

# A bioinformatics study of autophagy biomarkers associated with the prognosis of endometrial carcinoma

**Wei Liu**

Shandong University

**Rui Sun**

Shandong University

**Yixuan Bai**

Shandong University

**Yunkai Xie**

Shandong University

**Changzhong Li** (✉ [15168888909@163.com](mailto:15168888909@163.com))

Shandong Provincial Hospital

---

## Research article

**Keywords:** Endometrial cancer, Autophagy, Bioinformatics analysis, Prognosis

**Posted Date:** April 8th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-20960/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Autophagy plays a critical role in endometrial carcinoma (EC), but prognosis studies of differentially expressed autophagy-related genes (DEARGs) in EC are lacking. This study aimed to assess the prognostic value of autophagy-related genes (ARGs) in EC and to identify the potential characteristics of ARGs in predicting patient survival and guiding treatment.

**Methods:** The RNA-Seq data and clinical data were obtained from TCGA databases. The DEARGs were identified by “limma” package in R software. Clusterprofiler was used for functional analysis. Using STRING databases to construct a protein-protein interaction (PPI) network and plug-in MCODE to screen hub modules in Cytoscape. The degrees method of cytoHubba was used to select important hub genes. Co-expression analysis was performed using “limma” package and Metascape for functional analysis. Univariate and multivariate Cox proportional hazards regression analyses were performed to construct a prognostic signature. Kaplan–Meier curve analysis, ROC curve analysis and Gene set enrichment analysis (GSEA) were also performed. Validation was executed by UALCAN, CCLE and HPA databases.

**Results:** In total, 45 DEARGs were identified. Functional analysis showed that DEARGs were strikingly enriched in autophagy pathway. PPI network showed that CASP3 was the most central protein. CASP3 co-expression analysis revealed that co-expressed genes were mostly enriched in cell cycle. We identified a novel prognostic signature consisting of 3 genes (CDKN2A, PTK6 and GRID2). The risk score based on the prognostic signature was able to classify patients into high-risk and low-risk groups with significantly different overall survival. Furthermore, the prognostic signature is an independent prognostic predictor of survival and demonstrates superior prognostic performance compared with the clinicopathologic features for predicting 5-year survival. CDKN2A and PTK6 were further validated in UALCAN. GSEA suggested that the two genes played crucial roles in drug metabolism.

**Conclusion:** Using bioinformatics analyses, we identified DEARGs and determined

CDKN2A, PTK6 and DRID2 are tumor-associated genes and can be utilized as possible biomarkers in EC treatment.

## Background

Endometrial cancer (EC) is the most common gynecologic malignancy and it accounts for 20%-30% of female reproductive system malignancies. Over 300,000 new cases are diagnosed annually, accounting for approximately 8.2% of women's cancer incidence worldwide [1]. An investigation from the American Cancer Society (ACS) reveals that the death rate for endometrial cancer has doubled during the past 20 years [2]. Currently, there are still no validated biological markers to screen EC, early diagnosis is key to improving survival, which at 5 years is less than 20% in advanced disease and over 90% in early-stage disease [1, 3-5]. Therefore, it is urgent to develop prognostic or predictive biomarkers for risk assessment to identify high-risk or low-risk patients and to develop appropriate treatment options for them.

Autophagy is a mechanism that damages, degenerates or aging protein and organelles in cells are transported to the lysosome for digestion and degradation, leading to the basal turnover of cell components and providing energy and macromolecular precursors [6]. A growing body of research suggests that autophagy genes have a large potential to serve as promising markers in the diagnosis, prognosis, and personalized targeted therapies [7-9]. Autophagy mechanisms described in endometrial cancers, including the role of PI3K/AKT/mTOR, AMPK-mTOR, and p53 signaling pathways that trigger or inhibit this process, which may become a potential molecular target in clinical treatment methods [10]. Although a lot has been done during the past years, much remains unknown related to autophagy in endometrial cancer. Therefore, for the early detection and treatment of endometrial cancer, it is necessary to understand the autophagy molecular mechanism of the development of endometrial cancer and the identification of new biomarkers.

In recent years, based on the rapid development of high-throughput sequencing technology and increasingly complete public databases, The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database has collected a large number of clinical, pathological, and biological data from patients with malignancies [11, 12]. Through comprehensive bioinformatic analysis, we can study the mechanism of tumors in-depth and more accurately predict development trends, providing reliable research direction for treatment options [13, 14]. The prognostic evaluation of cancer patients is of great significance in guiding clinical treatment and elucidating the mechanism of tumorigenesis [15]. As far as we know, bioinformatics analysis of autophagy genes for survival prediction in endometrial carcinoma has not been reported. In this work, we identified the differential autophagy gene expression between endometrial cancer samples and matched normal endometrial samples by analyzing the high-throughput mRNA data downloaded from the TCGA database. Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, as well as protein-protein interactions (PPI) network analysis of DEARGs, were performed. Additionally, we evaluated the prognostic value of DEARGs and constructed a prognostic signature that effectively predicted patient survival.

## Materials And Methods

### Data acquisition and process

The RNA-Seq data (HTSeq-FPKM) of UCEC ( Uterine Corpus Endometrial Cancer) and corresponding clinical information were downloaded from the TCGA database (<https://cancergenome.nih.gov/>). A total of 425 samples were enrolled in this study, including 406 endometrial adenomas and adenocarcinomas samples and 19 matched normal samples. Autophagy related genes (ARGs) were obtained from HADb (Human Autophagy Database) (<http://www.autophagy.lu/>), a public repository containing information about the human genes described so far as involved in autophagy. The research design was illustrated with a flow chart (Fig. 1).

### Access to differentially expressed autophagy-related genes (DEARGs)

The differentially expressed autophagy mRNAs between endometrial cancer and normal samples were analyzed by “limma” package in R. The fold changes (FCs) in the expression of individual autophagy mRNA were calculated and differentially expressed autophagy mRNAs with  $\log_2|\text{FC}| > 1.0$  and  $\text{FDR} < 0.05$  were considered to be significant. Heatmap was completed by “pheatmap” package in R. Volcano plots and boxplots were also completed in R.

### **Enrichment analysis of functional categories**

DAVID (<https://david.ncifcrf.gov/>) is a web-based online bioinformatics resource designed to provide researchers with a comprehensive set of functional annotation tools to understand the biological mechanisms associated with a large number of genes/proteins. GO annotations of the screened DEARGs were performed using the DAVID online tool and the results were visualized using “clusterprofiler” package in R [16]. KEGG pathway analysis on DEARGs was performed using “clusterprofiler”.  $P_{\text{adjust}} < 0.05$  was set to be significant.

### **Construction and analysis of PPI network complex**

STRING online database (<http://string-db.org>) was used for analyzing the protein-protein interaction (PPI) of common DEARGs and Cytoscape software (<http://www.cytoscape.org/>) was employed for visualizing PPI network of common DEARGs. Cytoscape MCODE plug-in provided access to select hub modules of the PPI network. MCODEs were extracted when the Node score cut-off was 0.2 and K-core was 2. For genes in the MCODE, we use “clusterprofiler” again for functional enrichment analysis. The degrees method of cytoHubba of the Cytoscape was used to select important hub genes among DEARGs, the top 6 genes with high degrees were filtrated [17].

### **CASP3 co-expression and functional analysis**

Co-expression analysis of CASP3 was evaluated using “limma” package in R. The data obtained were RNA-Seq data from the TCGA database that included 406 endometrial cancer samples. |Pearson correlation coefficient|  $> 0.5$  and  $p\text{-value} > 0.001$  were used to select CASP3 co-expressed genes. Pearson correlation coefficient  $\geq 0.5$  was considered positively correlated and  $\leq -0.5$  was considered negatively correlated with CASP3. GO biological process and pathway enrichment analyses of co-expressed genes were performed using Metascape (<https://metascape.org/gp/index.html>) [18].

### **Construction of a prognostic signature**

A total of 401 UCEC patients with detailed follow-up time were included for subsequent analysis. The details of clinical and pathological characteristics were listed in Table 1. Univariate Cox regression analysis was used to select candidate prognostic DEARGs that were significantly related to overall survival [13]. For the optimal prognostic model, multivariate analysis with Cox proportional hazard was then applied to the candidate genes. Cox proportional hazards regression with a  $p < 0.05$  was set to the risk score of developing EC in each patient. A risk score was calculated based on the expression of gene and coefficient. According to the mean risk score, patients will be divided into low-risk and high-risk

groups. Kaplan-Meier analysis for the difference between the survival curves of high-risk group and low-risk group was compared. To evaluate independent prognostic values of the prognostic signature and the key clinical factors in survival prediction, univariate and multivariate Cox regression and analyses were conducted [19]. Hazard ratios (HRs) and 95% confidence intervals (CIs) were computed by the Cox analysis. Receiver operating characteristic (ROC) curve analysis was also performed to compare the five-year predictive value of the prognostic signature and the key clinical factors. The area under the ROC curve was calculated as a predictive value shown as sensitivity and specificity. Besides, we calculated the correlation between each prognostic genes and clinical factors. All statistical analyses were performed using R/Bioconductor, with the cut-off of  $p < 0.05$ .

### **Validation of prognostic genes**

The prognostic genes were finally validated in UALCAN (<http://ualcan.path.uab.edu/analysis.html>). The CCLE (<https://portals.broadinstitute.org/ccle>) was used to validate the mRNA expression of prognostic genes in a variety of tumors. The HPA (Human Protein Atlas) (<http://www.proteinatlas.org/>) was used to validate protein expression of prognostic genes in a variety of tumors.

### **Gene set enrichment analysis (GSEA)**

EC samples from TCGA were divided into 2 different groups according to the expression level of prognostic genes. Then GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was used to define the biological processes enriched in the gene rank derived from DEARGs between the two groups [20]. The NOM  $p$ -val  $< 0.05$  and normalized enrichment score  $|NES| \geq 1$  were used to sort the significant pathways enriched in each phenotype.

## **Results**

### **Identification of autophagy differential gene in TCGA.**

In the present study, a total of 425 samples were enrolled, including 406 endometrial adenomas and adenocarcinomas samples and 19 matched normal samples. First, the downloaded RNA-Seq data from TCGA is organized into an expression matrix. Then 202 known autophagy-related genes (ARGs) were extracted from the HADb database and the expression matrix of these genes was sorted out by R (Supplementary Table S1-S2). We identified the DEARGs using the “limma” package with  $FDR < 0.05$  and  $|\log_2 \text{fold change (FC)}| \geq 1$ . A total of 45 DEARGs were screened, including 23 up-regulated and 22 down-regulated genes (Table 2). CDKN2A is the most differential gene, which is up-regulated by 4.82 times compared with the normal control group. The detailed information of the differential genes was summarized in Supplementary Table S3. As shown in Fig. 2a, a heatmap was generated by R based on the expression levels of common DEARGs in TCGA. Each column in the heatmap represents a biological sample, and each row represents a gene. The depth of the color indicated the expression levels of genes between cancer samples and normal controls. Volcano plots displayed the distribution of the 45

differentially expressed genes (Fig. 2b). Boxplot displayed the expression level of the 45 differentially expressed genes in endometrial cancer and normal samples (Fig. 2c).

### **Gene ontology and KEGG pathway enrichment analysis**

To gain more biological insight, we performed gene ontology (GO) enrichment analysis using online databases, DAVID. The DEARGs were classified into three functional groups: biological process (BP), molecular function (MF) and cellular component (CC). The most enriched BP functions were autophagy, process utilizing autophagic mechanism, intrinsic apoptotic signaling pathway, neuron death, regulation of autophagy. For MF, ubiquitin protein ligase binding, ubiquitin-like protein ligase binding, chaperone binding and phosphatase binding were the most enriched. In the clusters of CC, autophagosome, membrane raft and nuclear envelope were the most enriched (Fig. 3a, b). KEGG analysis showed that the DEARGs were strikingly enriched in apoptosis, autophagy-animal and shigellosis pathway (Fig. 3c, d).

### **Protein-protein interaction network (PPI) and hub genes analysis**

The PPI network of DEARGs was constructed by using the STRING online database. STRING mapped 45 DEARGs into PPI network containing 45 nodes and 140 edges, average node degree is 6.22 (Fig. 4a). The interaction network was then imported into Cytoscape and MCODE plug-in was used to find hub modules in the network. Three MCODEs were calculated according to Node score cut-off = 0.2 and K-core = 2. Among them, MCODE1 contained 10 nodes and 40 edges, with the highest score (Fig. 4b), MCODE2 contained 5 nodes and 7 edges (Fig. 4c), MCODE3 contained 3 nodes and 3 edges (Fig. 4d).

We performed the functional analysis for the 3 MCODEs. In GO analysis, the DEARGs of MCODE1 were mostly enriched in response to corticosteroid, cellular response to inorganic substance and negative regulation of cyclin-dependent protein serine/threonine kinase activity (Fig. 5a); the DEARGs of MCODE2 in autophagy, process utilizing autophagic mechanism and macroautophagy (Fig. 5b); the DEARGs of MCODE3 in ERBB2 signaling pathway and ERBB signaling pathway (Fig. 5c). KEGG analysis showed that the DEARGs of MCODE1 were mostly enriched in Cellular senescence (Fig. 5d); the DEARGs of MCODE2 in Shigellosis (Fig. 5e); the DEARGs of MCODE3 in ErbB signaling pathway (Fig. 5f).

The interaction network was then imported into Cytoscape to screen the hub genes using the degree method scores (Degrees). The cytoHubba plug-in of the Cytoscape software was used to select genes with the highest degrees. The top 6 genes (CASP3, GAPDH, MYC, MAPK3, FOXO1, CDKN2A) with the highest degrees of connectivity are shown in Fig. 4e. CASP3 being the most important gene.

### **Genes co-expressed with CASP3 and enrichment of functions**

The Pearson correlation coefficients between the expression profiles of the CASP3 and mRNAs were calculated to determine the co-expression genes of the CASP3. In total, 2136 CASP3 co-expressed genes including 2115 positively correlated and 21 negatively correlated genes were selected with  $|\text{Pearson coefficient}| > 0.5$  and  $p\text{-value} > 0.001$  as criteria (Supplementary Table S4). The heatmap shows the top 20 positive correlated genes and the top 20 negative correlated genes (Fig. 6). Then we performed

enrichment analysis of GO biological processes and pathways of co-expressed genes by using Metascape. Pathway and Process Enrichment Analysis shows that co-expressed genes were mostly enriched in cell cycle, cell division and Metabolism of RNA (Fig. 7a-c).

### **Identification of prognostic signature**

To identify prognostic genes capable of distinguishing good survival from poor survival in UCEC patients, univariate Cox proportional hazard regression analysis was carried out on each DEARGs using the expression profile in TCGA. Of the 45 DEARGs, the univariate Cox proportional hazards regression analysis screened out 6 prognostic genes, including DLC1, CDKN2A, PTK6, GRID2, NRG3, BAK1 (Fig. 8a). Multivariate Cox proportional hazards regression analysis was further performed on the 6 genes, which screened CDKN2A, PTK6, GRI2D (Table 3). The risk score for predicting overall survival was calculated as follows: Risk score(patients)=0.202\*expression value of CDKN2+ 0.426 \*expression value of PTK6+1.190\*expression value of GRI2D. According to the risk score, patients were divided into low-risk and high-risk groups, and survival analysis showed that the overall survival (OS) of low-risk patients was longer than high-risk patients in TCGA. (Fig. 8b). The distribution of risk score, survival status, and the expression of 3 genes of each patient were also analyzed (Fig. 8c–e). The expression of 3 genes was significantly higher in the high-risk group than in the low-risk group.

### **Correlation between prognostic signature and other clinicopathologic characteristics**

To analyze the independent prognostic value of various factors in endometrial cancer, we analyzed the relationship between the different clinical parameters (age, grade) and the risk score based on prognostic signature. The univariate and multivariate Cox proportional hazards regression showed that age, grade and risk score based on prognostic signature were all independent prognostic indicator of EC (Fig. 9a, b). The HR of high-risk group versus low-risk group for overall survival were 1.167 ( $p = 0.002$ , 95% CI = 1.058–1.286). The HR of G3 group versus G1-2 group for overall survival were 2.012 ( $p = 0.003$ , 95% CI = 1.274–3.179). The HR of age>65 group versus age≤65 group for overall survival were 1.033 ( $p = 0.021$ , 95% CI = 1.005–1.063) . Analysis of the ROC curve was conducted to compare the sensitivity and specificity of survival prediction. The area under curve (AUC) for each of the prognostic factors was calculated and compared. As shown in Fig. 9c, the AUC of risk score was 0.709 and was greater than age (AUC = 0.640) and grade (AUC = 0.650). These results showed that the risk score based on prognostic signature had a better prognostic performance than other prognostic factors. Then, we analyzed the correlation of the three prognostic genes with the patient's age and found that CDKN2A, PTK6 and GRID2 expression were increased in patients over 65 years of age (Fig. 10a-c); the expressions of GRID2 was higher in G3 group than G1-2 group (Fig. 10d). The analysis also found that the risk scores increased with the patient's age and pathological grade (Fig. 10e, f).

### **Prognostic gene validation**

Using UALCAN, we found that the expression of CDKN2A and PTK6 in tumors was higher than normal samples (Fig. 11a, b), and they were negatively correlated with the overall survival of EC patients (Fig.

11c, d). Besides, both have higher expression levels in different subtypes of EC tissues, such as endometrioid adenocarcinoma, serous carcinoma, and mixed serous and endometrioid adenocarcinoma (Fig. 12a, d). Levels also increased in different stages of EC (Fig. 12b, e). The protein expression of CDKN2A and PTK6 in EC samples of the primary tumor are all higher than normal samples (Fig. 12c, f). We further analyzed the mRNA and protein expression of CDKN2A, PTK6 and GRID2 in a variety of tumor tissues using CCLE and HPA database. The analysis showed that CDKN2A and PTK6 were widely expressed in a variety of tumor tissues, whereas GRID2 expression was negative (Fig. S1).

### **Gene set enrichment analysis (GSEA)**

To identify the potential function of CDKN2A and PTK6 in EC, GSEA was conducted to search the enriched KEGG pathways. For CDKN2A, Metabolism of xenobiotics by cytochrome P450, Cardiac muscle contraction, Drug metabolism cytochrome P450 were enriched in three gene sets (Fig. 13a). For PTK6, Nitrogen metabolism, Nicotinate and nicotinamide metabolism, Glycosphingolipid biosynthesis - lacto and neolacto series were enriched in three gene sets (Fig. 13b).

## **Discussion**

Endometrial carcinoma (EC) is the most common gynecologic malignancy and is an increasing public health concern worldwide [2]. With the continuous emergence of new therapeutic targets, the mining of prognostic biomarkers has become a research hotspot. These biomarkers can be used to predict the risk of patients and guide personalized adjuvant therapy for early high-risk patients [21]. Autophagy genes can play a vital role as biomarkers in many cancers, such as melanoma [22], gastric cancer [23] and non-small cell lung cancer [24], etc. But the prognostic research of differentially expressed autophagy-related genes (DEARGs) in endometrial cancer is lacking. Therefore, it is necessary to understand the autophagy molecular mechanism and identification of new biomarkers for the development of endometrial cancer.

In this study, a total of 45 differentially expressed autophagy-related genes were identified based on the TCGA database, including 23 up-regulated DEARGs and 22 down-regulated DEARGs. Moreover, we predicted the biological functions and enrichment pathways of these genes using bioinformatics methods. GO and KEGG analysis showed that these DEARGs were significantly enriched in autophagic biological pathways, suggesting that dysregulation of autophagy played a key role in the development and progression of EC.

We also performed PPI network analysis and screened out three modules and six key genes (CASP3, GAPDH, MYC, MAPK3, FOXO1, CDKN2A), of which CASP3 was the most important. Co-expression analysis showed that co-expressed genes with CASP3 were mostly enriched in cell cycle, cell division and Metabolism of RNA. This indicates that CASP3 plays a central role in executing apoptosis and thus in the carcinogenic process.

A basic study shows that Caspase-3 (CASP3) plays an important role in promoting colon cancer cell invasion and metastasis and therefore serves as a potential target for colon cancer therapy [25]. Jia Lin et

al. found that CASP3 polymorphisms confer to the lung cancer susceptibility [26]. In human cancer treatment, several studies reported that patients with higher levels of procaspase-3 or active caspase-3 had a worse prognosis than patients with lower levels of procaspase-3 or active caspase-3 [27-30]. In this study, CASP3 as a central gene also showed its abnormal expression in EC development, and this finding can guide future exploration of EC mechanisms.

Next, we distinguished a prognostic signature consisting of three identified DEARGs (CDKN2A, PTK6, GRI2D), and most of them are harmful factors of the EC. The 3-DEARG signature divided the OS of EC into high- and low-risk subgroups. Kaplan–Meier analyses indicated that high-risk subgroups with poor survival. Notably, the univariate and multivariate regression analysis indicated that the risk score system of 3-DEARG signature could be regarded as an independent prognostic model to provide a more accurate assessment of OS in EC. Analysis of the ROC curve shows that the risk score is a better prognostic indicator compared with other clinical characteristics (age, grade) used for risk stratification in EC patients. Therefore, our biomarker may provide a simple and accurate prediction for the prognosis of endometrial cancer.

For clinical correlation analysis, we found that risk score increases with age and pathological grade of the patient and the expression of CDKN2A, PTK6, and GRI2D increases with age. GRI2D expression is higher in the high pathological grade group. High age and histopathological grade predicted poor prognosis for endometrial carcinoma. These results indicated that CDKN2A, PTK6, and GRI2D were highly EC-prognosis-related.

Using the online database, we verified that the expression of CDKN2A and PTK6 in tumors was higher than normal samples, and they were negatively correlated with the overall survival of EC patients. Besides, both have higher expression levels in different subtypes of EC tissues. Their expression levels also increased with EC stage. We further found that CDKN2A and PTK6 were widely expressed in a variety of tumor tissues, whereas GRID2 expression was negative. GSEA enrichment analysis showed that CDKN2A and PTK6 were mostly associated with Metabolism of xenobiotics by cytochrome P450 and Nitrogen metabolism, which may provide a new direction for EC research.

The CDKN2A is a gene type of cyclin-dependent kinase inhibitor, its mutations have been related to breast cancer, melanoma and squamous cell carcinoma [31-33]. Recent findings show that polymorphisms in the CDKN2A gene are associated with EC in postmenopausal women [34]. As for PTK6, an intracellular tyrosine kinase, in tumors overexpressing PTK6, this unusual soluble tyrosine kinase is emerging as a mediator of cancer cell phenotypes, including increased proliferation, survival, and migration [35]. In breast cancer cells, oncogenic PTK6 functions to protect cells from autophagic induced death under anchorage-independent conditions [36]. GRID2, human glutamate receptor delta-2, is a relatively new member of the family of ionotropic glutamate receptors which are the predominant excitatory neurotransmitter receptors in the mammalian brain. Recent research has demonstrated that GRID2 connected to several diseases including ataxia, inherited disease and cancer [37, 38]. In short,

almost all DEARGs in the signature are closely related to cancer. Therefore, there is reason to believe that these three DEARGs can be used in the clinic as a reliable and repeatable prognostic biomarker.

To our knowledge, this is the first landscape study on the prognosis of DEARGs signals in endometrial cancer based on large clinical data sets. However, there are some limitations in this study. For example, our research is actually an analysis based on previous data, lacking in vivo and in vitro experimental studies. Therefore, further investigations are still needed to verify the accuracy for estimating prognoses and to test its clinical utility in patient management.

## **Conclusion**

In conclusion, we comprehensively analyzed autophagy-related genes in endometrial carcinoma using bioinformatics analyses. In total, we identified 45 DEARGs, 3 MCODEs, 6 hub genes and CASP3 co-expressed genes, these findings provide comprehensive information for understanding the autophagy molecular mechanism in endometrial cancer. what's more, we identified a novel prognostic signature consisting of 3 prognostic genes through genome-wide integrated analysis of mRNA expression profiles and clinical data. Prognostic biomarkers identified can be used to strongly predict the survival of UCEC patients.

## **Abbreviations**

|        |   |
|--------|---|
| UCEC   | Uterine Corpus Endometrial Cancer                               |
| TCGA   | The Cancer Genome Atlas   |
| HADb   | Human Autophagy Database  |
| DEARGs | Differentially expressed autophagy-related genes                |
| DAVID  | Database for annotation, visualization and integrated discovery |
| GO     | Gene Ontology   |
| KEGG   | Kyoto Encyclopedia of Genes and Genomes                         |
| STRING | Search Tool for the Retrieval of Interacting Genes Database     |
| PPI    | Protein-protein interaction                                     |
| MCODE  | Molecular complex detection                                     |
| ROC    | Receiver operating characteristic                               |
| AUC    | Area under curve  |
| HR     | Hazard ratio  |
| CI     | Confidence interval   |
| GSEA   | Gene set enrichment analysis                                    |

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

Clinical and pathological information of UCEC patients were obtained from The Cancer Genome Atlas (TCGA) project (<https://cancergenome.nih.gov/>).

### Competing interests

The authors declare that they have no competing interests.

### Funding

None.

## Authors' contributions

WL designed the study and performed the data analysis; RS, YXB and YKX drafted the manuscript. CZL was responsible for the review. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable.

## References

1. Njoku K, Chiasserini D, Whetton AD, Crosbie EJ. Proteomic Biomarkers for the Detection of Endometrial Cancer. *Cancers (Basel)*. 2019, 11(10).
2. Sorosky JI. Endometrial cancer. *Obstet Gynecol*. 2012, 120(2 Pt 1):383-397.
3. Crusz SM, Miller RE. Targeted therapies in gynaecological cancers. *Histopathology*. 2020, 76(1):157-170.
4. Sur V, Arora A. Management of Endometrial Cancer A Review. *Reviews on Recent Clinical Trials*. 2015, 10, 309-316.
5. Lee YC, Lheureux S, Oza AM. Treatment strategies for endometrial cancer: current practice and perspective. *Curr Opin Obstet Gynecol*. 2017, 29(1):47-58.
6. Levy JMM, Towers CG, Thorburn A. Targeting autophagy in cancer. *Nat Rev Cancer*. 2017, 17(9):528-542.
7. Khan T, Relitti N, Brindisi M, Magnano S, Zisterer D, Gemma S, Butini S, Campiani G. Autophagy modulators for the treatment of oral and esophageal squamous cell carcinomas. *Med Res Rev*. 2019.
8. Kiruthiga C, Devi KP, Nabavi SM, Bishayee A. Autophagy: A Potential Therapeutic Target of Polyphenols in Hepatocellular Carcinoma. *Cancers (Basel)*. 2020, 12(3).
9. Lee JS, Jang EH, Woo HA, Lee K. Regulation of Autophagy Is a Novel Tumorigenesis-Related Activity of Multifunctional Translationally Controlled Tumor Protein. *Cells*. 2020, 9(1).
10. Nuñez-Olvera SI, Gallardo-Rincón D, Puente-Rivera J, Salinas-Vera YM, Marchat LA, Morales-Villegas R, López-Camarillo C. Autophagy Machinery as a Promising Therapeutic Target in Endometrial Cancer. *Frontiers in Oncology*. 2019, 9.
11. Tomczak K, Czerwińska P, Wiznerowicz M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkologia*. 2015, 1A:68-77.
12. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol*. 2008, 4(8):e1000117.
13. Liu J, Feng M, Li S, Nie S, Wang H, Wu S, Qiu J, Zhang J, Cheng W. Identification of molecular markers associated with the progression and prognosis of endometrial cancer: a bioinformatic study.

Cancer Cell Int. 2020, 20:59.

14. Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, Chen Y, Wang H, Xu M. ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer Management and Research*. 2018, Volume 10:663-670.
15. Zheng H, Zhang G, Zhang L, Wang Q, Li H, Han Y, Xie L, Yan Z, Li Y, An Y et al. Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. *Front Oncol*. 2020, 10:68.
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012, 16(5):284-287.
17. Jian L, Yang G. Identification of Key Genes Involved in Diabetic Peripheral Neuropathy Progression and Associated with Pancreatic Cancer. *Diabetes Metab Syndr Obes*. 2020, 13:463-476.
18. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019, 10(1):1523.
19. Zhou M, Zhang Z, Zhao H, Bao S, Sun J. A novel lncRNA-focus expression signature for survival prediction in endometrial carcinoma. *BMC Cancer*. 2018, 18(1):39.
20. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. 2007, 23(23):3251-3253.
21. Dienstmann R, Salazar R, Tabernero J: Personalizing colon cancer adjuvant therapy. selecting optimal treatments for individual patients. *J Clin Oncol*. 2015, 33(16):1787-1796.
22. Giatromanolaki AN, Charitoudis GS, Bechrakis NE, Kozobolis VP, Koukourakis MI, Foerster MH, Sivridis EL. Autophagy patterns and prognosis in uveal melanomas. *Mod Pathol*. 2011, 24(8):1036-1045.
23. Cao Y, Luo Y, Zou J, Ouyang J, Cai Z, Zeng X, Ling H, Zeng T. Autophagy and its role in gastric cancer. *Clin Chim Acta*. 2019, 489:10-20.
24. Liu J, Lin Y, Yang H, Deng Q, Chen G, He J. The expression of p33(ING1), p53, and autophagy-related gene Beclin1 in patients with non-small cell lung cancer. *Tumour Biol*. 2011, 32(6):1113-1121.
25. Zhou M, Liu X, Li Z, Huang Q, Li F, Li CY. Caspase-3 regulates the migration, invasion and metastasis of colon cancer cells. *Int J Cancer*. 2018, 143(4):921-930.
26. Lin J, Zhang Y, Wang H, Chang J, Wei L, Cao L, Zhang Z, Zhang X. Genetic Polymorphisms in the Apoptosis-Associated Gene CASP3 and the Risk of Lung Cancer in Chinese Population. *PLoS One*. 2016, 11(10):e0164358.
27. Flanagan L, Meyer M, Fay J, Curry S, Bacon O, Duesmann H, John K, Boland KC, McNamara DA, Kay EW et al. Low levels of Caspase-3 predict favourable response to 5FU-based chemotherapy in advanced colorectal cancer: Caspase-3 inhibition as a therapeutic approach. *Cell Death Dis*. 2016, 7:e2087.
28. Liu PF, Hu YC, Kang BH, Tseng YK, Wu PC, Liang CC, Hou YY, Fu TY, Liou HH, Hsieh IC et al. Expression levels of cleaved caspase-3 and caspase-3 in tumorigenesis and prognosis of oral tongue

- squamous cell carcinoma. PLoS One. 2017, 12(7):e0180620.
29. Pu X, Storr SJ, Zhang Y, Rakha EA, Green AR, Ellis IO, Martin SG. Caspase-3 and caspase-8 expression in breast cancer: caspase-3 is associated with survival. Apoptosis. 2017, 22(3):357-368.
  30. Huang KH, Fang WL, Li AF, Liang PH, Wu CW, Shyr YM, Yang MH. Caspase-3, a key apoptotic protein, as a prognostic marker in gastric cancer after curative surgery. Int J Surg. 2018, 52:258-263.
  31. Burke LS, Hyland PL, Pfeiffer RM, Prescott J, Wheeler W, Mirabello L, Savage SA, Burdette L, Yeager M, Chanock S et al. Telomere length and the risk of cutaneous malignant melanoma in melanoma-prone families with and without CDKN2A mutations. PLoS One. 2013, 8(8):e71121.
  32. Padhi SS, Roy S, Kar M, Saha A, Roy S, Adhya A, Baisakh M, Banerjee B. Role of CDKN2A/p16 expression in the prognostication of oral squamous cell carcinoma. Oral Oncol. 2017, 73:27-35.
  33. Aftab A, Shahzad S, Hussain HMJ, Khan R, Irum S, Tabassum S. CDKN2A/P16INK4A variants association with breast cancer and their in-silico analysis. Breast Cancer. 2019, 26(1):11-28.
  34. Wujcicka W, Zajac A, Szylo K, Smolarz B, Romanowicz H, Stachowiak G. Association of SNPs in CDKN2A (P14ARF) Tumour Suppressor Gene With Endometrial Cancer in Postmenopausal Women. In Vivo. 2020, 34(2):943-951.
  35. Ostrander JH, Daniel AR, Lange CA. Brk/PTK6 signaling in normal and cancer cell models. Curr Opin Pharmacol. 2010, 10(6):662-669.
  36. Harvey AJ, Pennington CJ, Porter S, Burmi RS, Edwards DR, Court W, Eccles SA, Crompton MR. Brk protects breast cancer cells from autophagic cell death induced by loss of anchorage. Am J Pathol. 2009, 175(3):1226-1234.
  37. Rozier L, El-Achkar E, Apiou F, Debatisse M. Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. Oncogene. 2004, 23(41):6872-6880.
  38. Ngollo M, Lebert A, Daures M, Judes G, Rifai K, Dubois L, Kemeny JL, Penault-Llorca F, Bignon YJ, Guy L et al. Global analysis of H3K27me3 as an epigenetic marker in prostate cancer progression. BMC Cancer. 2017, 17(1):261.

## Tables

**Table 1** Clinicopathological characteristics of UCEC patients used in this study

| Variables              | TCGA cases( <i>n</i> =401) |               |
|------------------------|----------------------------|---------------|
| Age, years (mean ± SD) | ≤65                        | 251(55.5±7.7) |
|                        | >65                        | 150(74.4±6.2) |
| Grade, no(%)           | 1                          | 97(24.2)      |
|                        | 2                          | 118(29.4)     |
|                        | 3                          | 186(46.4)     |
| Vital status, no(%)    | Alive                      | 358(89.3)     |
|                        | Dead                       | 43(10.7)      |

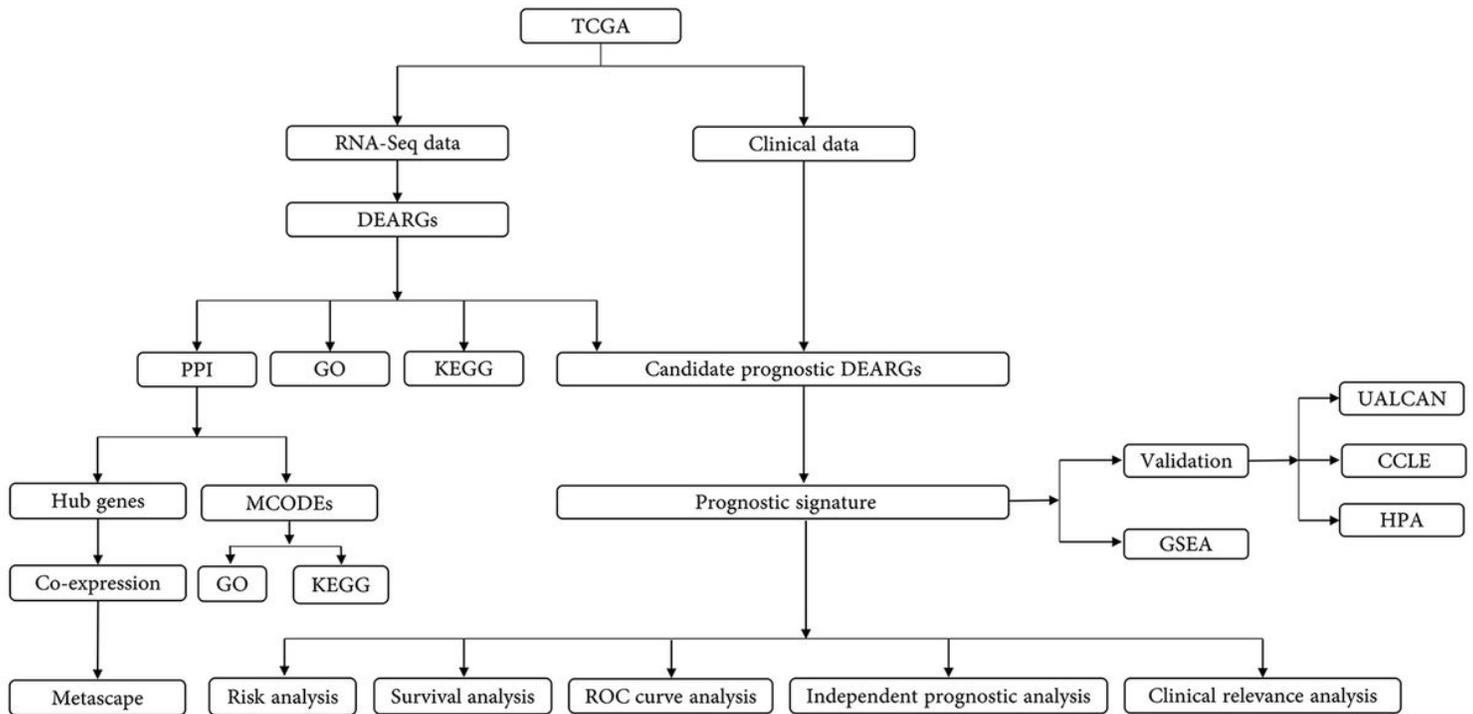
**Table 2** The common DEARGs identified in TCGA

| Common DEARGs        | Gene names   |
|----------------------|--|
| Upregulated DEARGs   | BID,PARP1,CLN3,CASP3,PRKCD,WIPI1,BAK1,CTSB,BAX,ATG4D,DAPK2,<br>ERO1A,BNIP3,P4HB,GAPDH,IKBKE,TP63,IFNG,EIF4EBP1,PTK6,BIRC5,<br>CDKN2A,SERPINA |
| Downregulated DEARGs | NRG3,ITPR1,FOS,GRID1,HSPB8,DLC1,FOXO1,BCL2,MYC,GABARAPL1,<br>PRKN,GRID2,TUSC1,CDKN1B,CALCOCO2,RAB33B,VAMP3,ST13,NRG2,<br>PRKAR1A,ITGB1,MAPK3 |

**Table 3** Prognostic signature of autophagy genes

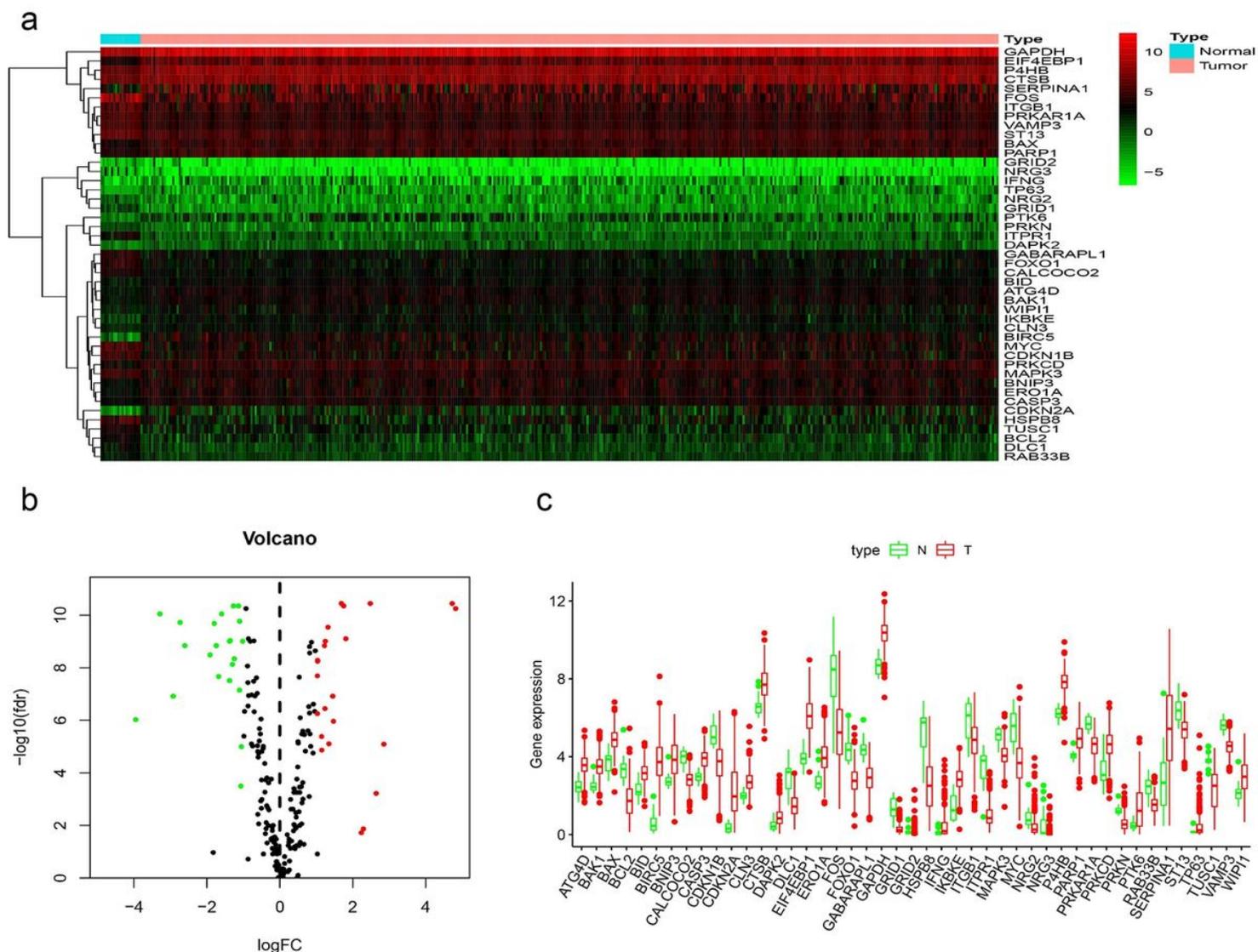
| Genes  | Coefficient | HR    | 95%CI         | p-value |
|--------|-------------|-------|---------------|---------|
| CDKN2A | 0.202       | 1.224 | 1.001 - 1.497 | 0.048   |
| PTK6   | 0.426       | 1.531 | 1.186 - 1.975 | 0.001   |
| GRID2  | 1.190       | 3.288 | 1.468 - 7.363 | 0.004   |

## Figures



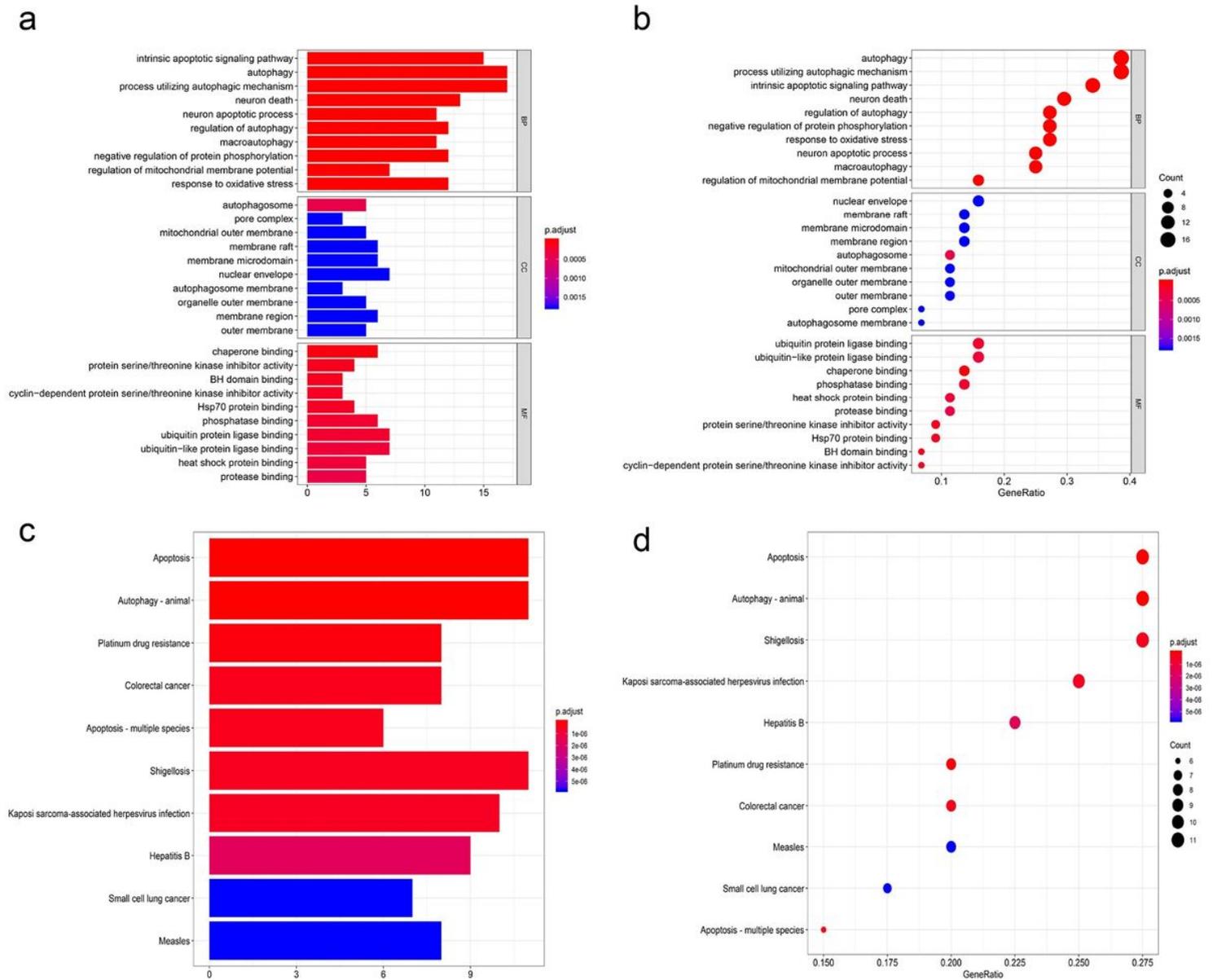
**Figure 1**

Flow chart of study design.



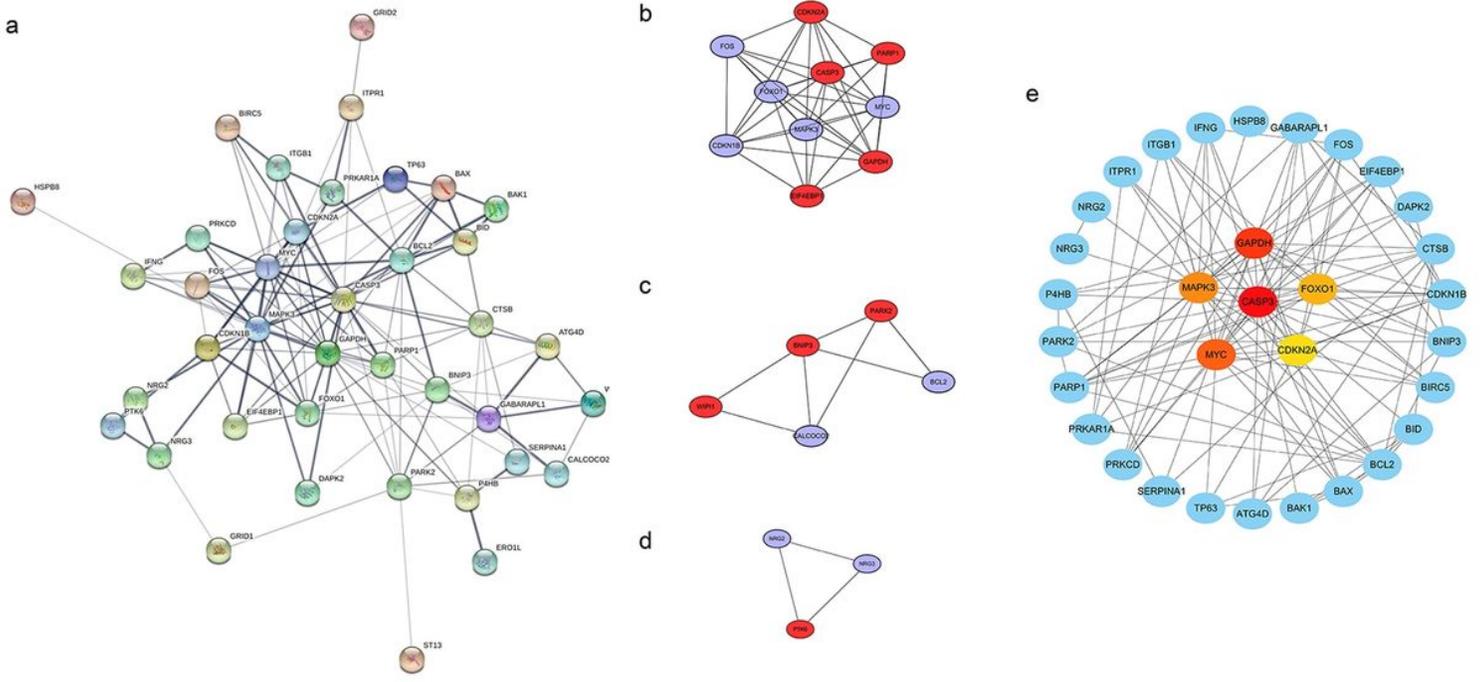
**Figure 2**

Identification of differentially expressed autophagy-related genes (DEARGs) in TCGA. a Heatmap of 45 DEARGs. Each column represented a biological sample and each row in the heatmap represents a gene. Genes shown in red are up-regulated and blue are down-regulated. b Volcano plot of the 202 autophagy genes. The red dot represents up-regulated genes, and green dot represents down-regulated genes. c Boxplot of the expression levels of 45 DEARGs between tumor and normal. N, normal samples; T, tumor samples. Statistically significant DEARGs were defined as  $\log_2|FC| > 1.0$  and  $FDR < 0.05$ .



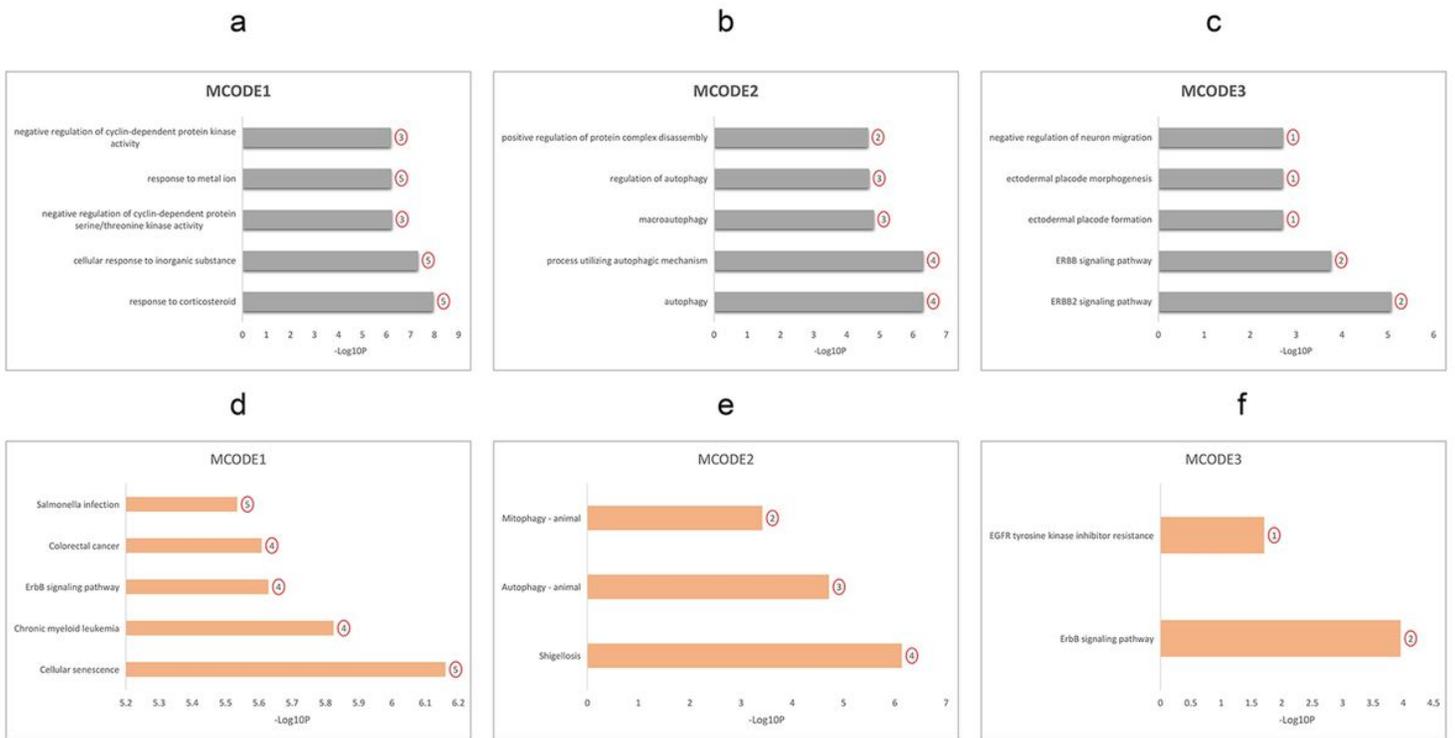
**Figure 3**

Functional enrichment analysis on the DEARGs. a, b GO analysis of the DEARGs. c, d KEGG analysis of the DEARGs.  $P_{\text{adjust}} < 0.05$  was considered statistically significant. BP, biological process; CC, cellular component; MF, molecular function.



**Figure 4**

PPI network and hub genes analysis. a PPI network of 45 DEARGs. PPI network contained 45 nodes and 140 edges. b-d 3 MCODEs identification in the PPI network. The red node represents the up-regulated gene and the blue node represents the down-regulated gene. e Hub genes screening. The darker the color of the cube, the more important it is. Hub genes were CASP3, MYC, MAPK3, GAPDH, FOXO1, CDKN2A.



**Figure 5**

GO and KEGG analyses on the 3 MCODEs. a GO analysis on MOCDE1. b GO analysis on MOCDE2. c GO analysis of MOCDE3. d KEGG analysis on MOCDE1. e KEGG analysis on MOCDE2. f KEGG analysis on MOCDE3.

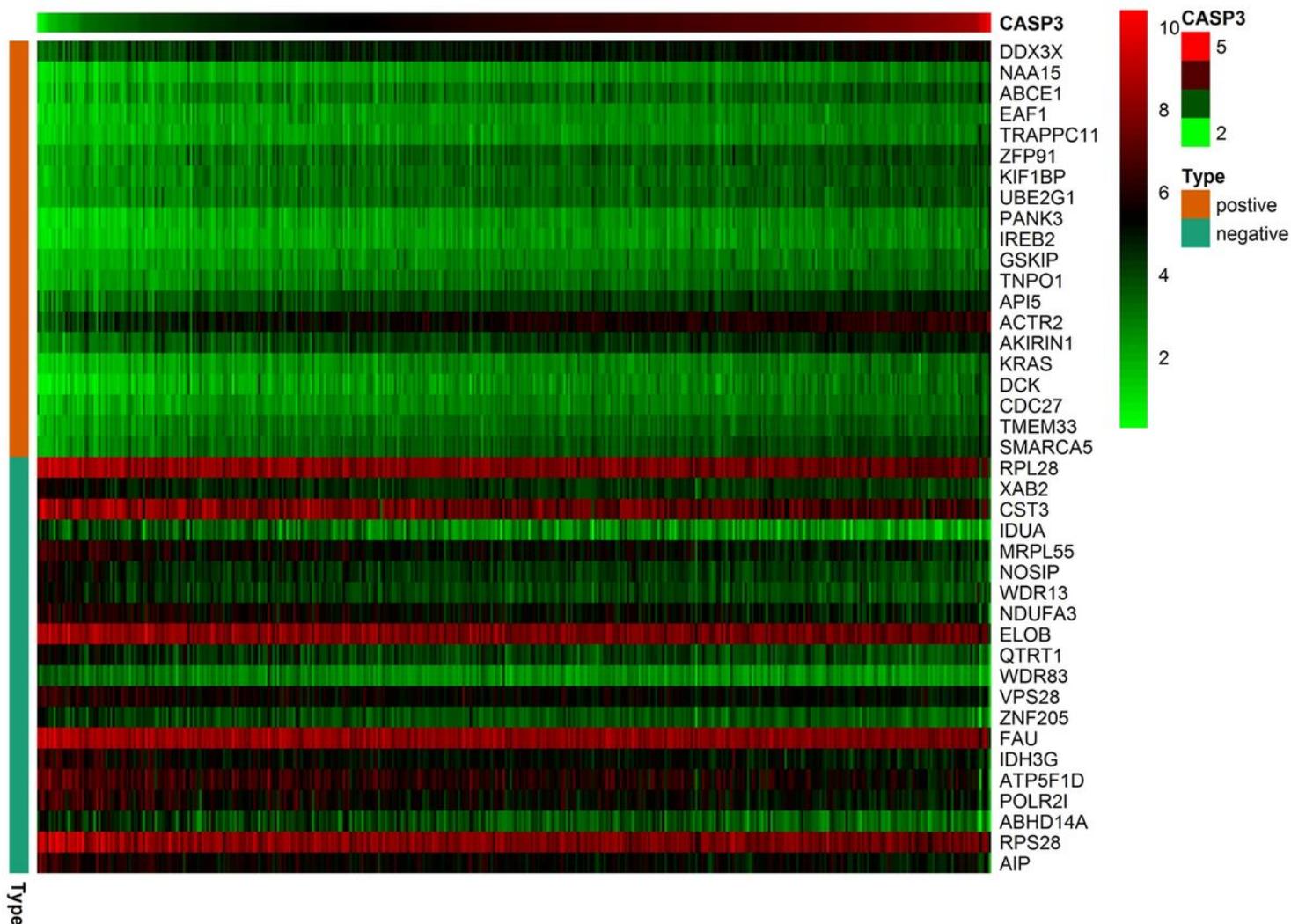
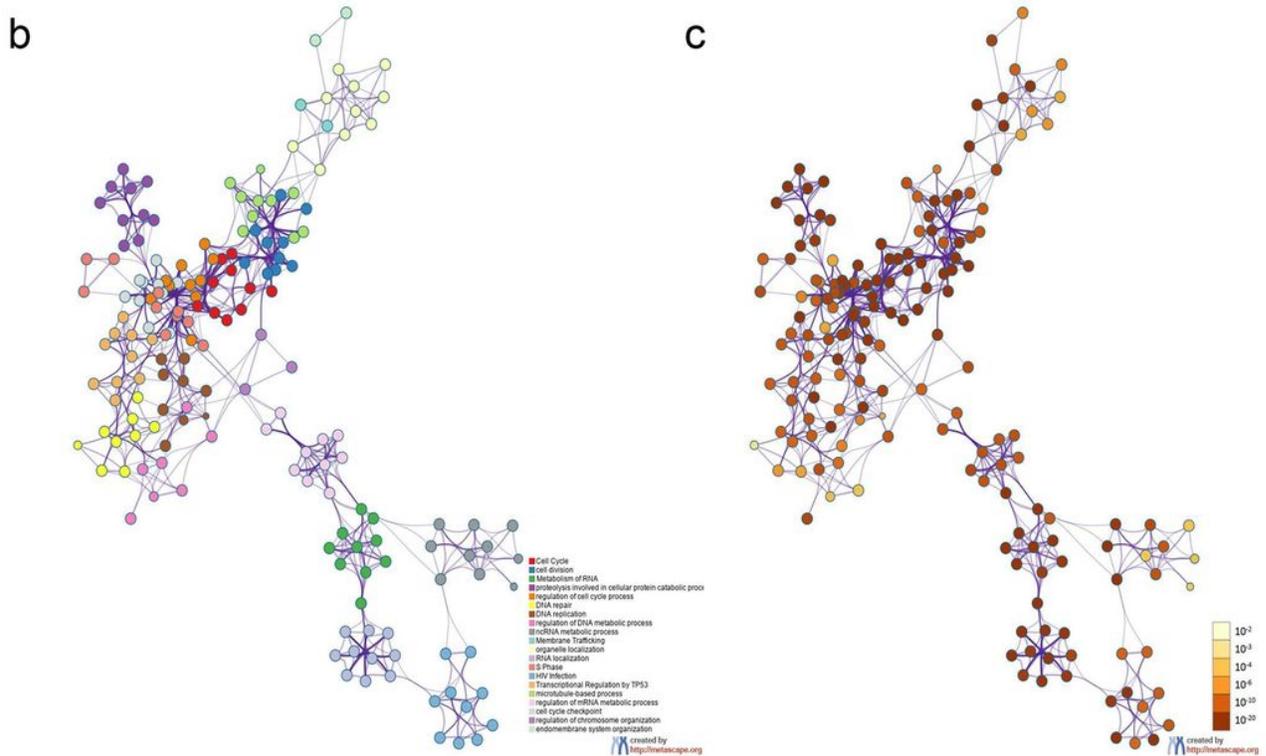
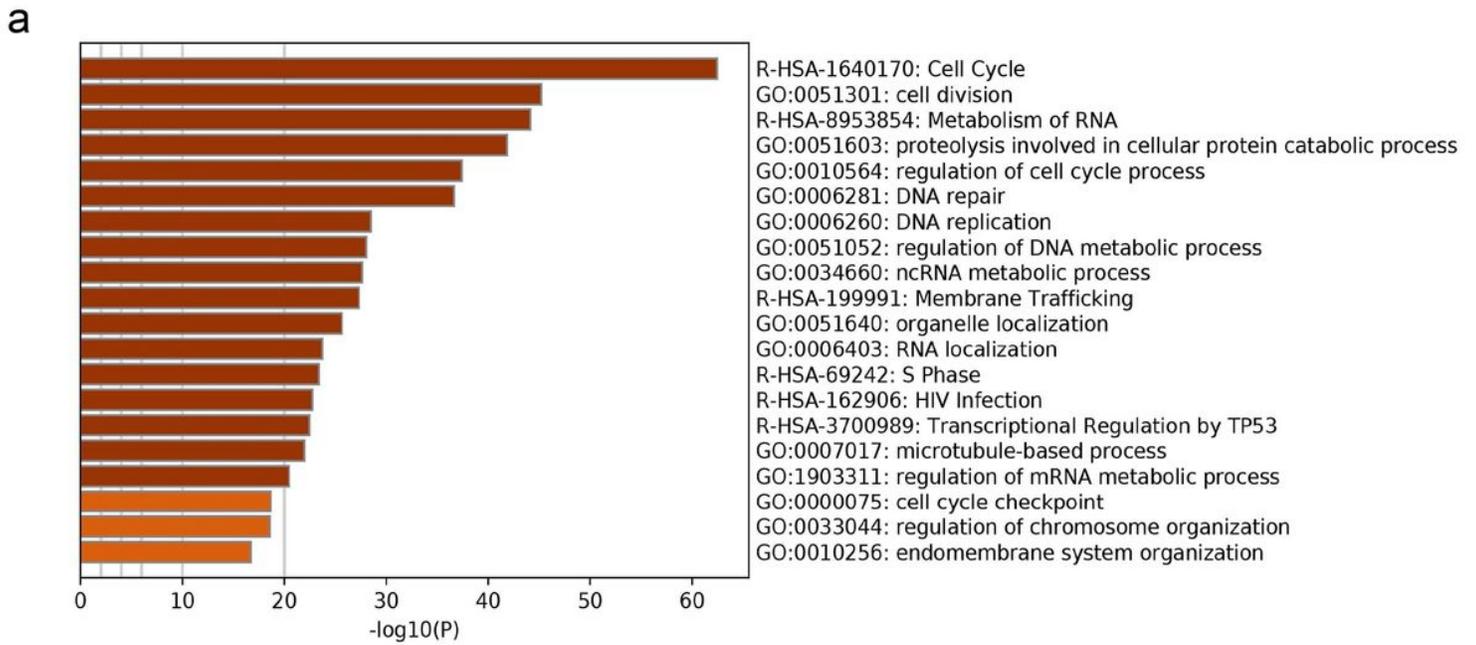


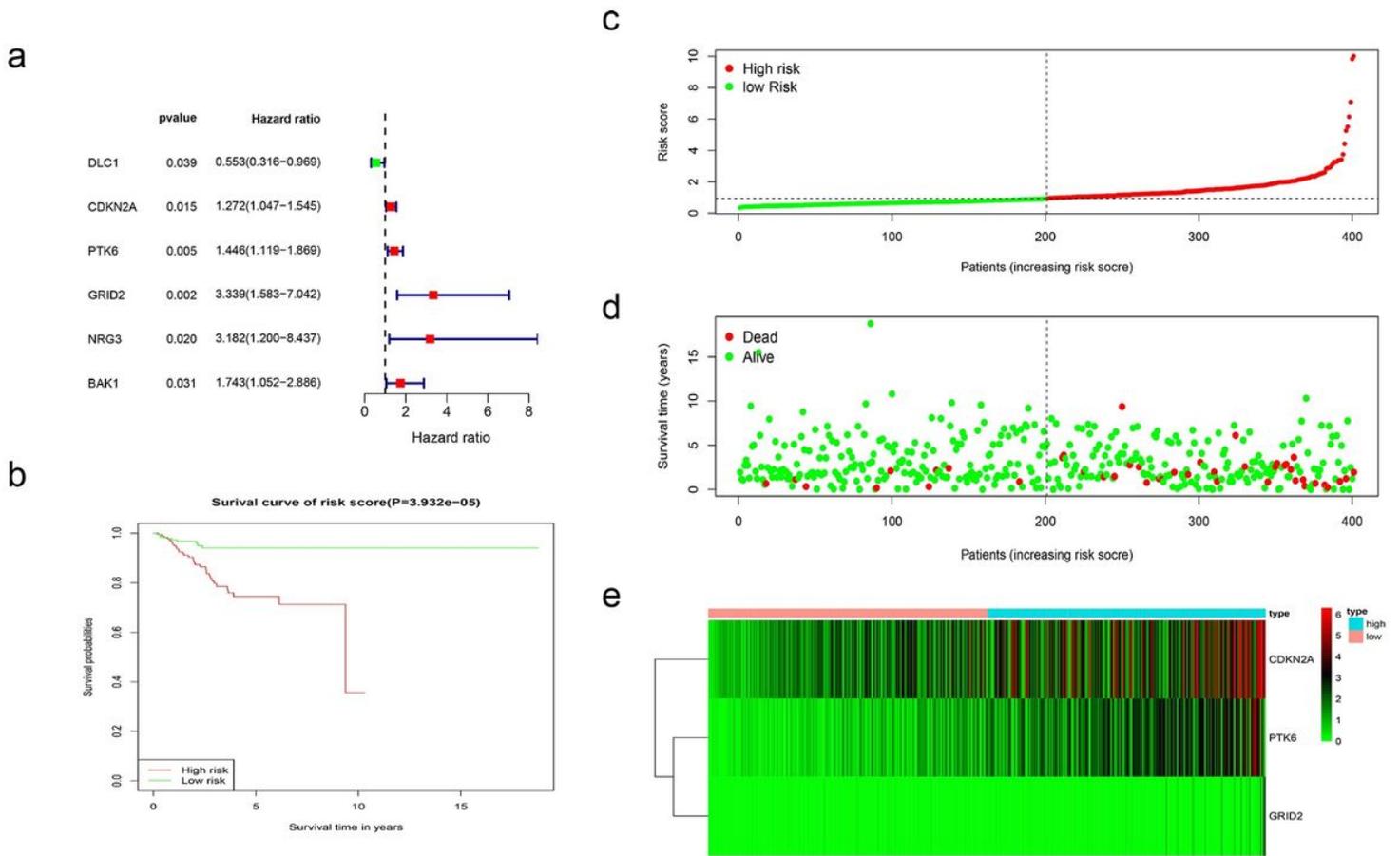
Figure 6

Genes co-expressed with CASP3 in TCGA. Heatmap of the top 20 positive correlated genes and the top 20 negative correlated genes. Each column represented a biological sample and each row in the heatmap represents a gene. Genes shown in red are up-regulated and blue are down-regulated.



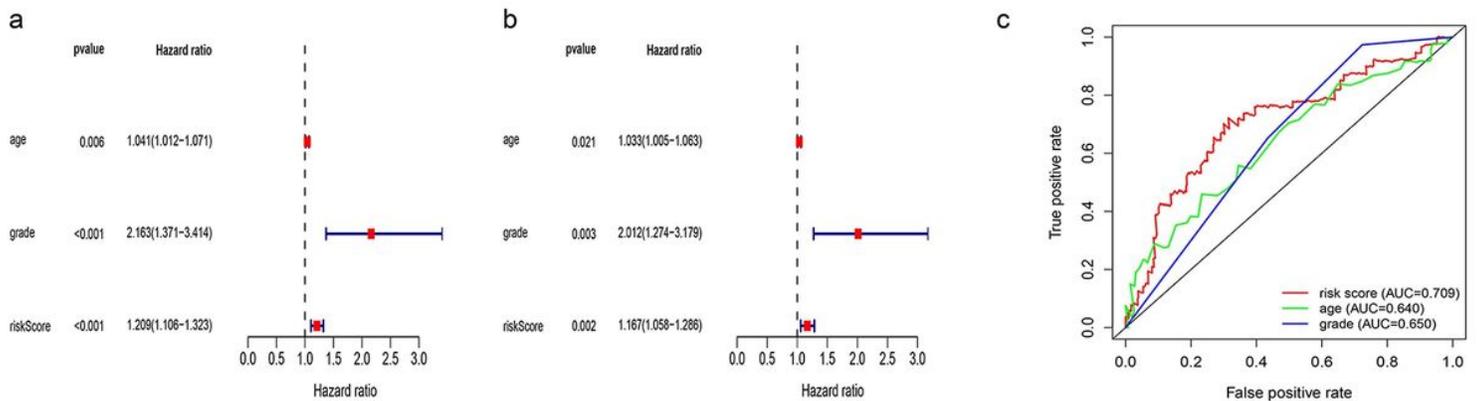
**Figure 7**

Metascape Gene List Analysis Report. a Bar graph of enriched terms across input gene lists, colored by p-values. Network of enriched terms: b colored by cluster ID, where nodes that share the same cluster ID are typically close to each other; c colored by p-value, where terms containing more genes tend to have a more significant p-value.



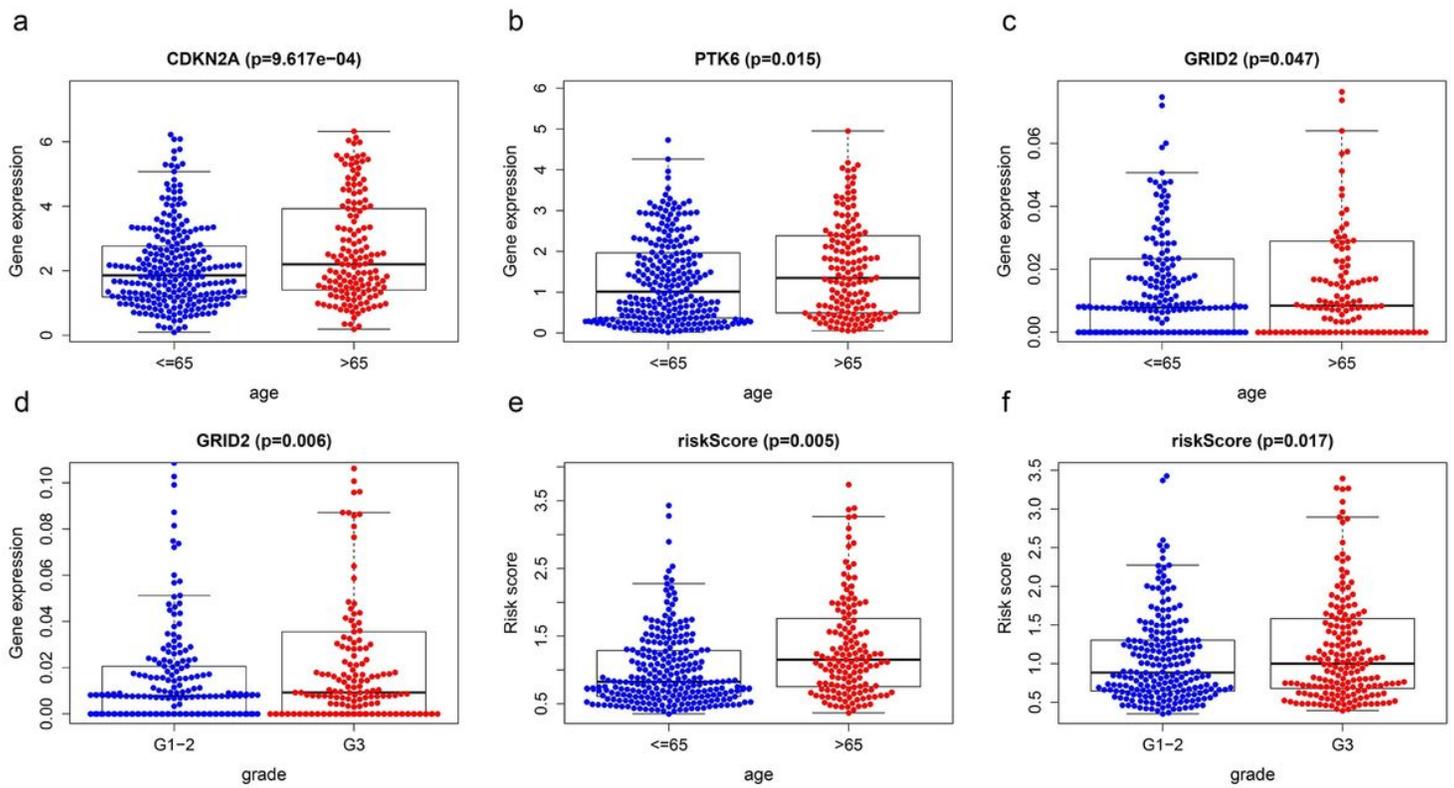
**Figure 8**

Survival assessment of the prognostic signature in the TCGA. a The univariate Cox proportional hazards regression analysis screened out 6 EC relative genes. b Kaplan-Meier curves for overall survival of patients classified into high-risk and low-risk groups using the prognostic signature in TCGA. c, d The distribution of risk score and patients' survival status marked as low risk (blue) or high risk (red), as divided by the threshold (vertical black line). e Gene expression and risk score distribution in TCGA by z-score, with red indicating higher expression and light blue indicating lower expression.



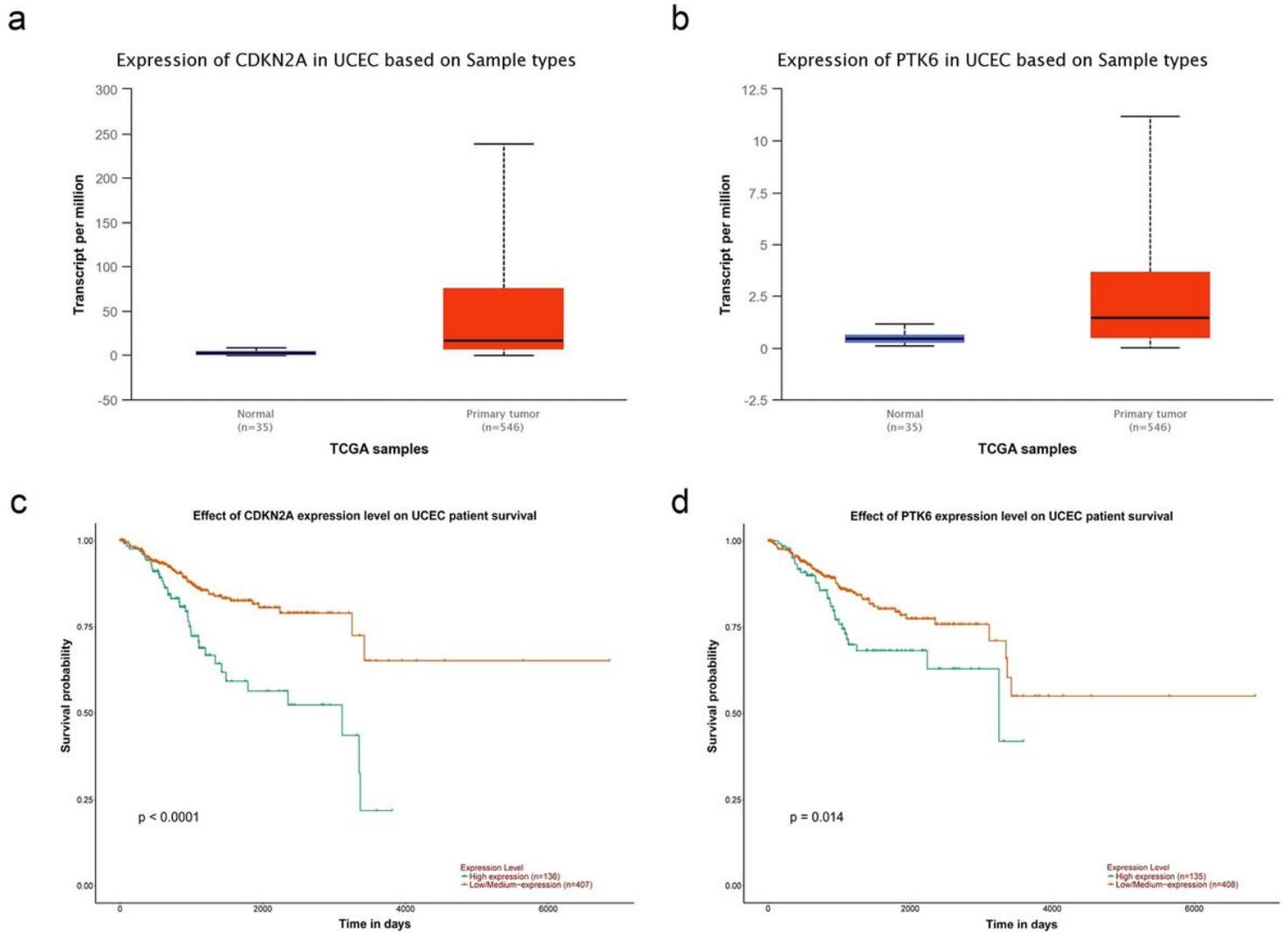
**Figure 9**

Univariate and multivariate independent prognostic analysis. a Univariate independent prognostic analysis. b Multivariate independent prognostic analysis. c Multi-index ROC curve. The grouping settings are as follows: Age (age>65 vs age≤65); Grade (G3 vs G1-2); Risk score (high-risk vs low-risk).



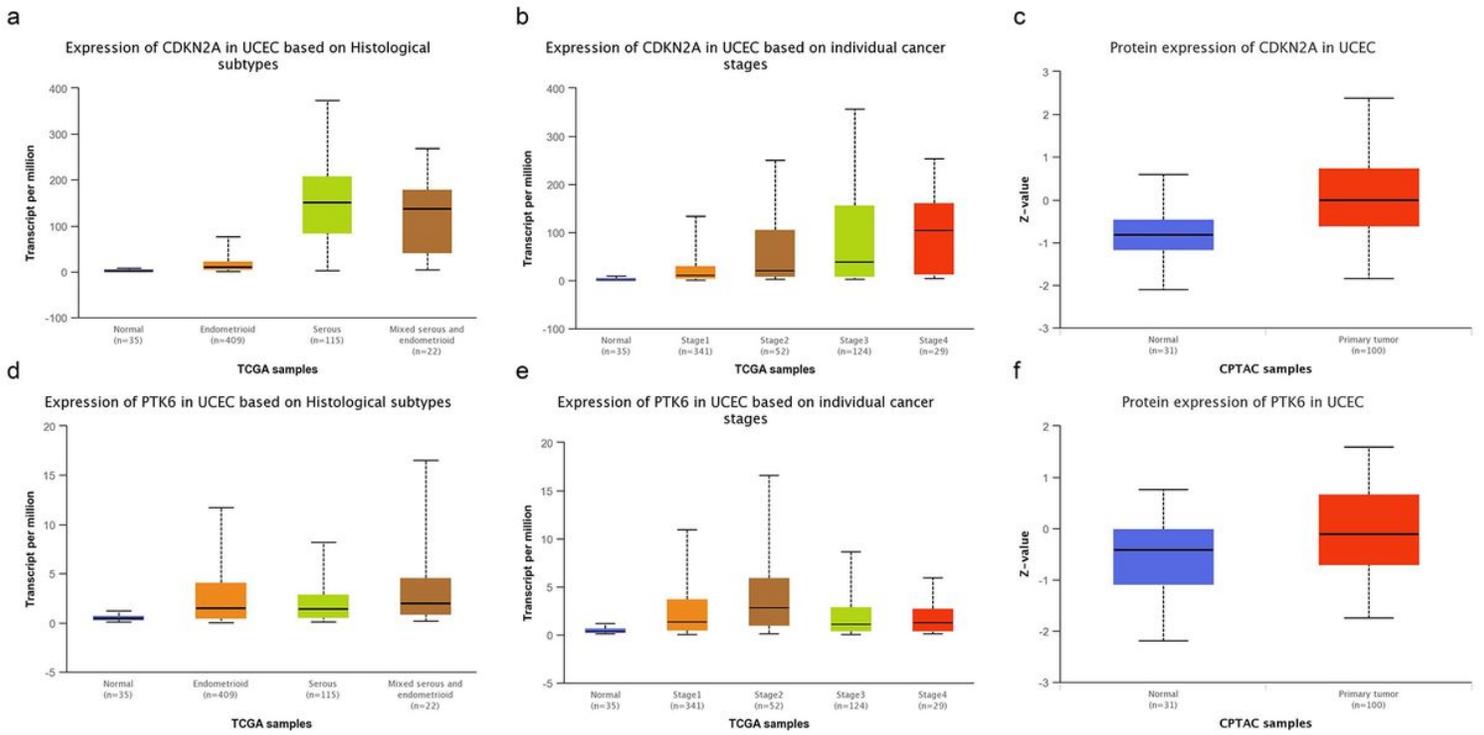
**Figure 10**

Clinical correlation analysis. a-c Gene expression in different age groups. d GRID2 in different pathological grades. e Risk score in different age groups. f Risk score in different pathological grades.



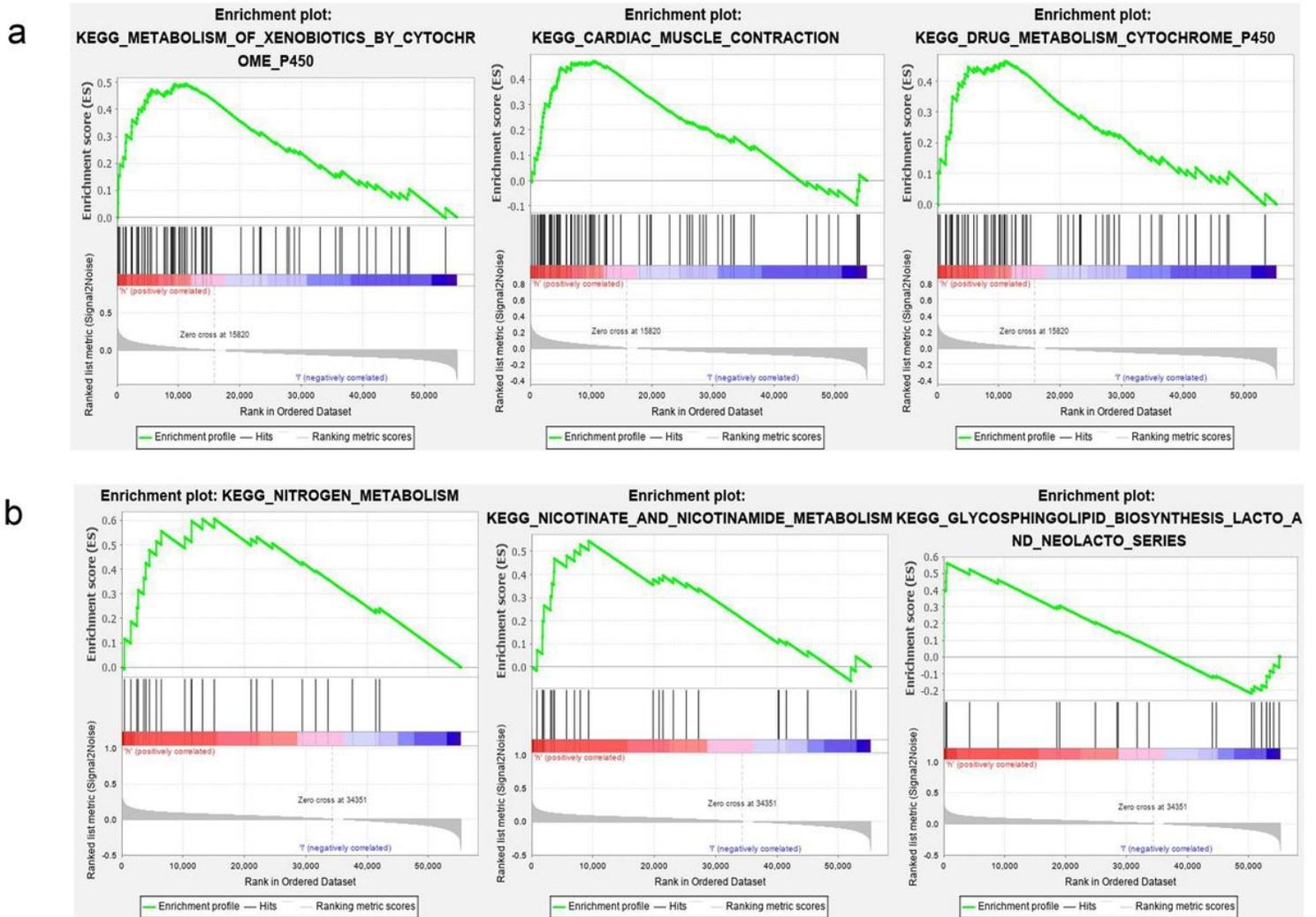
**Figure 11**

Validation of UALCAN website. a, b The expression of CDKN2A and PTK6 in UCEC samples. c, d Survival analysis of CDKN2A and PTK6.



**Figure 12**

Validation of UALCAN website. a, d The expression of CDKN2A and PTK6 in UCEC samples of different histological subtypes. b, e The expression of CDKN2A and PTK6 in UCEC samples at different stages. c, f The protein expression of CDKN2A and PTK6 in UCEC samples.



**Figure 13**

Gene set enrichment analysis (GSEA) of high versus low-expression groups. a The three most functional gene sets enriched in EC samples with CDKN2A highly expressed. b The three most functional gene sets enriched in EC samples with PTK6 highly expressed.

## Supplementary Files

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

- [TableS3.xls](#)
- [TableS4.xls](#)
- [TableS1.txt](#)
- [TableS2.txt](#)