOK3* ( $r=0.45$ ,  $P<0.001$ ), *DOK2* and *DOK3* ( $r=0.66$ ,  $P<0.001$ ) have the highest correlation among these seven genes, indicating that they may share some common features or functions (Fig.6E). The results of the GEPIA database also confirmed that there is a strong correlation between the *DOK* family genes (Fig.6F-H).

### **Discussion**

*DOK* family genes disorders have been reported in many cancers [9, 18-20]. Although *DOK* family genes expression is closely associated related with pathological progression and poorer prognosis in some tumors, there has been no comprehensive analysis of *DOK* family genes in different cancer. This study is the first to explore the expression and prognostic value of *DOK* family genes in human tumors, as well as the mechanisms underlying tumor progression.

As the first family member discovered, *DOK1* was named p62dok according to its molecular weight. Because *DOK1* can bind to p120 rasGAP, thereby inhibiting the Ras-Raf-MEK-ERK signaling pathway[21, 22], *DOK1* is considered to act as a tumor suppressor gene in most tumor[23-25]. Initially, *DOK1* was only related to leukemia and Burkitt lymphoma. However, in recent years, studies have found that the low expression of *DOK1* mRNA is also associated with non-hematological tumors. For example, knocking out *DOK1* in epithelial ovarian cancer (EOC) can promote EOC cell migration, proliferation and cisplatin sensitivity[24]. In the study of mouse lung cancer

models, it was found that knocking out *DOK1*, *DOK2*, and *DOK3* alone can promote the occurrence of lung cancer. In addition, knocking out *DOK1*, *DOK2*, and *DOK3* at the same time can significantly promote the formation of lung cancer, so *DOK1-DOK3* can synergistically inhibit lung tumors. Our study found that *DOK1* has a better prognosis in LUAD, MESO, SKCM, and UCEC, and may play a tumor suppressor function in these tumors. However, *DOK1* does not play a tumor suppressor function in all tumors, and can play a tumor-promoting function in other tumors. For example, in glioma cells, *DOK1* promotes the invasion of glioma cells through the p130Cas-Rap1 signaling pathway. Our study also found that *DOK1* is associated with poor prognosis of KIRC, LGG, and UVM. In the study of tumor staging, it was found that *DOK1* was highly expressed in low-grade LUAD and SKCM, but the opposite was true in KIRC. It is further confirmed that *DOK1* plays a role of suppressing or promoting cancer in different tumors.

*DOK2* and *DOK3* are similar to *DOK1* and act as tumor suppressor genes in most tumors, including lung cancer, gastric cancer, colorectal cancer, and acute myeloid leukemia[26-28]. Here, we found that *DOK2* expression indicates an improved prognosis in HNSC, LUAD, SARC, SKCM, and UCEC, but a poor prognosis in GBM, LAML, and UVM. Similarly, *DOK3* expression was linked to a better prognosis in CESC, HNSC, PCPG, SARC, and UCEC, but a worse prognosis in GBM, KIRC, LGG, and THYM. Therefore, *DOK2* and *DOK3* do not act exclusively as tumor suppressor genes and may promote the proliferation of certain tumors. In support of this hypothesis, we found that *DOK2* expression could also be linked to tumor stage with both early stage COAD and TGCT exhibiting increased *DOK2* expression and BLCA and KIRC demonstrating increased *DOK2* expression in their later stages of development. *DOK3* expression was higher in LUAD and TGCT in the earlier stages of diseases but higher in the later stages of both KIRP and STAD.

Since *DOK* family proteins have no obvious catalytic activity, their biological effects are likely attributed to the activity of their binding partners. Therefore, *DOK1-DOK3* may interact with other proteins which affect tumor proliferation. Our study found that *DOK1*, *DOK2*, and *DOK3* expression were all low in LUAD tumors and that this

reduction was linked to poorer prognosis. The results of the correlation analysis showed that the expression of *DOK1* and *DOK2* are correlated and that *DOK2* and *DOK3* expression are also correlated. Niki[29] found that mice with a single deletion in *DOK1* or *DOK2* maintained normal hematopoiesis, but successive deletions in *DOK1* and *DOK2* led to abnormal hematopoiesis and activation of Ras/mitogen activated protein kinase in mice. In addition, mice with deleted *DOK1* and *DOK2* exhibited increased cellular proliferation and decreased apoptosis, thus facilitating the production of transplantable chronic myeloid leukemia-like bone marrow and external myeloproliferative diseases. Berger [26] confirmed that *DOK1*, *DOK2*, and *DOK3* act as lung cancer suppressor genes and that *DOK1*, *DOK2*, and *DOK3* single and (or) combined gene knockout promotes the occurrence of lung cancer in mice, while *DOK1*, *DOK2*, and *DOK3* exerted a synergistic effect in suppressing lung cancer in these models. In addition, studies have found that the *DOK1*, *DOK2*, and *DOK3* are closely associated with immunity, which we confirmed in our study when *DOK1*, *DOK2*, and *DOK3* expression were all positively correlated with stromal scores and immune cell infiltration levels in a variety of cancers. However, the underlying mechanism allowing *DOK1-DOK3* interactions to exert synergistic effects in tumor progression/suppression remains unknown. *DOK4 -DOK6* have been less studied in tumors. Previous studies have shown that *DOK4* and *DOK6* are weakly expressed in breast cancer and play a tumor suppressor function[14, 30]. However, *DOK5* and *DOK6* are highly expressed in AML and are associated with poor prognosis[31]. Our study also found that UVM has a poor prognosis when *DOK4-DOK6* are expressed and STAD has a poor prognosis when *DOK4* and *DOK6* are expressed. In contrast LGG has a better prognosis when *DOK4* and *DOK6* are expressed. Our research found that the expression of *DOK5* and *DOK6* are positively correlated with each other.

*DOK7* is different from the other members of the *DOK* family in that it is preferentially expressed in muscle tissue and participates in the production and maintenance of neuromuscular junctions, making it essential for neuronal postsynaptic differentiation[32]. Therefore, research on *DOK7* has mainly focused on neuromuscular junction diseases, such as myasthenia gravis[33, 34]. However, recent studies have

suggested that the *DOK7* is closely related to the occurrence and development of lung and breast cancer[35, 36]. *DOK7* expression decreases significantly in these tissues resulting in increased tumor cell proliferation and cloning ability, thereby accelerating pathological progression and worsening the prognosis. Our study showed that reduced expression of *DOK7* was associated with poor prognosis in KIRC, UVM, and promoted a survival advantage in BLCA, BRCA, STAD, and THCA. The expression of *DOK7* is also related to changes in the clinicopathological characteristics of several cancers. *DOK7* is highly expressed in patients with stage III-IV CHOL, KIRC, THCA, and UVM, but reduced in stage I-II of these cancers. In contrast, BLCA and KIRP present with high levels of *DOK7* in stage I–II which gradually reduces in stage III–IV. Based on this data, we can conclude that the *DOK* play an important role in suppressing or promoting cancer in different tumors. The reason for the differences in their activity may be explained by the heterogeneity among *DOK* family members, which results in differences in effector partners and signaling efficiencies. In addition, the expression level of a gene is related to the gene itself as well as to the upstream and downstream regulatory sequences (including promoters, enhancers, etc.) and introns of the gene. For example, it has been demonstrated that *E2F1* and *ATRA* have a significant positive effect on the *DOK1* promoter and promote *DOK1* expression, while *CREB1*, *SP1* and *p53* can inhibit *DOK1* expression [37].

In addition, the analysis of 118 breast cancer samples revealed that the expression levels of *DOK1* mRNA were significantly reduced in breast cancer which was associated with changes in the clinicopathological characteristics of this samples, but in 6 of the 118 breast cancer samples, four of the coding sequences appeared to change[23]. These changes were located in the functional domain of the protein and only exist in tumor tissues. Therefore, these mutations may affect the function and/or cellular localization of the protein, which may promote the development of cancer. In addition, *DOK1* mutations have also been found in a small number of colorectal cancer cases[38]. However, no obvious mutation in *DOK2* were found in leukemia, colorectal cancers, and gastric cancer [27, 39]. Meanwhile, mutations in *DOK7* have mainly been shown to be related to congenital myasthenic syndrome, which has not been studied in

tumors[40]. Our study found that the expression of *DOK* family genes is not only related to tumor mutations, but also related to CNV, SNV, and promoter methylation. These changes may be one of the important mechanisms leading to the disturbance of *DOK* family genes expression.

Cancer stem cells are a small population of tumor cells which present with stem cell characteristics. They maintain self-renewal and multidirectional differentiation potential through asymmetric division, leading to continuous proliferation of tumor cells and promoting the production of tumor heterogeneity and diversity. These cells have strong tumorigenesis and invasive ability, and may be the root cause of tumor resistance, recurrence after treatment, and tumor metastasis[41, 42]. Previous studies have found that knocking out *DOK1* in epithelial ovarian cancer (EOC) cells can increase their sensitivity to cisplatin[24]. Therefore, it is necessary to further evaluate the relationship between *DOK* family genes, cancer stem cells, and chemotherapeutic drug sensitivity. Our research found that the expression of *DOK5* and *DOK6* was the lowest in tumor tissues and that *DOK5* and *DOK6* expression was negatively correlated with RNAs and DNAs ( $P < 0.05$ ), with the relationship between *DOK6* and RNAs being the most obvious ( $r = -0.73$ ). *DOK5* is not linked to chemotherapy drug sensitivity but *DOK6* is closely related to the sensitivity of multiple chemotherapy drugs, including ethinyl estradiol, estramustine, etoposide, carmustine, and 7-hydroxystaurosporine. *DOK2*, *DOK3*, and *DOK4* are negatively correlated with RNAs, but positively or negatively correlated with DNAs in different tumors. Increased expression of *DOK2* may increase tumor sensitivity to many chemotherapy drugs, including chelerythrine, hydroxyurea, fenretinide, idarubicin, uracil mustard, and chlorambucil. Previous studies have confirmed that in ovarian cancer, *DOK2* deletion increases carboplatin resistance by reducing apoptosis[43, 44]. The increase in *DOK4* expression reduces the sensitivity of cancer cells to multiple chemotherapy drugs. However, increased expression of *DOK4* is related to resistance to several drugs, including carfilzomib, depsipeptide, pipamperone, ribulin mesilate, vinblastine, and actinomycin D. Therefore, *DOK4* may play a role in drug resistance in some tumors and affect the survival of patients, but given that this is only a database analysis these hypotheses will need to be

evaluated in *in vivo* and *in vitro* studies.

The tumor microenvironment is composed of tumor cells and infiltrating immune cells around the tumor, cancer-associated fibroblasts (CAFs), extracellular matrix, new blood vessels, and endothelial cells[45]. There are many factors in this microenvironment that facilitate the interaction between tumor tissues and the immune system, including immune cells which may identify and eliminate tumor cells inhibiting disease progression. Tumor cells express a variety of inhibitory molecules on their cell surface and secrete tumor-related cytokines to weaken the anti-tumor activity of these immune cells, thereby mediating the body's immune tolerance to tumors and enabling immune escape. CD4 + and CD8 + T cells are important members of TME and participate in specific anti-tumor immune responses. DCs are the most important antigen-presenting cells in the body, and can initiate immunity and determine the final development of the immune response channeling the environment towards immune activation or tolerance[46]. Neutrophils secrete MMP9 into TME, which contributes to angiogenesis, tumor progression and metastasis in mouse transplantation models. The degree of enrichment of mature dendritic cells and CD8+T cells is closely related to the survival rate of tumor patients. The more DC and CD8+T cells, the higher the survival rate. However, our analysis suggests that *DOK1-DOK3* expression exhibits a significant correlation with the immune score, with *DOK2* having the strongest correlation. *DOK4-DOK6* were both positively and negatively correlated with immune score in different tumors, while *DOK7* was shown to have a low degree of correlation with immune score in the microenvironment. In addition, we found that *DOK1-DOK3* have a significant, positive correlation with B cell, CD4 + T cell, CD8 + T cell, neutrophil, macrophage, and dendritic cell (DCs) infiltration. Among them, the correlation was strongest between the DOK and DCs. *DOK1-DOK3* may affect the development of tumors by regulating DCs and CD8+T cells in some tumors[47]. Therefore, *DOK1-DOK3* may play an important role in tumor immunity.

Studies have found that *DOK2* and *DOK5* expression is linked to changes in the tumor microenvironment[48, 49], but the correlation between other members of the *DOK* family gene and the tumor microenvironment have not been studied.

Tumor-associated macrophages (TAMs) occupy a large proportion of the immune cell population in the tumor microenvironment. The classical typology of TAMs can be divided into M1-type and M2-type macrophages. In most cases, the M1-type is thought to play a predominantly anti-tumor role, while the M2-type plays a role in promoting tumor progression[50]. Recent studies have found that *DOK1* activates NF- $\kappa$ B conduction in macrophages and inhibits PD-L1 expression, thereby affecting the prognosis of gastric cancer patients[51]. Our research also found that there is a significant correlation between *DOK1-DOK3* and *DOK5* expression and macrophage activation. In addition, we also found that *DOK1-DOK3* has a significant correlation with immune checkpoints. Our research further clarified that *DOK1-DOK3* has broader tumor applicability, and confirmed that the expression of *DOK1-DOK3* is closely related to the biological processes of immune cells and immune-related molecules in most cancers.

In summary, we found massive heterogeneity in *DOK* expression levels in different types of tumors. Among all the *DOK* family members, *DOK4* showed the most abundant expression, and the highest degree of heterogeneity among the various tumors analyzed and was the most obvious marker for prognostic evaluations. Each gene in the *DOK* family members demonstrated a different expressional profile in each of the 18 cancers evaluated with most genes exhibiting some degree of upregulation in GBM, KIRC, CHOL, THCA, KIRP, LIHC and STAD, and downregulation in LUSC and LUAD, but there were some exceptions. We went on to analyzed the relationship between expression *DOK* family genes and the overall survival rate of patients in 30 cancer types. The relationship between the expression level of *DOK* family genes and the prognosis depended on the type of cancer. However, in general, *DOK* family genes were upregulated in GBM, KIRC, CHOL, THCA, KIRP, LIHC, and STAD, and linked to poorer prognosis, while they mainly downregulated in LUSC and LUAD, and linked to improved survival rate. We then used the ESTIMATE algorithm to evaluate the links between *DOK* family genes expression and stromal and immune cell infiltration. The correlation between *DOK* family genes expression and tumor stemness score and drug sensitivity score indicated that *DOK4* reduces tumor the sensitivity for multiple

chemotherapy drugs in tumors, while *DOK2*, *DOK6*, and *DOK7* increased tumor sensitivity to multiple chemotherapy drugs. Although we conducted a comprehensive analysis of *DOK* family genes in human tumors, there are certain limitations. First, our data came from the TCGA database, which could not guarantee the quality of the samples included. Second, our current research is based on bioinformatic analysis, and many results have not been reported. Therefore, further clinical and laboratory studies are in urgently needed to verify our findings.

## Conclusion

In summary, our research revealed the *DOK* family genes is not entirely a tumor suppressor gene, but also a tumor-promoting gene in some tumors. *DOK* family genes are significantly associated with poor prognosis of UVM. *DOK1-DOK3* showed a significant correlation with tumor immunity. *DOK2* and *DOK6* were negatively correlated with tumor stemness and could increase the sensitivity of cancer cells to chemotherapy drugs, while *DOK4* reduced the sensitivity of cancer cells to multiple chemotherapy drugs.

## Conflict of interest

The authors declare that they have no competing interests.

## Data availability

The following information was supplied regarding data availability: TCGA raw data is available at UCSC xena browser(<https://xenabrowser.net/datapages/>). NCI-60 cell line data is available at CellMiner (<https://discover.nci.nih.gov/cellminer/home.do>).

## References

- [1] J.A. Janas, L. Van Aelst, Oncogenic tyrosine kinases target Dok-1 for ubiquitin-mediated proteasomal degradation to promote cell transformation, *Mol Cell Biol* 31(13) (2011) 2552-2565.

- [2] R. Mashima, S. Arimura, S. Kajikawa, H. Oda, S. Nakae, Y. Yamanashi, Dok adaptors play anti-inflammatory roles in pulmonary homeostasis, *Genes Cells* 18(1) (2013) 56-65.
- [3] N. Carpino, D. Wisniewski, A. Strife, D. Marshak, R. Kobayashi, B. Stillman, et al. p62(dok): a constitutively tyrosine-phosphorylated, GAP-associated protein in chronic myelogenous leukemia progenitor cells, *Cell* 88(2) (1997) 197-204.
- [4] Y. Yamanashi, D. Baltimore, Identification of the Abl- and rasGAP-associated 62 kDa protein as a docking protein, *Dok, Cell* 88(2) (1997) 205-211.
- [5] R. Mashima, Y. Hishida, T. Tezuka, Y. Yamanashi, The roles of Dok family adapters in immunoreceptor signaling, *Immunol Rev* 232(1) (2009) 273-285.
- [6] G. Guittard, P. Pontarotti, S. Granjeaud, M. Rodrigues, L. Abi-Rached, J.A. Nunès, 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* 84 (2018) 193-198.
- [7] J. Grimm, M. Sachs, S. Britsch, S. Di Cesare, T. Schwarz-Romond, K. Alitalo, et al. Novel p62dok family members, dok-4 and dok-5, are substrates of the c-Ret receptor tyrosine kinase and mediate neuronal differentiation, *J Cell Biol* 154(2) (2001) 345-354.
- [8] C. Favre, A. Gérard, E. Clauzier, P. Pontarotti, D. Olive, J.A. Nunès, DOK4 and DOK5: new Dok-related genes expressed in human T cells, *Genes Immun* 4(1) (2003) 40-45.
- [9] P.-F. He, Z.-J. Xu, J.-D. Zhou, X.-X. Li, W. Zhang, D.-H. Wu et al. Methylation-associated DOK1 and DOK2 down-regulation: Potential biomarkers for predicting adverse prognosis in acute myeloid leukemia, *J Cell Physiol* 233(9) (2018) 6604-6614.
- [10] M. Chen, J. Zhang, A.H. Berger, M.S. DiIombi, C. Ng, J. et al. Bronson, M. Castillo-Martin,

- T.H. Thin, C. Cordon-Cardo, R. Plevin, P.P. Pandolfi, Compound haploinsufficiency of Dok2 and Dusp4 promotes lung tumorigenesis, *J Clin Invest* 129(1) (2019) 215-222.
- [11] C.H. An, M.S. Kim, N.J. Yoo, S.H. Lee, Mutational and expressional analysis of a haploinsufficient tumor suppressor gene DOK2 in gastric and colorectal cancers, *APMIS* 119(8) (2011) 562-564.
- [12] T. Friedrich, M. Söhn, T. Gutting, K.-P. Janssen, H.-M. Behrens, C. Röcken, M.P.A. et al. Burgermeister, Subcellular compartmentalization of docking protein-1 contributes to progression in colorectal cancer, *EBioMedicine* 8 (2016) 159-172.
- [13] T. Ghanem, J. Bracken, A. Kasem, W.G. Jiang, K. Mokbel, mRNA expression of DOK1-6 in human breast cancer, *World J Clin Oncol* 5(2) (2014) 156-163.
- [14] A. Heidarizadi, M. Salimi, H. Mozdarani, Study of DOK4 gene expression and promoter methylation in sporadic breast cancer, *Neoplasma* 67(4) (2020) 916-921.
- [15] P. Wu, Z.J. Heins, J.T. Muller, L. Katsnelson, I. de Bruijn, A.A. Abeshouse et al. Integration and Analysis of CPTAC Proteomics Data in the Context of Cancer Genomics in the cBioPortal, *Mol Cell Proteomics* 18(9) (2019) 1893-1898.
- [16] T. Li, J. Fan, B. Wang, N. Traugh, Q. Chen, J.S. Liu et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells, *Cancer Res* 77(21) (2017) e108-e110.
- [17] A. Gérard, M. Ghiotto, C. Fos, G. Guittard, D. Compagno, A. Galy et al. Dok-4 is a novel negative regulator of T cell activation, *J Immunol* 182(12) (2009) 7681-7689.
- [18] J. Huang, X. Peng, K. Zhang, C. Li, B. Su, Y. Zhang et al. Co-expression and significance of Dok2 and Ras p21 protein activator 1 in breast cancer, *Oncol Lett* 14(5) (2017) 5386-

5392.

- [19] F. Fang, H. Cardenas, H. Huang, G. Jiang, S.M. Perkins, C. Zhang, et al. Genomic and Epigenomic Signatures in Ovarian Cancer Associated with Resensitization to Platinum Drugs, *Cancer Res* 78(3) (2018) 631-644.
- [20] R.P. Deshpande, Y.B.V.K. Chandra Sekhar, M. Panigrahi, P.P. Babu, Region-Specific Dok2 Overexpression Associates with Poor Prognosis in Human Astrocytoma, *Mol Neurobiol* 55(1) (2018) 402-408.
- [21] E.J. Downer, D.G. Johnston, M.A. Lynch, Differential role of Dok1 and Dok2 in TLR2-induced inflammatory signaling in glia, *Mol Cell Neurosci* 56 (2013) 148-58.
- [22] S. Oki, A. Limnander, P.M. Yao, M. Niki, P.P. Pandolfi, P.B. Rothman, Dok1 and SHIP act as negative regulators of v-Abl-induced pre-B cell transformation, proliferation and Ras/Erk activation, *Cell Cycle* 4(2) (2005) 310-4.
- [23] E. Tuna, Y.E. Ersoy, P. Bulut, F. Ozdemir, N. Buyru, Analysis of the DOK1 gene in breast cancer, *Mol Biol Rep* 47(3) (2020) 1605-1612.
- [24] P.L. Mercier, M. Bachvarova, M. Plante, J. Gregoire, M.C. Renaud, K. Ghani et al. Characterization of DOK1, a candidate tumor suppressor gene, in epithelial ovarian cancer, *Mol Oncol* 5(5) (2011) 438-53.
- [25] A. Saulnier, T. Vaissière, J. Yue, M. Siouda, M. Malfroy, R. Accard et al. Inactivation of the putative suppressor gene DOK1 by promoter hypermethylation in primary human cancers, *Int J Cancer* 130(11) (2012) 2484-94.
- [26] A.H. Berger, M. Niki, A. Morotti, B.S. Taylor, N.D. Socci, A. Viale, et al. Identification of DOK genes as lung tumor suppressors, *Nat Genet* 42(3) (2010) 216-23.

- [27] C.H. An, M.S. Kim, N.J. Yoo, S.H. Lee, Mutational and expressional analysis of a haploinsufficient tumor suppressor gene DOK2 in gastric and colorectal cancers, *Apmis* 119(8) (2011) 562-4.
- [28] F. Carneiro, C. Oliveira, M. Leite, R. Seruca, Molecular targets and biological modifiers in gastric cancer, *Semin Diagn Pathol* 25(4) (2008) 274-87.
- [29] M. Niki, A. Di Cristofano, M. Zhao, H. Honda, H. Hirai, L. Van Aelst, et al. Role of Dok-1 and Dok-2 in leukemia suppression, *J Exp Med* 200(12) (2004) 1689-95.
- [30] T. Ghanem, J. Bracken, A. Kasem, W.G. Jiang, K. Mokbel, mRNA expression of DOK1-6 in human breast cancer, *World J Clin Oncol* 5(2) (2014) 156-63.
- [31] L. Zhang, R. Li, K. Hu, Y. Dai, Y. Pang, Y. Jiao et al. Prognostic role of DOK family adapters in acute myeloid leukemia, *Cancer Gene Ther* 26(9-10) (2019) 305-312.
- [32] R. Ueta, S. Sugita, Y. Minegishi, A. Shimotoyodome, N. Ota, N. Ogiso, et al. DOK7 Gene Therapy Enhances Neuromuscular Junction Innervation and Motor Function in Aged Mice, *iScience* 23(8) (2020) 101385.
- [33] V. Bissay, R.A. Maselli, Phenotypic Differences in 2 Unrelated Cases Carrying Identical DOK7 Mutations, *J Clin Neuromuscul Dis* 21(1) (2019) 30-34.
- [34] J. Finsterer, Congenital myasthenic syndromes, *Orphanet J Rare Dis* 14(1) (2019) 57.
- [35] G. Chen, H. Yu, L. Satherley, C. Zabkiewicz, J. Resaul, H. Zhao, et al. The downstream of tyrosine kinase 7 is reduced in lung cancer and is associated with poor survival of patients with lung cancer, *Oncol Rep* 37(5) (2017) 2695-2701.
- [36] H. Heyn, F.J. Carmona, A. Gomez, H.J. Ferreira, J.T. Bell, S. Sayols et al. DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel

- epigenetic biomarker, *Carcinogenesis* 34(1) (2013) 102-8.
- [37] M. Siouda, J. Yue, R. Shukla, S. Guillemier, Z. Herceg, M. Creveaux, R. Accardi, et al. Transcriptional regulation of the human tumor suppressor DOK1 by E2F1, *Mol Cell Biol* 32(23) (2012) 4877-90.
- [38] E.J. Choi, J.H. Lee, M.S. Kim, S.Y. Song, N.J. Yoo, S.H. Lee, Intratumoral Heterogeneity of Somatic Mutations for NRIP1, DOK1, ULK1, ULK2, DLGAP3, PARD3 and PRKCI in Colon Cancers, *Pathol Oncol Res* 24(4) (2018) 827-832.
- [39] M.S. Kim, N.G. Chung, N.J. Yoo, S.H. Lee, Mutational analysis of DOK2 tumor suppressor gene in acute leukemias, *Leuk Res* 35(6) (2011) e87-8.
- [40] J.A. Bevilacqua, M. Lara, J. Díaz, M. Campero, J. Vázquez, R.A. Maselli, Congenital Myasthenic Syndrome due to DOK7 mutations in a family from Chile, *Eur J Transl Myol* 27(3) (2017) 6832.
- [41] X. Li, M.T. Lewis, J. Huang, C. Gutierrez, C.K. Osborne, M.F. Wu, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy, *J Natl Cancer Inst* 100(9) (2008) 672-9.
- [42] C.J. Creighton, X. Li, M. Landis, J.M. Dixon, V.M. Neumeister, A. Sjolund, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features, *Proc Natl Acad Sci U S A* 106(33) (2009) 13820-5.
- [43] E. Lum, M. Vigliotti, N. Banerjee, N. Cutter, K.O. Wrzeszczynski, S. Khan, et al. Loss of DOK2 induces carboplatin resistance in ovarian cancer via suppression of apoptosis, *Gynecol Oncol* 130(2) (2013) 369-76.
- [44] F. Fang, H. Cardenas, H. Huang, G. Jiang, S.M. Perkins, C. Zhang, H.N. Keer, et al. Genomic and Epigenomic Signatures in Ovarian Cancer Associated with

- Resensitization to Platinum Drugs, *Cancer Res* 78(3) (2018) 631-644.
- [45] F. Klemm, J.A. Joyce, Microenvironmental regulation of therapeutic response in cancer, *Trends Cell Biol* 25(4) (2015) 198-213.
- [46] R.M. Steinman, M.C. Nussenzweig, Avoiding horror autotoxicus: the importance of dendritic cells in peripheral T cell tolerance, *Proc Natl Acad Sci U S A* 99(1) (2002) 351-8.
- [47] J.D. Phillips, L.M. Knab, N.R. Blatner, L. Haghi, M.M. DeCamp, S.L. Meyerson et al. Preferential expansion of pro-inflammatory Tregs in human non-small cell lung cancer, *Cancer Immunol Immunother* 64(9) (2015) 1185-91.
- [48] L. Pan, J. Fang, M.Y. Chen, S.T. Zhai, B. Zhang, Z.Y. Jiang et al. Promising key genes associated with tumor microenvironments and prognosis of hepatocellular carcinoma, *World J Gastroenterol* 26(8) (2020) 789-803.
- [49] Z.Y. Xu, M. Zhao, W. Chen, K. Li, F. Qin, W.W. Xiang, et al. Analysis of prognostic genes in the tumor microenvironment of lung adenocarcinoma, *PeerJ* 8 (2020) e9530.
- [50] A.M. Murri, M. Hilmy, J. Bell, C. Wilson, A.M. McNicol, A. Lannigan, et al. The relationship between the systemic inflammatory response, tumour proliferative activity, T-lymphocytic and macrophage infiltration, microvessel density and survival in patients with primary operable breast cancer, *Br J Cancer* 99(7) (2008) 1013-9.
- [51] T. Li, B. Li, A. Sara, C. Ay, W.Y. Leung, Y. Zhang, et al. Docking protein-1 promotes inflammatory macrophage signaling in gastric cancer, *Oncoimmunology* 8(11) (2019) e1649961.

### Figure Legend

**Figure 1** The expression level of DOK family gene in cancer tissues and adjacent normal tissues. (A) The box plot shows the distribution of DOK family gene expression

in all 33 cancer types. (B) The expression level of DOK family gene in cancer obtained from the GCSALite. (C) The heat map shows the difference in DOK expression between the primary tumor and adjacent normal tissues based on  $\log_2(\text{fold change})$ , for more than 5 adjacent normal samples of 18 cancer types. (D-J) Expression levels of DOK family gene in cancerous and adjacent normal tissues for all 18 cancer types. Boxplots represent the distribution of the DOK expression levels in primary tumor and normal tissues (if available) of different cancer types for each of the DOK. The band inside the box is the median expression values for the gene. (K) Association of DOK family gene expression with patient overall survival for different cancer types. The forest plots with the hazard ratios and 95% confidence intervals for overall survival for different cancer types to show survival advantage and disadvantage with increased gene expression of DOK family gene. Univariate Cox proportional hazard regression models were used for the association tests. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

**Figure 2** Immunohistochemical expression of DOK1, DOK2, DOK4, and DOK6 in human lung cancer, liver cancer, and colorectal cancer. (A-C) The expression of DOK1, DOK2, DOK4, and DOK6 by IHC staining (Magnification  $\times 200$ ). (D-F) Compare the MOD of all groups DOK1, DOK2, and DOK6. (G) The expression of DOK1 is correlated with the prognosis of lung cancer. (H) the expression of DOK2 is correlated with the prognosis of lung cancer. (I) the expression of DOK6 is correlated with the prognosis of lung cancer. (J) the expression of DOK2 is correlated with the prognosis of liver cancer. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

**Figure 3** The alteration frequency of DOK family gene in different cancers, and the relationship between DOK family gene expression. (A-C) The alteration frequency of DOK family gene in different cancers obtained from the cBioPortal. (D) Mutation landscape of DOK family gene in cancer cell lines obtained from the CCLE. (E) The relationship between the Copy Number Variation (CNV) of DOK family genes and the expression level in different cancers. (F) SNV frequency of DOK family gene in different cancers. (G) The relationship between promoter methylation and expression

levels of DOK family gene in different cancers. (H) The difference in promoter methylation of DOK family gene in different cancers.

**Figure 4** The correlation of DOK family gene expression with tumor microenvironment and immune infiltration. (A-C) The correlation matrix plots to show the correlation between DOK family gene expression based on the ESTIMATE algorithm and the immune score, estimated score, and stromal score of 33 different cancer types. Spearman correlation is used for testing. The size of the dot represents the absolute value of the correlation coefficients. The larger the size, the higher the correlation (the higher the absolute correlation coefficient). This also applies to Figure 6A-B. (D-F) The correlation between DOK1, DOK2, and DOK3 expression and immune infiltration levels in 32 cancers of TIMER. Red, positive correlation ( $P < 0.05$ ); Green, negative correlation ( $P < 0.05$ ); Gray, not significant ( $P > 0.05$ ).

**Figure 5** Correlation between DOK family gene mRNA expression levels and acknowledged immune checkpoints' mRNA expression in multiple tumors from TCGA database. The lower triangle in each tile indicates coefficients calculated by Pearson's correlation test, and the upper triangle indicates log<sub>10</sub> transformed  $P$ -value. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\*  $P < 0.001$ .

**Figure 6** Association of DOK family gene expression with tumor stemness, drug sensitivity, and signaling pathways. (A-B) Based on the Spearman correlation test, the correlation matrix between DOK family gene expression and cancer severity is RNAss (A) and DNAss (B). (C) Scatter plots to show the association between DOK family gene expression and drug sensitivity (Z-score from CellMiner interface) tested with Pearson Correlation using NCI-60 cell line data. (D) DOK family gene are associated with the activation and inhibition of 10 cancer pathways. The pie chart shows the correlation between DOK family gene and cancer pathways. Solid line indicates activation and dashed line indicates inhibition. (E) Correlation plot based on Spearman correlation test results, showing the correlation of gene expression among DOK family

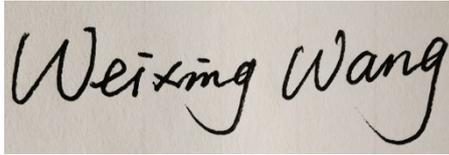
gene in all 33 cancer types. (F) Correlation analysis between DOK2 and DOK3 in GEPIA. (G) Correlation analysis between DOK1 and DOK2 in GEPIA. (H) Correlation analysis between DOK1 and DOK3 in GEPIA.

Journal Pre-proofs

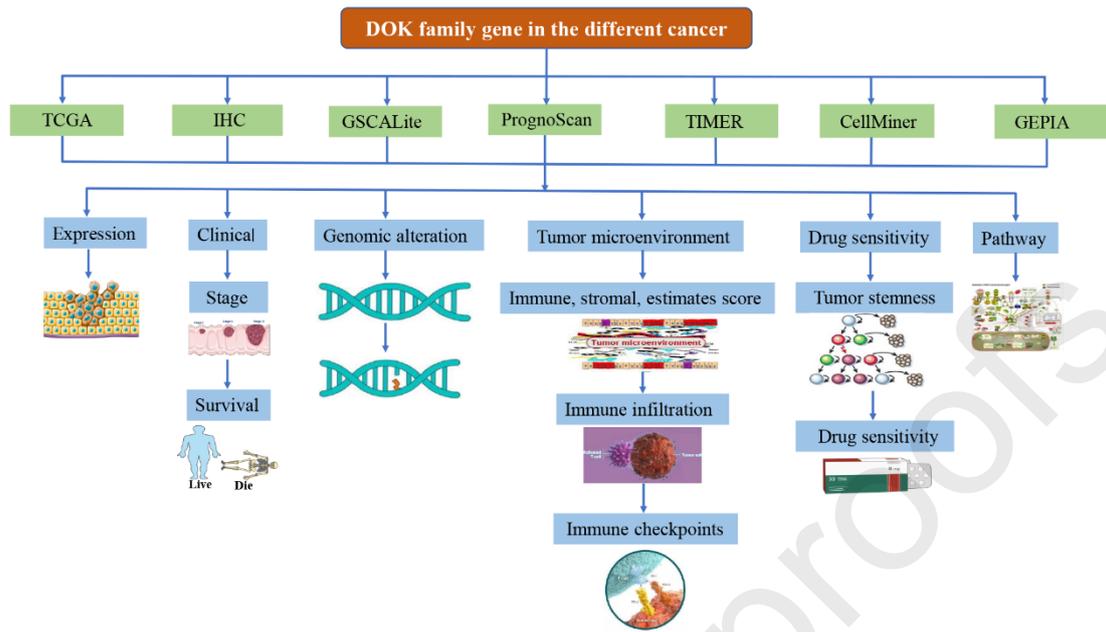
**Credit author statement**

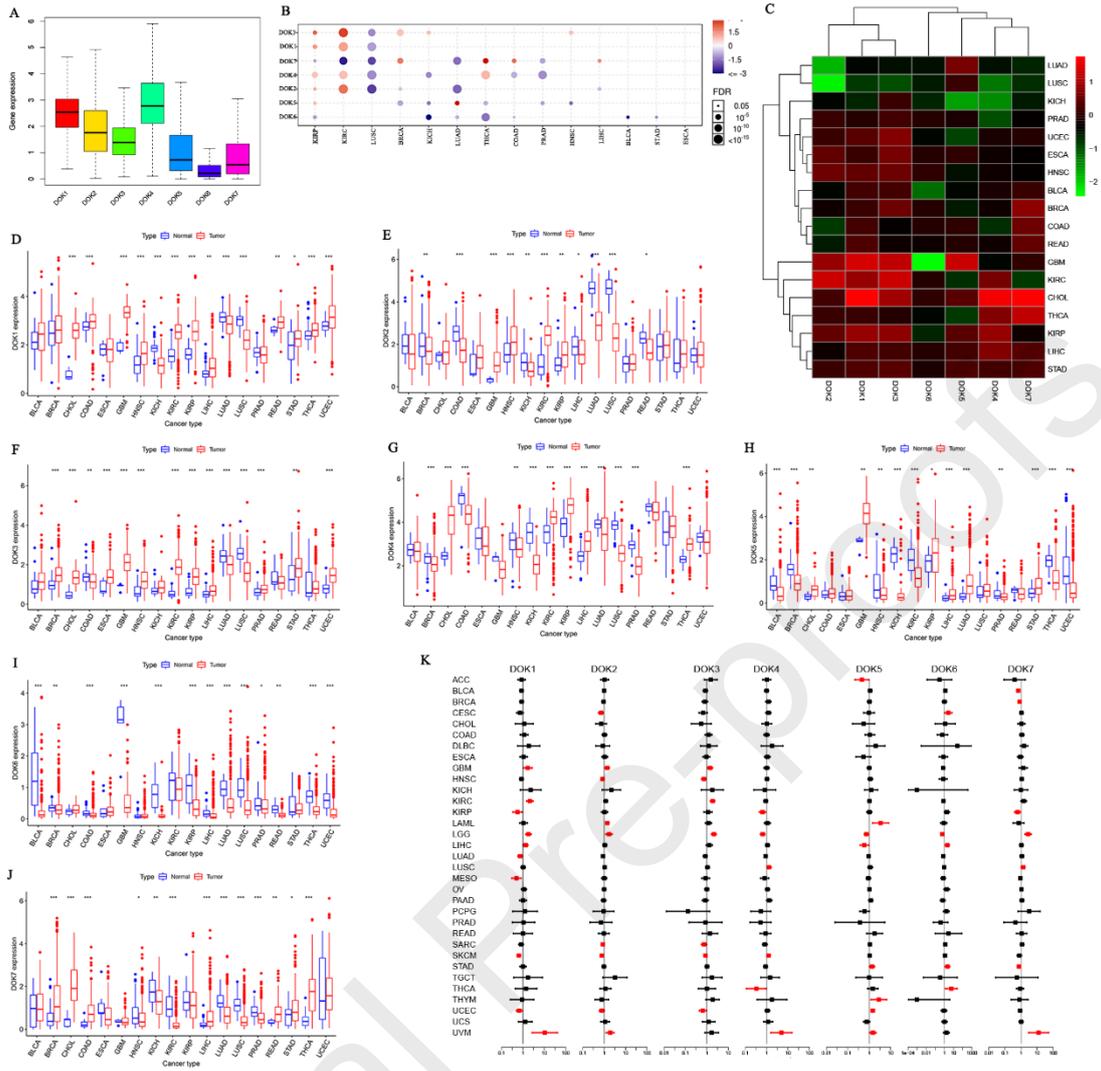
*The authors have declared no conflict of interest*

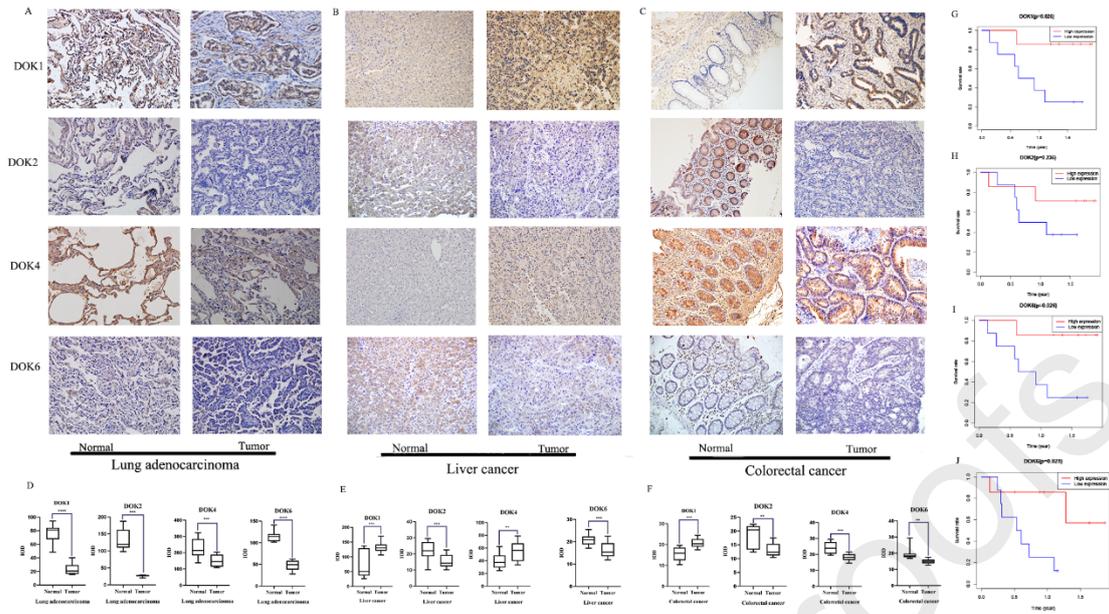
**Corresponding author**

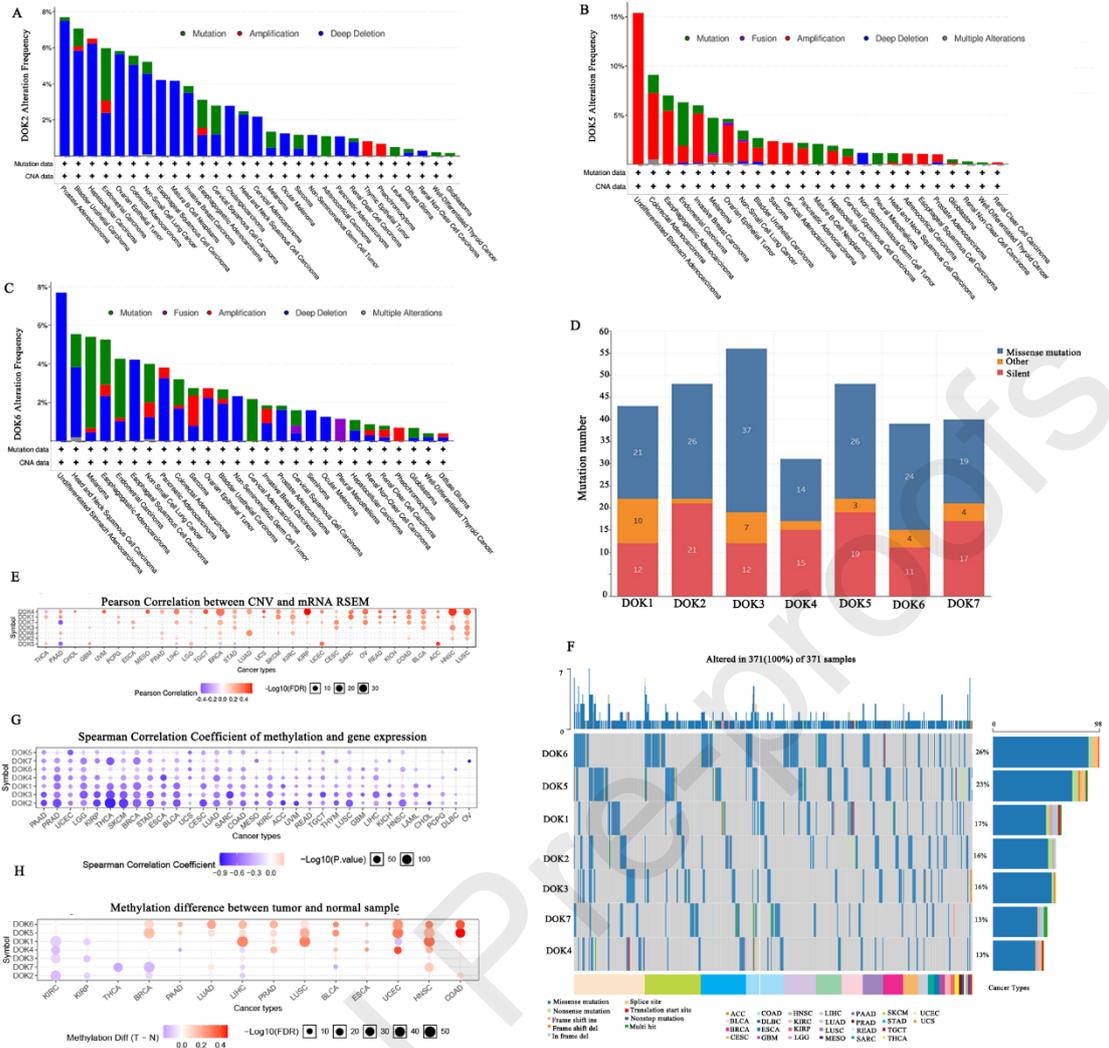
A rectangular image showing a handwritten signature in black ink on a light-colored background. The signature reads "Weixing Wang" in a cursive script.

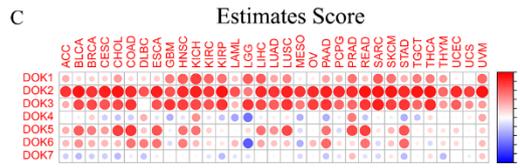
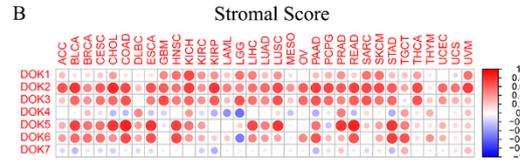
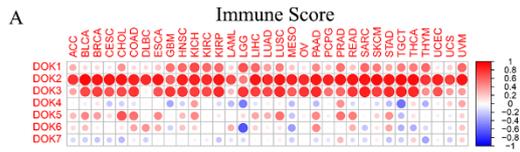
Journal Pre-proofs











**E DOK2**

Cancer	Purity	B cell	CD8+Tcell	CD4+Tcell	Macrophage	Neutrophil	Dendritic cell
ACC	-0.921	0.257	0.361	0.29	0.471	0.569	0.634
BLCA	-0.686	-0.043	0.294	0.42	0.24	0.665	0.682
BRCA	-0.524	0.442	0.437	0.653	0.258	0.641	0.779
CESC	-0.427	0.382	0.612	0.525	0.326	0.664	0.788
CHOL	-0.701	0.492	0.528	0.519	0.549	0.377	0.772
COAD	-0.445	0.237	0.357	0.475	0.554	0.698	0.756
DLBC	-0.682	-0.299	0.439	0.627	-0.16	0.318	-0.179
ESCA	-0.438	0.373	0.049	0.494	0.564	0.378	0.343
GBM	-0.363	-0.121	-0.266	-0.113	-0.007	-0.043	0.33
HNSC	-0.281	0.395	0.654	0.654	0.722	0.724	0.828
KICH	-0.564	0.374	-0.017	0.31	0.269	0.008	0.579
KIRC	-0.388	0.423	0.553	0.42	0.477	0.608	0.735
KIRP	-0.372	0.562	0.617	0.594	0.555	0.527	0.789
LGG	0.046	0.139	0.203	0.326	0.329	0.337	0.387
LIHC	-0.525	0.591	0.67	0.32	0.581	0.465	0.729
LUAD	-0.468	0.401	0.308	0.536	0.559	0.686	0.785
LUSC	-0.465	0.39	0.494	0.519	0.553	0.578	0.815
MESO	-0.381	0.4	0.033	0.37	0.077	0.221	0.662
OV	0.471	0.096	0.373	0.313	0.117	0.399	0.416
PAAD	-0.307	0.333	0.343	0.504	0.532	0.656	0.725
PCPG	-0.437	0.142	-0.072	0.462	0.521	0.301	0.449
PRAD	-0.407	0.486	0.227	0.681	0.359	0.622	0.64
READ	-0.434	0.176	0.024	0.448	0.209	0.194	0.745
SARC	-0.543	0.223	0.381	0.595	0.65	0.518	0.737
SKCM	-0.641	0.3	0.577	0.451	0.533	0.651	0.787
STAD	-0.257	-0.267	0.668	0.345	0.442	0.681	0.806
TGCT	-0.85	0.495	0.166	0.586	0.423	0.703	0.672
THCA	-0.142	0.429	0.51	0.241	0.412	0.721	0.825
THYM	-0.121	0.29	0.051	0.659	0.036	0.129	0.616
UCEC	-0.289	0.663	0.288	0.621	0.438	0.464	0.632
UCS	-0.276	0.144	-0.004	0.527	0.233	0.335	0.7
UVM	0.032	-0.273	0.23	0.067	0.12	-0.291	0.039

**D DOK1**

Cancer	Purity	B cell	CD8+Tcell	CD4+Tcell	Macrophage	Neutrophil	Dendritic cell
ACC	-0.203	0.533	0.147	0.033	0.398	0.283	0.316
BLCA	-0.027	0.125	-0.029	0.173	0.196	0.142	0.166
BRCA	-0.112	-0.01	-0.028	0.193	0.142	0.082	0.082
CESC	-0.043	0.317	0.143	0.165	0.412	0.114	0.035
CHOL	-0.225	0.463	0.27	0.354	0.389	0.277	0.553
COAD	-0.018	-0.031	0	0.304	0.149	0.218	0.235
DLBC	-0.281	0.067	0.245	0.17	-0.406	0.296	0.329
ESCA	0.057	0.367	0.043	0.279	0.344	0.007	-0.219
GBM	-0.172	0.176	-0.198	0.166	0.123	0.111	0.325
HNSC	-0.133	0.468	0.465	0.558	0.691	0.412	0.654
KICH	-0.552	0.438	0.182	0.234	0.479	-0.052	0.684
KIRC	-0.29	0.35	0.346	0.494	0.493	0.515	0.586
KIRP	-0.26	0.247	0.274	0.595	0.348	0.476	0.521
LGG	-0.217	0.421	-0.043	0.647	0.579	0.492	0.587
LIHC	-0.386	0.548	0.525	0.527	0.659	0.468	0.679
LUAD	-0.152	0.311	0.067	0.424	0.367	0.326	0.476
LUSC	-0.149	0.32	0.302	0.369	0.389	0.295	0.53
MESO	-0.024	-0.046	-0.385	0.273	-0.024	0.06	0.055
OV	0.004	0.085	0.077	0.058	0.077	0.11	0.133
PAAD	-0.15	0.337	0.181	0.429	0.273	0.367	0.472
PCPG	-0.254	0.076	0.159	0.347	0.366	0.087	0.325
PRAD	-0.303	0.323	0.118	0.539	0.446	0.395	0.407
READ	-0.127	0.057	-0.062	0.337	0.258	0.119	0.149
SARC	-0.345	0.095	0.073	0.435	0.499	0.306	0.407
SKCM	-0.323	0.217	0.249	0.291	0.408	0.467	0.424
STAD	-0.085	-0.106	0.186	0.288	0.162	0.221	0.29
TGCT	-0.106	0.324	0.093	0.45	0.353	0.406	0.466
THCA	-0.177	0.392	0.1	0.416	0.358	0.516	0.517
THYM	-0.16	0.172	0.13	0.273	0.19	0.304	0.261
UCEC	-0.085	0.071	0.026	0.097	0.205	-0.113	0.019
UCS	-0.027	0.411	0.011	0.152	0.189	0.252	0.109
UVM	0.119	-0.254	0.087	0.149	0.157	-0.274	0.099

**F DOK3**

Cancer	Purity	B cell	CD8+Tcell	CD4+Tcell	Macrophage	Neutrophil	Dendritic cell
ACC	-0.342	0.104	-0.022	0.338	0.088	0.193	0.227
BLCA	-0.546	0.114	0.037	0.495	0.208	0.454	0.41
BRCA	-0.371	0.526	0.308	0.594	0.366	0.681	0.7
CESC	-0.442	0.327	0.315	0.45	0.367	0.456	0.552
CHOL	-0.658	0.458	0.205	0.397	0.54	0.286	0.457
COAD	-0.417	0.227	0.208	0.559	0.51	0.267	0.673
DLBC	-0.123	0.322	-0.196	0.112	0.144	-0.032	-0.067
ESCA	-0.395	0.55	0.023	0.45	0.517	0.25	0.178
GBM	-0.313	0.084	-0.396	-0.061	-0.003	0.184	0.266
HNSC	-0.305	0.43	0.591	0.67	0.724	0.691	0.799
KICH	-0.492	0.44	-0.085	0.314	0.164	0.07	0.576
KIRC	-0.335	0.292	0.275	0.498	0.429	0.566	0.525
KIRP	-0.367	0.564	0.416	0.641	0.402	0.632	0.743
LGG	-0.267	0.619	0.135	0.818	0.736	0.765	0.801
LIHC	-0.51	0.619	0.596	0.532	0.682	0.56	0.747
LUAD	-0.457	0.544	0.182	0.602	0.397	0.591	0.614
LUSC	-0.522	0.475	0.42	0.588	0.492	0.54	0.746
MESO	-0.22	0.368	-0.151	0.322	0.22	0.131	0.395
OV	-0.098	0.015	0.078	0.142	-0.08	0.022	0.043
PAAD	-0.291	0.306	0.205	0.601	0.504	0.834	0.645
PCPG	-0.384	0.012	-0.062	0.496	0.371	0.026	0.45
PRAD	-0.182	0.49	0.238	0.577	0.35	0.482	0.576
READ	-0.436	0.154	-0.044	0.488	0.117	0.225	0.632
SARC	-0.588	0.299	0.306	0.592	0.666	0.611	0.715
SKCM	-0.557	0.309	0.344	0.352	0.357	0.448	0.592
STAD	0.28	-0.09	0.479	0.435	0.414	0.656	0.751
TGCT	-0.657	0.567	0.327	0.459	0.239	0.587	0.573
THCA	-0.104	0.397	0.494	0.224	0.377	0.659	0.76
THYM	-0.082	-0.175	-0.2	-0.252	-0.209	0.345	-0.256
UCEC	-0.199	0.597	0.184	0.494	0.426	0.411	0.476
UCS	-0.318	-0.045	-0.125	0.339	-0.044	0.036	0.529
UVM	-0.063	-0.283	0.114	-0.042	-0.117	-0.242	0.239



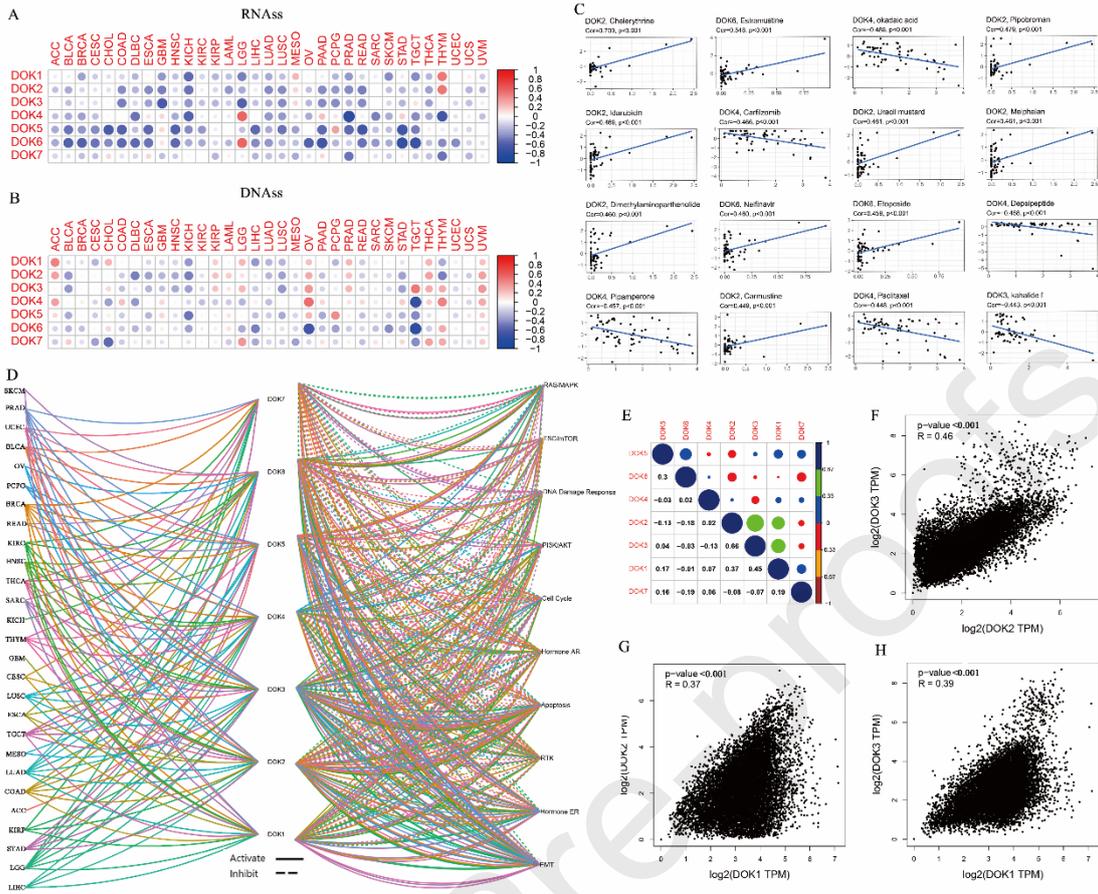


Table 1 Cancer types included in our study from TCGA database

TCGA ID	Primary disease types	Total	Tumor	Normal
ACC	Adrenocortical carcinoma	79	79	0
BLCA	Bladder urothelial carcinoma	430	411	19
BRCA	Breast invasive carcinoma	1217	1097	120
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	309	306	3
CHOL	Cholangiocarcinoma	45	36	9
COAD	Colon adenocarcinoma	512	471	41
DLBC	Lymphoid neoplasm diffuse large B-cell lymphoma	48	48	0
ESCA	Esophageal carcinoma	173	162	11
GBM	Glioblastoma multiforme	173	168	5
HNSC	Head and Neck squamous cell carcinoma	546	502	44
KICH	Kidney chromophobe	89	65	24
KIRC	Kidney renal clear cell carcinoma	607	535	72
KIRP	Kidney renal papillary cell carcinoma	321	289	32
LAML	Acute Myeloid Leukemia	152	152	0
LGG	Brain lower grade glioma	529	529	0
LIHC	Liver hepatocellular carcinoma	424	374	50
LUAD	Lung adenocarcinoma	585	526	59
LUSC	Lung squamous cell carcinoma	550	501	49
MESO	Mesothelioma	86	86	0
OV	Ovarian serous cystadenocarcinoma	379	379	0
PAAD	Pancreatic adenocarcinoma	182	178	4
PCPG	Pheochromocytoma and paraganglioma	186	183	3
PRAD	Prostate adenocarcinoma	551	499	52
READ	Rectum adenocarcinoma	177	167	10
SARC	Sarcoma	265	263	2
SKCM	Skin cutaneous melanoma	472	471	1
STAD	Stomach adenocarcinoma	407	375	32
TGCT	Testicular germ cell tumors	156	156	0
THCA	Thyroid carcinoma	568	510	58
THYM	Thymoma	121	119	2
UCEC	Uterine corpus endometrial carcinoma	583	548	35
UCS	Uterine carcinosarcoma	56	56	0
UVM	Uveal melanoma	80	80	0
Total		11058	10321	737

Table 2 DOK1 expression was associated with the prognosis of different cancers in PrognScan.

Data set	Cancer Type	Endpoint	COX p-value	HR	95% CI (low - high)
GSE12417- GPL570	Blood cancer (AML)	OS	0.005908	3.09	1.38-6.91
GSE5122	Blood cancer (AML)	OS	0.049976	1.41	1.00-1.99
MGH- glioma	Brain cancer (Glioma)	OS	0.026259	1.87	1.08-3.24
GSE4412- GPL96	Brain cancer (Glioma)	OS	0.004309	5.61	1.72-18.31
GES12276	Breast cancer	RFS	0.000316	0.70	0.58-0.85
GSE11121	Breast cancer	DMFS	0.000363	0.38	0.23-0.65
GSE1378	Breast cancer	RFS	0.048402	0.71	0.51-1.00
GSE1456- GPL96	Breast cancer	DSS	0.015867	0.57	0.37-0.90
GSE1456- GPL96	Breast cancer	RFS	0.011870	0.30	0.12-0.77
GSE7390	Breast cancer	OS	0.040569	0.73	0.54-0.99
GES11595	Esophagus cancer	OS	0.003069	0.25	0.10-0.62
Jacob- 00182-HLM	Lung cancer (Adenocarcinoma)	OS	0.006034	0.27	0.11-0.69
Jacob- 00182-MSK	Lung cancer (Adenocarcinoma)	OS	0.028547	0.20	0.05-0.84
GSE13213	Lung cancer (Adenocarcinoma)	OS	0.028070	0.69	0.49-0.96
GSE3120	Lung cancer (Adenocarcinoma)	RFS	0.012020	5.07	1.43-18.02
Jacob- 00182-UM	Lung cancer (Adenocarcinoma)	OS	0.032661	0.47	0.23-0.94
GSE9891	Ovarian cancer	OS	0.036597	0.58	0.34-0.97
GSE26712	Ovarian cancer	DFS	0.003239	0.59	0.41-0.84
GSE26712	Ovarian cancer	OS	0.003871	0.57	0.39-0.83
GSE19234	Skin cancer (Melanoma)	OS	0.041078	0.16	0.03-0.93

Annotation: OS, overall survival; RFS, relapse free survival; DSS, disease specific survival; DFS, Disease Free Survival; DMFS, Distant Metastasis Free Survival.

Table 3 The relationship between DOK expression and drug sensitivity

Gene	Drug	correlation	p-value
DOK2	Chelerythrine	0.7085977	2.39E-10
DOK6	Estramustine	0.5482401	5.77E-06
DOK4	okadaic acid	-0.488041	7.64E-05
DOK2	Pipobroman	0.4788433	0.0001087

DOK2	Idarubicin	0.4686688	0.0001589
DOK4	Carfilzomib	-0.466327	0.0001731
DOK2	Uracil mustard	0.4610959	0.0002091
DOK2	Melphalan	0.4607304	0.0002119
DOK2	Dimethylaminoparthenolide	0.4603258	0.000215
DOK6	Nelfinavir	0.4600371	0.0002172
DOK6	Etoposide	0.4585132	0.0002293
DOK4	Depsipeptide	-0.458461	0.0002298
DOK4	Pipamperone	-0.457402	0.0002386
DOK2	Carmustine	0.4488611	0.0003217
DOK4	Paclitaxel	-0.447562	0.0003365
DOK3	kahalide f	-0.44348	0.0003868
DOK2	Fludarabine	0.4415362	0.0004132
DOK2	Thiotepa	0.440709	0.0004249
DOK6	Carmustine	0.4405261	0.0004275
DOK2	Carboplatin	0.4359063	0.0004989
DOK2	Asparaginase	0.4344543	0.0005235
DOK2	Arsenic trioxide	0.4304928	0.0005962
DOK2	PX-316	0.4280999	0.0006445
DOK4	Eribulin mesilate	-0.426006	0.0006896
DOK4	Vinblastine	-0.425417	0.0007028
DOK6	7-Hydroxystaurosporine	0.4248457	0.0007158
DOK4	Actinomycin D	-0.420248	0.0008288
DOK2	Etoposide	0.4176907	0.0008984
DOK2	Ifosfamide	0.4162463	0.00094
DOK2	Triethylenemelamine	0.4146959	0.0009865
DOK1	Gemcitabine	0.4113606	0.0010938
DOK6	Megestrol acetate	0.4074377	0.0012334
DOK4	Vinorelbine	-0.400945	0.0014998
DOK4	Dolastatin 10	-0.394018	0.00184
DOK6	Salinomycin	0.3939486	0.0018438
DOK6	Teniposide	0.3923908	0.0019293
DOK6	AP-26113	0.3909877	0.0020094
DOK2	Parthenolide	0.3900241	0.0020661
DOK6	Ethinyl estradiol	0.3855035	0.0023517
DOK2	Lomustine	0.3831947	0.0025107
DOK6	Arsenic trioxide	0.3824409	0.0025647
DOK6	Epirubicin	0.3809366	0.0026755
DOK2	Nitrogen mustard	0.3794911	0.0027859
DOK2	Digoxin	0.3773697	0.0029553
DOK2	Cytarabine	0.3771083	0.0029768
DOK7	Pipamperone	-0.375646	0.0030997
DOK6	Isotretinoin	0.3743589	0.0032115
DOK2	XK-469	0.3682948	0.0037879

DOK7	BN-2629	-0.365256	0.0041097
DOK2	Cisplatin	0.3616868	0.0045185
DOK6	Irofulven	-0.359741	0.0047561
DOK7	Midostaurin	-0.359726	0.004758
DOK1	Triapine	0.3589469	0.0048562
DOK7	Epirubicin	-0.358925	0.004859
DOK1	Triethylenemelamine	0.357197	0.0050833
DOK4	Tyrothricin	-0.356618	0.0051605
DOK2	BN-2629	0.356401	0.0051897
DOK3	ABT-199	0.35528	0.0053428
DOK1	5-fluoro deoxy uridine 10mer	0.3537224	0.0055622
DOK3	Carboplatin	0.3534247	0.005605
DOK4	Carmustine	-0.352144	0.0057926
DOK4	Arsenic trioxide	-0.351821	0.0058407
DOK2	Dexrazoxane	0.3498989	0.0061345
DOK1	Asparaginase	0.3494209	0.0062096
DOK6	Ifosfamide	0.3484718	0.006361
DOK1	LMP-400	0.3483937	0.0063736
DOK1	Thiotepa	0.3474983	0.0065197
DOK2	Cladribine	0.3462837	0.0067225
DOK2	Batracylin	0.3459751	0.0067749
DOK7	Bortezomib	-0.344944	0.0069526
DOK1	Cladribine	0.34487	0.0069655
DOK4	Lomustine	-0.342182	0.0074488
DOK7	Carmustine	-0.341285	0.0076165
DOK4	Mithramycin	-0.340635	0.0077401
DOK2	Calusterone	0.3405351	0.0077592
DOK4	Ethinyl estradiol	-0.340444	0.0077768
DOK1	Uracil mustard	0.3401498	0.0078335
DOK2	Teniposide	0.3400573	0.0078514