Expression and prognosis of CDC45 in cervical cancer based on the GEO database

Zikang He [1], **Xiaojin Wang** [1], **Zhiming Yang** [2], **Ying Jiang** [1], **Luhui Li** [1], **Xingyun Wang** [1], **Zheyao Song** [1], **Xiuli Wang** [1,3], **Jiahui Wan** [4], **Shijun Jiang** [1,5], **Naiwen Zhang** [1], **Rongjun Cui** [Corresp. 1]

[1] Department of Biochemistry and Molecular Biology, Mudanjiang Medical University, Mudanjiang, China

[2] Department of Clinical Laboratory, Handan central Hospital, Handan, China

[3] Department of Clinical Laboratory, The Seventh Hospital in Qiqihar, Qiqihar, China

[4] Department of Clinical Laboratory, Harbin Public Security Hospital, Harbin, China

[5] Department of Clinical Laboratory, Daqing Medical College, Daqing, China

Corresponding Author: Rongjun Cui
Email address: cuirongjun@mdjmu.edu.cn

Cervical cancer is one of the most common malignant tumors in women, and its morbidity and mortality are increasing year by year worldwide. Therefore, an urgent and challenging task is to identify potential biomarkers for cervical cancer. This study aims to identify the hub genes based on the GEO database and then validate their prognostic values in cervical cancer by multiple databases. By analysis, we obtained 83 co-expressed differential genes from the GEO database (GSE63514, GSE67522, and GSE39001). GO and KEGG enrichment analysis showed that these 83 co-expressed it mainly involved differential genes in DNA replication, cell division, cell cycle, etc.. The PPI network was constructed and top 10 genes with protein-protein interaction were selected. Then, we validated ten genes using some databases such as TCGA, GTEx and oncomine. Survival analysis demonstrated significant differences in CDC45, RFC4, TOP2A. Differential expression analysis showed that these genes were highly expressed in cervical cancer tissues. Furthermore, univariate and multivariate cox regression analysis indicated that CDC45 and clinical stage IV were independent prognostic factors for cervical cancer. In addition, the HPA database validated the protein expression level of CDC45 in cervical cancer. Further studies investigated the relationship between CDC45 and tumor-infiltrating immune cells via CIBERSORT. Finally, gene set enrichment analysis (GSEA) showed CDC45 related genes were mainly enriched in cell cycle, chromosome, catalytic activity acting on DNA, etc.. These results suggested CDC45 may be a potential biomarker associated with the prognosis of cervical cancer.

Expression and Prognosis of CDC45 in Cervical Cancer based on the GEO database

3   **Zikang He[1], Xiaojin Wang[1], Zhiming Yang[2], Ying Jiang[1], Luhui Li[1], Xingyun Wang[1],**

4   **Zheyao Song[1], Xiuli Wang[1,3], Jiahui Wan[4], Shijun Jiang[1,5], Naiwen Zhang[1], Rongjun Cui[1]**

5   [1]Department of Biochemistry and Molecular Biology, Mudanjiang Medical University,

6   Mudanjiang, Heilongjiang, 157011, China.

7   [2]Department of Clinical Laboratory, Handan Central Hospital, Handan, Hebei, 056001, China.

8   [3]Department of Clinical Laboratory, The Seventh Hospital in Qiqihar, Qiqihar, Heilongjiang,

9   161000, China.

10   [4]Department of Clinical Laboratory, Harbin Public Security Hospital,Harbin, Heilongjiang,

11   150000, China.

12   [5]Department of Clinical Laboratory, Daqing Medical College, Daqing, Heilongjiang, 163311,

13   China.

14   Corresponding author:

15   Rongjun Cui[1]

16   E-mail: cuirongjun@mdjmu.edu.cn

21   **Abstract**

22 Cervical cancer is one of the most common malignant tumors in women, and its morbidity and

23 mortality are increasing year by year worldwide. Therefore, an urgent and challenging task is to

24 identify potential biomarkers for cervical cancer. This study aims to identify the hub genes based

25 on the GEO database and then validate their prognostic values in cervical cancer by multiple

26 databases. By analysis, we obtained 83 co-expressed differential genes from the GEO database

27 (GSE63514, GSE67522, and GSE39001). GO and KEGG enrichment analysis showed that these

28 83 co-expressed it mainly involved differential genes in DNA replication, cell division, cell cycle,

29 etc.. The PPI network was constructed and top 10 genes with protein-protein interaction were

30 selected. Then, we validated ten genes using some databases such as TCGA, GTEx and

31 oncomine. Survival analysis demonstrated significant differences in CDC45, RFC4, TOP2A.

32 Differential expression analysis showed that these genes were highly expressed in cervical

33 cancer tissues. Furthermore, univariate and multivariate cox regression analysis indicated that

34 CDC45 and clinical stage IV were independent prognostic factors for cervical cancer. In

35 addition, the HPA database validated the protein expression level of CDC45 in cervical cancer.

36 Further studies investigated the relationship between CDC45 and tumor-infiltrating immune cells

37 via CIBERSORT. Finally, gene set enrichment analysis (GSEA) showed CDC45 related genes

38 were mainly enriched in cell cycle, chromosome, catalytic activity acting on DNA, etc.. These

39 results suggested CDC45 may be a potential biomarker associated with the prognosis of cervical

40 cancer.

41 **Introduction**

42   Cervical cancer (CC) is one of the most prevalent gynecological malignancy in women, and

43   its incidence and mortality are second only to breast cancer. It is estimated that 570,000 cases

44   and 311,000 deaths from cervical cancer worldwide occurred in 2018[1]. The progression of CC

45   takes approximately 10 to 20 years from a benign to a malignant disease, with squamous cell

46   carcinoma being its most common subtype[2]. Despite significant advances in screening and

47   various treatments, such as surgery, radiotherapy, and chemotherapy, deficiencies remain. Some

48   studies indicated that over 90% of cases are caused by persistent infection with human

49   papillomavirus (HPV), the main subtypes of which are HPV16 and HPV18[3]. The genetic

50   sensitivity of CC is caused by HPV infection, which leads to genetic mutations. For instance, the

51   polymorphism of GSTM1 is associated with high-risk HPV infection[4]. In recent years, with the

52   popularization and sharing of biomedical big data, the screening of effective molecular targets

53   related to CC has become possible through bioinformatics methods. The previous studies have

54   reported some targeted molecules regarding CC treatment [5, 6], but the clinical applications are

55   very limited or even almost none. Therefore, an urgent and challenging task is to continue to

56   explore early biomarkers in CC.

57   Cell division cycle (CDC45) is one of the proteins essential for the initiation and extension

58   of DNA replication and for regulating DNA replication. It has been found that CDC45, mini-

59   chromosome maintenance protein complex (MCM) and Go-Ichi-Ni-San (GINS) forms a "super

60   complex" that is the central to eukaryotic replicons and has been shown to have helicase

61   activity[7]. It binds to DNA molecules and unwinds double-stranded DNA to form a replication

62   fork structure throughout the entire DNA replication process[8, 9]. The previous studies reported

63   that CDC45 may be a proliferation-associated antigen and contribute to the progression of

64   malignant tumors[10]. However, the expression and function of CDC45 in CC still remains

65   unknown.

66   In this study, we screened differentially expressed genes (DEGs) between cervical cancer

67   tissues and normal or adjacent non-cancerous tissues based on the Gene Expression Omnibus

68   (GEO) database, and collated and analyzed DEGs by a series of bioinformatics methods. Finally,

69   we confirmed the important role of CDC45 in the development and prognosis of cervical cancer.

70   **Material and Methods**

71   **1. Data collection and data processing**

72   Using the keyword "cervical cancer" search on the GEO database[11] (https://www.ncbi.nlm.

73   nih.gov/geo/). The gene expression microarrays of GSE63514[12], GSE67522[13, 14], GSE39001[15]

74   and GSE52903[16] were downloaded. The GSE63514 dataset included 28 cancer tissues and 24

75   non-cancerous tissues. GSE67522 contained 20 cancer tissues and 22 non-cancerous tissues.

76   GSE39001 contained 43 cancer tissues and 12 non-cancerous tissues. GSE52903 contained 55

77   cancer tissues and 17 non-cancerous tissues. Among them, the first three data sets are used as

78   training sets and the last data set is used as a validation set. As the data come from the online

79   database, no further approval from the Ethics Committee was required.

80   The differentially expressed genes (DEGs) between cancer tissues and non-cancerous

81   tissues were screened out using GEO2R[17] (http://www.ncbi. nlm.nih.gov/geo/geo2r). Probe sets

82   with no corresponding gene symbols or genes with multiple sets of probe were removed or

83   averaged, separately. $|LogFC| > 1$ and FDR < 0.05 were selected statistically significant ($|logFC|$

84   stands for absolute value of the log fold change and FDR stands for false discovery rate). Co-

85   expressed genes were obtained by intersection of DEGs from three datasets using Draw Venn

86   Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/).

87   **2. GO and KEGG pathway enrichment analyses of DEGs**

88   To identify DEGs associated pathways and function annotations, Gene Ontology (GO) and

89   The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted

90   by DAVID online database [18, 19] (DAVID; https://david.ncifcrf.gov). GO is a widely used

91   ontology in the field of bioinformatics, which covers three aspects of biology: biological process

92   (BP), cellular component (CC), and molecular function (MF)[20]. KEGG is one of the most

93   commonly used bioinformatics tools in the world for understanding advanced functional and

94   high-throughput experimental technologies of biological systems[21]. P <0.05 indicated

95   statistically significant difference.

96   **3. Construction of PPI network and screening of Key genes**

97   The PPI network is to analyze the functional interactions between proteins using STRING

98   online database, which is helpful to mine the core regulatory genes for the mechanisms of

99   generation or development of diseases[22] (STRING; https://string-db.org/). In this study, we

100   constructed to PPI network of DEGs and selected one interaction that was statistically significant

101   with a composite score > 0.4. Then, the PPI networks were mapped using Cytoscape 3.7.2 [23]

102   (https://cytoscape.org/), and the top 10 genes with the protein-protein interaction among the

103   network were identified using CytoHubba which is a plug-in to Cytoscape.

104   **4. Validation of hub genes by the TCGA, GTEx, and Oncomine databases.**

105   The transcriptome profiling counts data of CC were downloaded from the Cancer Genome

106   Altas (TCGA) database[24] (https://portal.gdc.cancer.gov/) with all subtypes of the project

107   TCGA-CESC as inclusion criteria. Normal cervix tissues were downloaded from the Genotype-

108   Tissue Expression (GTEx) database[25] (https://www.gtexportal.org/home/index.html). Data from

109   both databases were combined and normalized. A total of 319 samples of CC included 306

110   cancer tissues and 13 adjacent non-cancerous tissues. We performed differential analysis of these

111   samples and plotted the volcanic map and heat map using Perl[26] (http://www.perl.org/, version

112   5.32.0) and edgeR [27, 28] (http://bioinf.wehi.edu.au/edgeR/, version 4.0.2). |LogFC| > 1 and FDR

113   < 0.05 were selected for statistical significance. Next, we performed overall survival (OS)

114   analysis of 10 key genes using Gene Expression Profiling Interactive Analysis (GEPIA) tool[29]

115   (http://gepia.cancer-pku.cn/), which is a new web server for analyzing the RNA sequencing

116   expression data from the TCGA and the GTEx databases. The expression level of hub genes in

117   tumor and normal tissues was shown using GEPIA and Oncomine online database[30]

118   (https://www.onco mine.org/resource/login.html).

119   **5. Cox proportional hazards regression analyses**

120   Clinical information data on CC was downloaded from TCGA database. We analyzed the

121   association with clinical information with genes expression by univariate and multivariate cox

122   regression analysis and evaluated the influence of hub genes and clinicopathological factors on

123   CC. P-value < 0.05 was set as the cut-off standard.

124   **6. ROC and DCA curve analysis**

125   Receiver Operating Characteristics (ROC) curve analysis was performed based on pROC[31]

126   and ggplot2 (https://ggplot2.tidyverse.org) packages in R software (version 4.0.2) for evaluating

127   the sensitivity and specificity of CDC45 expression in CC diagnosis. The area under curve

128   (AUC) is calculated to assess the veracity and reliability of diagnosis. Decision curve analysis

129   (DCA) is a novel method for evaluating clinical outcome by comparing all-or-none clinical net-

130   benefits[32, 33]. DCA curve was performed using ggDCA (https://CRAN.R-

131   project.org/package=ggDCA) and survival (https://CRAN.R-project.org/package=survival)

132   package in R software (version 4.0.2) based on clinical data from TCGA database.

133   **7. Analysis of CDC45 protein expression level by HPA database**

134   The Human Protein Atlas (HPA) database[34] (https://www.proteinatlas.org/) can spatially

135   localize proteins at the single-cell level and detect more than 90% of the putative protein-coding

136   genes. In the present study, we validate the protein level of CDC45 in normal cervix tissue and

137   cervical cancer tissue by HPA.

138   **8. Relationship between CDC45 expression and tumor-infiltrating immune cells via**

139   **CIBERSORT**

140   CIBERSORT[35] (http://cibersort.stanford.edu/) is a deconvolution algorithm based on gene

141   expression, which estimates the P-value for deconvolution of each sample by Monte Carlo

142   sampling, establishing a measure of confidence in the results. To explore the potential

143   relationship between the CDC45 expression and tumor-infiltrating immune cells in CC, the

144   mRNA expression matrix was standardized and the content of 22 human immune cells was

145   calculated using CIBERSORT. We then divided the CC samples into two groups according to

146   the median value and visualized the data using the vioplot [36]package in R software (version

147   4.0.2). P-value < 0.05 was considered statistically significant.

148   **9. Gene set enrichment analysis**

149   Based on the correlation between gene pathways and CDC45 expression, we used Gene set

150   enrichment analysis[37, 38] (GSEA; https://www.gsea-msigdb.org/gsea /index.jsp) to generated a

151   list of gene classifications and realized graphic visualization.

152   **Results**

153   **1. Screening out DEG in Cervical Cancer**

154   Differentially expressed genes (DEGs) from the three datasets (GSE63514, GSE67522, and

155   GSE39001) were identified using GEO2R standardizing gene microarray on the GEO database.

156   Volcano plot and heatmap analysis showed that gene expression profiles from GSE63514

157   identified 4608 differentially expressed genes with 3053 up-regulated genes and 1555 down-

158   regulated genes in cervical cancer tissues when compared with normal cervical tissues. Gene

159   expression profiles from GSE67522 identified 1327 differentially expressed genes with 588 up-

160   regulated genes and 739 down-regulated genes. Gene expression profiles from GSE39001

161   identified 620 differentially expressed genes with 323 up-regulated genes and 297 down-

162   regulated genes (Fig. 1A and 1B). Overlapping DEGs in the three datasets and plotting the Venn

163   diagram (Fig. 1C) revealed that 83 co-expressed genes, among which 31 were highly expressed

164   and 52 were low-expressed.

165   **2. GO and KEGG pathway enrichment analysis**

166   GO and KEGG pathway enrichment analysis predicted the biological functions of DEGs.

167   KEGG pathway analysis revealed that the DEGs were mainly enriched in DNA replication and

168   cell cycle (Fig. 2A). GO enrichment analysis showed that changes in the biological processes

169   (BP) of DEGs were significantly enriched in DNA replication, DNA-dependent DNA

170   replication, and telomere maintenance via semi-conservative replication (Fig. 2B). Changes in

171   cell component (CC) were mainly involved in condensed chromosome, chromosomal region, and

172   centrosome (Fig. 2C). Genes associated with molecular function (MF) were mainly related to

173   single-stranded DNA-dependent ATPase activity, DNA-dependent ATPase activity, and

174   chemokine receptor binding (Fig. 2D).

175   **3. Construction of PPI network and screening of Hub genes**

176   The PPI network of DEGs was constructed using STRING online database (Fig. 3A). The

177   top 10 genes in the module with protein-protein interactions network were got using Cytoscape

178   (Fig. 3B), including CDC45, RFC4, TOP2A, CCNA2, CCNB2, MCM6, KIF11, KIF20A,

179   UBE2C and FEN1. We hypothesized they are hub genes that may play important roles in the

180   development of CC.

181   **4. Verification and analysis of Hub genes in multiple databases.**

182   In order to validate above viewpoint, we collected mRNA sequencing data in CC from the

183   TCGA and GTEx databases. Gene expression profiles identified 7652 differentially expressed

184   genes with 4221 upregulated genes and 3431 downregulated genes (FDR < 0.05, |logFC| > 1;).

185   The heat map and volcano plot showed the distribution of differentially expressed genes (Fig. 4A

186   and 4B). The survival analysis of 10 key genes showed that CDC45, RFC4 and TOP2A were of

187   statistically significant (P-value < 0.05) (Fig. 5A and Supplementary Fig. 1). Expression analysis

188   by Oncomine database and GEPIA showed that three genes were over-expressed in CC tissues

189   compared with normal tissues (Fig. 5B and 5C), suggesting that the expression of CDC45,

190   RFC4, and TOP2A was associated with prognosis of CC. According to clinical data from the

191   TCGA database, univariate cox proportional regression analysis revealed that clinical stage IV

192   and CDC45 are significantly associated with the development and progression of CC. Clinical

193   stage IV and CDC45 were independent prognostic factors for CC in multivariate cox

194   proportional regression analysis (Table 1 and Fig. 6).

195   **5. ROC and DCA curve analysis on CDC45**

196   Based on CDC45 expression in CC from three datasets of the GEO database (GSE63514,

197   GSE67522, and GSE39001), we performed ROC curve analysis separately to assess its

198   specificity for CC diagnosis. The cut-off values were 10.281 for GSE63514, 335.486 for

199   GSE67522, and 5.362 for GSE39001. The AUC of all three data sets was > 0.8 (Fig. 7A-C),

200   indicating that CDC45 has significant sensitivity and specificity for CC diagnosis. Similarly,

201   DCA curve has also shown that the net benefit of CDC45 exceeds that of the reference model

202   over the entire range of thresholds (Fig.7D). These results suggested that CDC45 could be used

203   as a potential biomarker for the diagnosis of CC.

204   **6. Validation of CDC45 by the GEO database**.

205   To further validate CDC45 expression in CC, we screened a gene microarray data set of CC

206   (GSE52903) from the GEO database for differential expression analysis and plotted volcano and

207   heatmap. The results found that CDC45 expression remained significant (Supplementary Fig. 2A

208   and 2B). Box plots were plotted using the unpaired t-test method by GraphPrism software. ROC

209   curve analysis showed that the cut-off value was 0.721 and the AUC was 0.9. These results

210   suggested that CDC45 was significant in the development of CC (Supplementary Fig. 2C and

211   2D).

212   **7. Protein expression level of CDC45 on the HPA database.**

213   Immunohistochemistry (IHC) staining obtained by the HPA database demonstrated the

214   expression status of the CDC45 and the patient clinical data (Fig. 8). The result showed that the

215   protein expression level of CDC45 was positively correlated with disease status and it was up-

216   regulated in CC tissue, which suggested that the effect of CDC45 that we found as reliable.

217   **8. Correlation analysis between CDC45 expression and tumor-infiltrating immune cells in**

218   **CC.**

219   Previous studies suggest that tumor-infiltrating lymphocytes are independent predictors of

220   sentinel lymph node status and survival in cancers[39]. Therefore, we investigated whether

221   CDC45 expression was associated with tumor-infiltrating immune cells in CC using

222   CIBERSORT. Fig.9A and 9B showed that the relative content distribution of 22 immune cells

223   and the correlation between 22 immune cells in CC. Then, we divided 306 CC samples into high

224   and low groups based on the median value, and calculated the difference and correlation of

225   CDC45 expression in 22 immune cells. The results showed that activated memory CD4+ T cells

226   (R=0.28, P=3.2e-05) and follicular helper T cells (R=0.2, P=0.0026) were positive relation with

227   the expression of CDC45. Naive B cell (R=-0.19, P=0.0044) and resting memory CD4+ T cells

228   (R=-0.49, P=0.0049) were negative relation with the expression of CDC45 (Fig. 9C and 9D).

9. Gene sets enriched in CDC45 expression phenotype

229

230   CDC45 related signaling pathways were analyzed base on GSEA to identify the signaling

231   pathways with significant differences (FDR < 0.05, NOM P-value < 0.05) in GO and KEGG

232   enrichment of the highly expression data sets in CC (Table 2).

233   5 KEGG items including purine metabolism, cell cycle, oocyte meiosis, pyrimidine

234   metabolism, DNA replication were showed significantly differential enrichment in the CDC45

235   high expression phenotype (Fig. 10). GO items results displayed that the biological process of

236   the CDC45 high expression phenotype was mainly enriched in the chromosome, nuclear

237   chromosome, chromosome region, catalytic complex, and microtubule cytoskeleton. The cellular

238   component of the CDC45 high expression phenotype was mostly enriched in the cell cycle, cell

239   cycle process, DNA replication, DNA metabolic process, and cellular response to DNA damage

240   stimulus. The molecular function of the CDC45 high expression phenotype was chiefly enriched

241   in catalytic activity action on DNA, chromatin binding, ubiquitin-like protein binding, ATPase

242   activity, and hydrolase activity action on acid anhydrides.

Discussion

244   Cervical cancer (CC) is the fourth most common malignant tumor in women. It has been

245   reported that there was a high mortality rate owing to cervical cancer worldwide[1]. In recent

246   years, the incidence of cervical cancer is younger and younger, resulting in a shorter life

247   expectancy[40, 41]. Although the clinical therapy of cervical cancer has achieved substantial

248   progress, it remains high in the advanced mortality rate. Therefore, it is an extremely urgent that

249   identify the potential biomarkers for cervical cancer and clinical treatment and prognosis.

250   In our present study, we investigated whether CDC45 has a potentially risk effect on CC.

251   To elucidate the effect of CDC45; we performed a large number of data mining and analysis by

252   some online databases to detect expression levels of CDC45 in CC. We found CDC45 was

253   highly expressed in CC. These results showed that high expression of CDC45 may play a crucial

254   role in the progression and prognosis of CC.

255   Here, we first found that 83 candidate genes of cervical cancer are mainly enriched in DNA

256   replication, cell cycle and cell division by GO and KEGG pathway analysis. This suggested that

257   the development of cervical cancer may be related to abnormal changes in the cell cycle.

258   Survival analysis was performed on the top 10 genes with protein-protein interactions showed

259   that the overall survival (OS) time of CDC45, RFC4, and TOP2A was significantly correlated

260   with the prognosis of cervical cancer. Indeed, we found that the expression of three genes was

261   up-regulated in cervical cancer tissues compared with normal or adjacent tissues through the

262   validation with multiple databases. Furthermore, we performed univariate and multivariate cox

263   regression analysis in combination with the clinical information of the patients. The results

264   showed that the CDC45 and clinical stage IV of cervical cancer were significantly associated

265   with cervical cancer, suggesting that CDC45 and clinical stage IV may be independent risk

266   factors for the development of cervical cancer. The ROC and DCA curve analysis were also

267   consistent with the above results. Moreover, CDC45 is common highly expressed in pan-cancer

268   (Supplementary Fig. 3).

269   In recent years, cancer immunotherapy is a hot topic in cancer treatment[42, 43]. We therefore

270   investigated whether the CDC45 expression was associated with tumor-infiltration immune cells

271   in CC. Interestingly, expression of CDC45 was positively associated with activated memory

272   CD4+ T cells and T follicular helper cells. Previous studies have suggested that dysregulation of

273   memory CD4 + T cells is promoting the progression of malignancy[44, 45]. The crucial role of T

274   follicular helper cells is to help B cells produce antibodies and participate in humoral

275   immunity[46]. Our researches indicated that activated memory CD4+ T cells and T follicular

276   helper cells have better prognostic value in patients with CC consistent with previous results.

277   However, here is a limitation that further studies are needed to illustrate the molecular characters

278   of activated memory CD4 + T cells and T follicular helper cells to explain their prognostic

279   potential.

280   CDC45 can act as a DNA replication initiation factor[47]. It was first proposed in 1997 that

281   the gene has a genetic correlation with the DNA replicators MCM5/CDC46, MCM7/CDC47 and

282   ORC genes previously discovered, and is specifically related to the stability of G1/S mRNA[48-

283   50]. CDC45 is believed to be involved in the development and progression of different tumors and

284   serves as a potential therapeutic target. For instance, Huang et al. found that the low expression

285   of CDC45 can suppress cell proliferation in non-small cell lung cancer (NSCLC), resulting in the

286   cell were stagnated in G2/M phase of cell cycle[51]. This result shows that CDC45 is supporting

287   the carcinogenic effects. Sun et al. found that the expression of CDC45 was up-regulation in

288   papillary thyroid cancer (PTC), which promoted the proliferation of cancer cell in vitro and

289   tumor growth in vivo[52]. To confirm the function of the CDC45 in cervical cancer, we performed

290   single gene enrichment analysis by GSEA. The results showed that purine metabolism, cell

291   cycle, oocyte meiosis, pyrimidine metabolism, and DNA replication in KEGG, chromosome,

292   nuclear chromosome, chromosome region, catalytic complex, and microtubule cytoskeleton in

293   the biological process of GO, cell cycle, cell cycle process, DNA replication, DNA metabolic

294   process and cellular response to DNA damage stimulus in the cell cycle of GO, catalytic activity

295   acting on DNA, chromatin binding, ubiquitin-like protein binding, ATPase activity and

296   hydrolase activity acting on acid anhydrides in the molecular function of GO are significantly

297   enriched in CDC45 high expression phenotype. However, those pathways are no significantly

298   enriched in the CDC45 low expression phenotype (no show). The results indicated that highly

299   expressed CDC45 can be used as a potential biomarker of prognosis and therapeutic target in CC

300   patients. Furthermore, previous studies have also found that CDC45 was indeed associated with

301   prognosis in cervical cancer, which strongly demonstrates the reliability of our results[53].

302   In this paper, we acknowledged our researches are only limited to mining and analysis of

303   the online database without wet experiment verification. For instance, CDC45 expression in CC

304   was examined at the cellular level by the methods of real-time PCR, and MTT, and so on. We

305   strongly recommend further research in this area to increasing evidence for the biological effect

306   of CDC45.

307   **Conclusions**

308   Based on GEO and other multi-database biological big data mining, we found that CDC45

309   can be involved in the development of cervical cancer as an independent prognostic factor. This

310   study provides a new potential target for the clinical diagnosis of cervical cancer. Meanwhile, the

311   relationship between the CDC45 expression and immune-infiltrating cells suggests that

312   immunotherapy may facilitate the treatment of cervical cancer.

313 **Acknowledgements**

314 We acknowledge TCGA and GEO database for providing their platforms and contributors for up

315 loading their meaningful datasets.

316 **References**

317 [1] Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A, Jemal A. 2018. Global cancer statistics 2018:

318   GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer

319   journal for clinicians. **68**: 394-424 DOI: 10.3322/caac.21492.

320 [2] Rajitha B, Malla R R, Vadde R, Kasa P, Prasad G L V, Farran B, Kumari S, Pavitra E, Kamal M A, Raju G

321   S R, Peela S, Nagaraju G P. 2021. Horizons of nanotechnology applications in female specific cancers, in

322   Seminars in cancer biology. **69**: 376-390 DOI: 10.1016/j.semcancer.2019.07.005.

323 [3] Schiffman M, Wentzensen N, Wacholder S, Kinney W, Gage J C, Castle P E. 2011. Human papillomavirus

324   testing in the prevention of cervical cancer. Journal of the National Cancer Institute. **103**: 368-383 DOI:

325   10.1093/jnci/djq562.

326 [4] Lee S-A, Kim J W, Roh J W, Choi J Y, Lee K-M, Yoo K-Y, Song Y S, Kang D. 2004. Genetic

327   polymorphisms of GSTM1, p21, p53 and HPV infection with cervical cancer in Korean women.

328   Gynecologic oncology. **93**: 14-18 DOI: 10.1016/j.ygyno.2003.11.045.

329 [5] Jiao X, Zhang S, Jiao J, Zhang T, Qu W, Muloye G M, Kong B, Zhang Q, Cui B. 2019. Promoter

330   methylation of SEPT9 as a potential biomarker for early detection of cervical cancer and its overexpression

331   predicts radioresistance. Clinical epigenetics. **11**: 120 DOI: 10.1186/s13148-019-0719-9.

332 [6] Wu X, Peng L, Zhang Y, Chen S, Lei Q, Li G, Zhang C. 2019. Identification of Key Genes and Pathways

333   in Cervical Cancer by Bioinformatics Analysis. International journal of medical sciences. **16**: 800-812 DOI:

334   10.7150/ijms.34172.

335 [7] Masai H, You Z, Arai K-i. 2005. Control of DNA replication: regulation and activation of eukaryotic

336   replicative helicase, MCM. IUBMB life. **57**: 323-335 DOI: 10.1080/15216540500092419.

337   [8] Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan M R, Berger J M. 2011. The structural basis

338   for MCM2-7 helicase activation by GINS and Cdc45. Nature structural & molecular biology. **18**: 471-477

339   DOI: 10.1038/nsmb.2004.

340   [9] Simon A C, Sannino V, Costanzo V, Pellegrini L. 2016. Structure of human Cdc45 and implications for

341   CMG helicase function. Nature communications. **7**: 11638 DOI: 10.1038/ncomms11638.

342   [10] Pollok S, Bauerschmidt C, Sänger J, Nasheuer H P, Grosse F. 2007. Human Cdc45 is a proliferation-

343   associated antigen. The FEBS journal. **274**: 3669-3684 DOI: 10.1111/j.1742-4658.2007.05900.x.

344   [11] Edgar R, Domrachev M, Lash A E. 2002. Gene Expression Omnibus: NCBI gene expression and

345   hybridization array data repository. Nucleic acids research. **30**: 207-210 DOI: 10.1093/nar/30.1.207.

346   [12] den Boon J A, Pyeon D, Wang S S, Horswill M, Schiffman M, Sherman M, Zuna R E, Wang Z, Hewitt S

347   M, Pearson R, Schott M, Chung L, He Q, Lambert P, Walker J, Newton M A, Wentzensen N, Ahlquist P.

348   2015. Molecular transitions from papillomavirus infection to cervical precancer and cancer: Role of stromal

349   estrogen receptor signaling. Proceedings of the National Academy of Sciences of the United States of

350   America. **112**: E3255-E3264 DOI: 10.1073/pnas.1509322112.

351   [13] Sharma S, Mandal P, Sadhukhan T, Roy Chowdhury R, Ranjan Mondal N, Chakravarty B, Chatterjee T,

352   Roy S, Sengupta S. 2015. Bridging Links between Long Noncoding RNA HOTAIR and HPV Oncoprotein

353   E7 in Cervical Cancer Pathogenesis. Scientific reports. **5**: 11724 DOI: 10.1038/srep11724.

354   [14] Saha S S, Chowdhury R R, Mondal N R, Roy S, Sengupta S. 2017. Expression signatures of HOX cluster

355   genes in cervical cancer pathogenesis: Impact of human papillomavirus type 16 oncoprotein E7.

356   Oncotarget. **8**: 36591-36602 DOI: 10.18632/oncotarget.16619.

357   [15] Espinosa A M, Alfaro A, Roman-Basaure E, Guardado-Estrada M, Palma Í, Serralde C, Medina I, Juárez

358   E, Bermúdez M, Márquez E, Borges-Ibáñez M, Muñoz-Cortez S, Alcántara-Vázquez A, Alonso P, Curiel-

359   Valdez J, Kofman S, Villegas N, Berumen J. 2013. Mitosis is a source of potential markers for screening

360   and survival and therapeutic targets in cervical cancer. PloS one. **8**: e55975 DOI:

361   10.1371/journal.pone.0055975.

362   [16] Medina-Martinez I, Barrón V, Roman-Bassaure E, Juárez-Torres E, Guardado-Estrada M, Espinosa A M,

363   Bermudez M, Fernández F, Venegas-Vega C, Orozco L, Zenteno E, Kofman S, Berumen J. 2014. Impact of

364   gene dosage on gene expression, biological processes and survival in cervical cancer: a genome-wide

365   follow-up study. PloS one. **9**: e97842 DOI: 10.1371/journal.pone.0097842.

366   [17] Davis S, Meltzer P S. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and

367   BioConductor. Bioinformatics (Oxford, England). **23**: 1846-1847 DOI: 10.1093/bioinformatics/btm254.

368   [18] Huang D W, Sherman B T, Lempicki R A. 2009. Bioinformatics enrichment tools: paths toward the

369   comprehensive functional analysis of large gene lists. Nucleic acids research. **37**: 1-13 DOI:

370   10.1093/nar/gkn923.

371   [19] Huang D W, Sherman B T, Lempicki R A. 2009. Systematic and integrative analysis of large gene lists

372   using DAVID bioinformatics resources. Nature protocols. **4**: 44-57 DOI: 10.1038/nprot.2008.211.

373   [20] Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S

374   S, Eppig J T, Harris M A, Hill D P, Issel-Tarver L, Kasarskis A, Lewis S, Matese J C, Richardson J E,

375   Ringwald M, Rubin G M, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene

376   Ontology Consortium. Nature genetics. **25**: 25-29 DOI: 10.1038/75556.

377   [21] Kanehisa M. 2002. The KEGG database. Novartis Foundation symposium. **247**: 91-252.

378   [22] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P,

379   von Mering C, Jensen L J. 2013. STRING v9.1: protein-protein interaction networks, with increased

380   coverage and integration. Nucleic acids research. **41**: D808-D815 DOI: 10.1093/nar/gks1094.

381   [23] Smoot M E, Ono K, Ruscheinski J, Wang P-L, Ideker T. 2011. Cytoscape 2.8: new features for data

382   integration and network visualization. Bioinformatics (Oxford, England). **27**: 431-432 DOI:

383   10.1093/bioinformatics/btq675.

384   [24] Zhang Z, Li H, Jiang S, Li R, Li W, Chen H, Bo X. 2019. A survey and evaluation of Web-based

385   tools/databases for variant analysis of TCGA data. Briefings in bioinformatics. **20**: 1524-1541 DOI:

386   10.1093/bib/bby023.

387   [25] Battle A, Brown C D, Engelhardt B E, Montgomery S B. 2017. Genetic effects on gene expression across

388   human tissues. Nature. **550**: 204-213 DOI: 10.1038/nature24277.

389 [26] Lindbom L, Ribbing J, Jonsson E N. 2004. Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM

390 related programming. Computer methods and programs in biomedicine. **75**: 85-94 DOI:

391 10.1016/j.cmpb.2003.11.003.

392 [27] Robinson M D, McCarthy D J, Smyth G K. 2010. edgeR: a Bioconductor package for differential

393 expression analysis of digital gene expression data. Bioinformatics (Oxford, England). **26**: 139-140 DOI:

394 10.1093/bioinformatics/btp616.

395 [28] McCarthy D J, Chen Y, Smyth G K. 2012. Differential expression analysis of multifactor RNA-Seq

396 experiments with respect to biological variation. Nucleic acids research. **40**: 4288-4297 DOI:

397 10.1093/nar/gks042.

398 [29] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. 2017. GEPIA: a web server for cancer and normal gene

399 expression profiling and interactive analyses. Nucleic acids research. **45**(W1): W98–W102 DOI:

400 10.1093/nar/gkx247.

401 [30] Rhodes D R, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan

402 A M. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia

403 (New York, N.Y.). **6**: 1-6 DOI: 10.1016/s1476-5586(04)80047-2.

404 [31] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an open-

405 source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics. **12**: 77 DOI:

406 10.1186/1471-2105-12-77.

407 [32] Vickers A J, Elkin E B. 2006. Decision curve analysis: a novel method for evaluating prediction models.

408 Medical decision making : an international journal of the Society for Medical Decision Making. **26**: 565-

409 574 DOI: 10.1177/0272989X06295361.

410 [33] Van Calster B, Wynants L, Verbeek J F M, Verbakel J Y, Christodoulou E, Vickers A J, Roobol M J,

411 Steyerberg E W. 2018. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators.

412 European urology. **74**: 796-804 DOI: 10.1016/j.eururo.2018.08.038.

413 [34] Uhlén M, Fagerberg L, Hallström B M, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C,

414 Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto C A-K, Odeberg J,

415 Djureinovic D, Takanen J O, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J,

416   Nilsson P, Schwenk J M, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M,

417   von Heijne G, Nielsen J, Pontén F. 2015. Proteomics. Tissue-based map of the human proteome. Science

418   (New York, N.Y.). **347**: 1260419 DOI: 10.1126/science.1260419.

419   [35] Gentles A J, Newman A M, Liu C L, Bratman S V, Feng W, Kim D, Nair V S, Xu Y, Khuong A, Hoang

420   C D, Diehn M, West R B, Plevritis S K, Alizadeh A A. 2015. The prognostic landscape of genes and

421   infiltrating immune cells across human cancers. Nature medicine. **21**: 938-945 DOI: 10.1038/nm.3909.

422   [36] Hu K. 2020. Become Competent within One Day in Generating Boxplots and Violin Plots for a Novice

423   without Prior R Experience. Methods and protocols. **3**: 64 DOI: 10.3390/mps3040064.

424   [37] Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov J P, 2007, GSEA-P: a desktop application for

425   Gene Set Enrichment Analysis. Bioinformatics (Oxford, England). **23**: 3251-3253 DOI:

426   10.1093/bioinformatics/btm369.

427   [38] Subramanian A, Tamayo P, Mootha V K, Mukherjee S, Ebert B L, Gillette M A, Paulovich A, Pomeroy S

428   L, Golub T R, Lander E S, Mesirov J P. 2005. Gene set enrichment analysis: a knowledge-based approach

429   for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the

430   United States of America. **102**: 15545-15550 DOI: 10.1073/pnas.0506580102.

431   [39] Azimi F, Scolyer R A, Rumcheva P, Moncrieff M, Murali R, McCarthy S W, Saw R P, Thompson J F.

432   2012. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and

433   survival in patients with cutaneous melanoma. Journal of clinical oncology : official journal of the

434   American Society of Clinical Oncology. **30**: 2678-2683 DOI: 10.1200/JCO.2011.37.8539.

435   [40] Li H, Wu X, Cheng X. 2016. Advances in diagnosis and treatment of metastatic cervical cancer. Journal

436   of gynecologic oncology. **27**: e43 DOI: 10.3802/jgo.2016.27.e43.

437   [41] Kong Y, Zong L, Yang J, Wu M, Xiang Y. 2019. Cervical cancer in women aged 25 years or younger: a

438   retrospective study. Cancer management and research. **11**: 2051-2058 DOI: 10.2147/CMAR.S195098.

439   [42] Bader J E, Voss K, Rathmell J C. 2020. Targeting Metabolism to Improve the Tumor Microenvironment

440   for Cancer Immunotherapy. Molecular cell. **78**: 1019-1033 DOI: 10.1016/j.molcel.2020.05.034.

441   [43] Yang Y. 2015. Cancer immunotherapy: harnessing the immune system to battle cancer. The Journal of

442   clinical investigation. **125**: 3335-3337 DOI: 10.1172/JCI83871.

443 [44] Gasper D J, Tejera M M, Suresh M. 2014. CD4 T-cell memory generation and maintenance. Critical

444   reviews in immunology. **34**: 121-146 DOI: 10.1615/critrevimmunol.2014010373

445 [45] MacLeod M K L, Clambey E T, Kappler J W, Marrack P. 2009. CD4 memory T cells: what are they and

446   what can they do? Seminars in immunology. **21**: 53-61 DOI: 10.1016/j.smim.2009.02.006.

447 [46] Crotty S. 2019. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. Immunity. **50**:

448   1132-1148 DOI: 10.1016/j.immuni.2019.04.011.

449 [47] Hennessy K M, Lee A, Chen E, Botstein D. 1991. A group of interacting yeast DNA replication genes.

450   Genes & development. **5**: 958-969 DOI: 10.1101/gad.5.6.958.

451 [48] Zou L, Mitchell J, Stillman B. 1997. CDC45, a novel yeast gene that functions with the origin recognition

452   complex and Mcm proteins in initiation of DNA replication. Molecular and cellular biology. **17**: 553-563

453   DOI: 10.1128/mcb.17.2.553.

454 [49] Hardy C F. 1997. Identification of Cdc45p, an essential factor required for DNA replication. Gene. **187**:

455   239-246 DOI: 10.1016/s0378-1119(96)00761-5.

456 [50] Hopwood B, Dalton S. 1996. Cdc45p assembles into a complex with Cdc46p/Mcm5p, is required for

457   minichromosome maintenance, and is essential for chromosomal DNA replication. Proceedings of the

458   National Academy of Sciences of the United States of America. **93**: 12309-12314 DOI:

459   10.1073/pnas.93.22.12309.

460 [51] Huang J, Li Y, Lu Z, Che Y, Sun S, Mao S, Lei Y, Zang R, Li N, Zheng S, Liu C, Wang X, Sun N, He J.

461   2019. Analysis of functional hub genes identifies CDC45 as an oncogene in non-small cell lung cancer - a

462   short report. Cellular oncology (Dordrecht). **42**: 571-578 DOI: 10.1007/s13402-019-00438-y.

463 [52] Sun J, Shi R, Zhao S, Li X, Lu S, Bu H, Ma X. 2017. Cell division cycle 45 promotes papillary thyroid

464   cancer progression via regulating cell cycle. Tumour biology : the journal of the International Society for

465   Oncodevelopmental Biology and Medicine. **39**: 1010428317705342 DOI: 10.1177/1010428317705342.

466 [53] Qiu H-Z, Huang J, Xiang C-C, Li R, Zuo E-D, Zhang Y, Shan L, Cheng X. 2020. Screening and

467   Discovery of New Potential Biomarkers and Small Molecule Drugs for Cervical Cancer: A Bioinformatics

468   Analysis. Technology in cancer research & treatment. **19**: 1533033820980112 DOI:

469   10.1177/1533033820980112.

470

Figure 1

Identification of differentially expressed genes in CC based on the GEO database.

(A) Volcano plot of the expression level of differentially expressed genes in normal and cancer tissues from GSE63514, GSE67522 and GSE39001. Yellow dots represent a high expression of genes and blue dots represent a low expression of genes. (B) Heatmap of the expression level of differential expressed genes between normal and cancer tissues from the three data sets. The abscissa indicates the sample names, and the ordinate shows the gene names. High expression of genes is shown in red and low expression of genes is shown in blue. LFC stands for log Fold Change. DEGs were defined with FDR < 0.05 (-log10 FDR > 1.301) and |logFC| > 1. (C) The 3 datasets showed an overlap of 83 genes using Venn diagram.

A

B

C

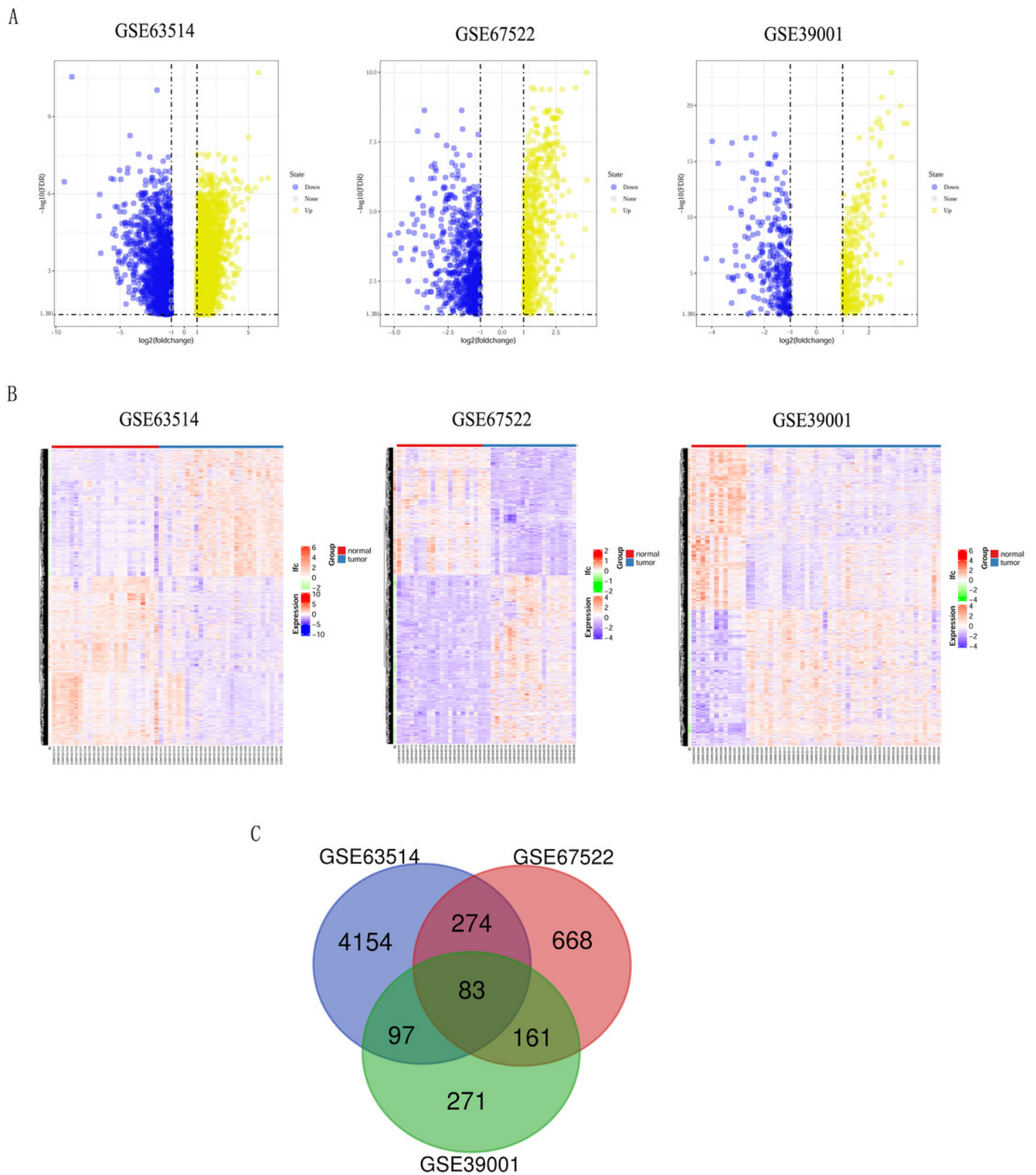



Figure 2

Significantly functional enrichment pathway of 83 DEGs.

(A) KEGG pathway enrichment analysis. (B-D) the top 10 terms significantly enriched in the three GO categories: (B) biological process; (C) cellular component and (D) molecular function. P <0.05 was set as the threshold.

A

All pathway_KEGG

B

Top10 geneontology_Biological_Process

C

Top10 geneontology_Cellular_Component

D

Top10 geneontology_Molecular_Function

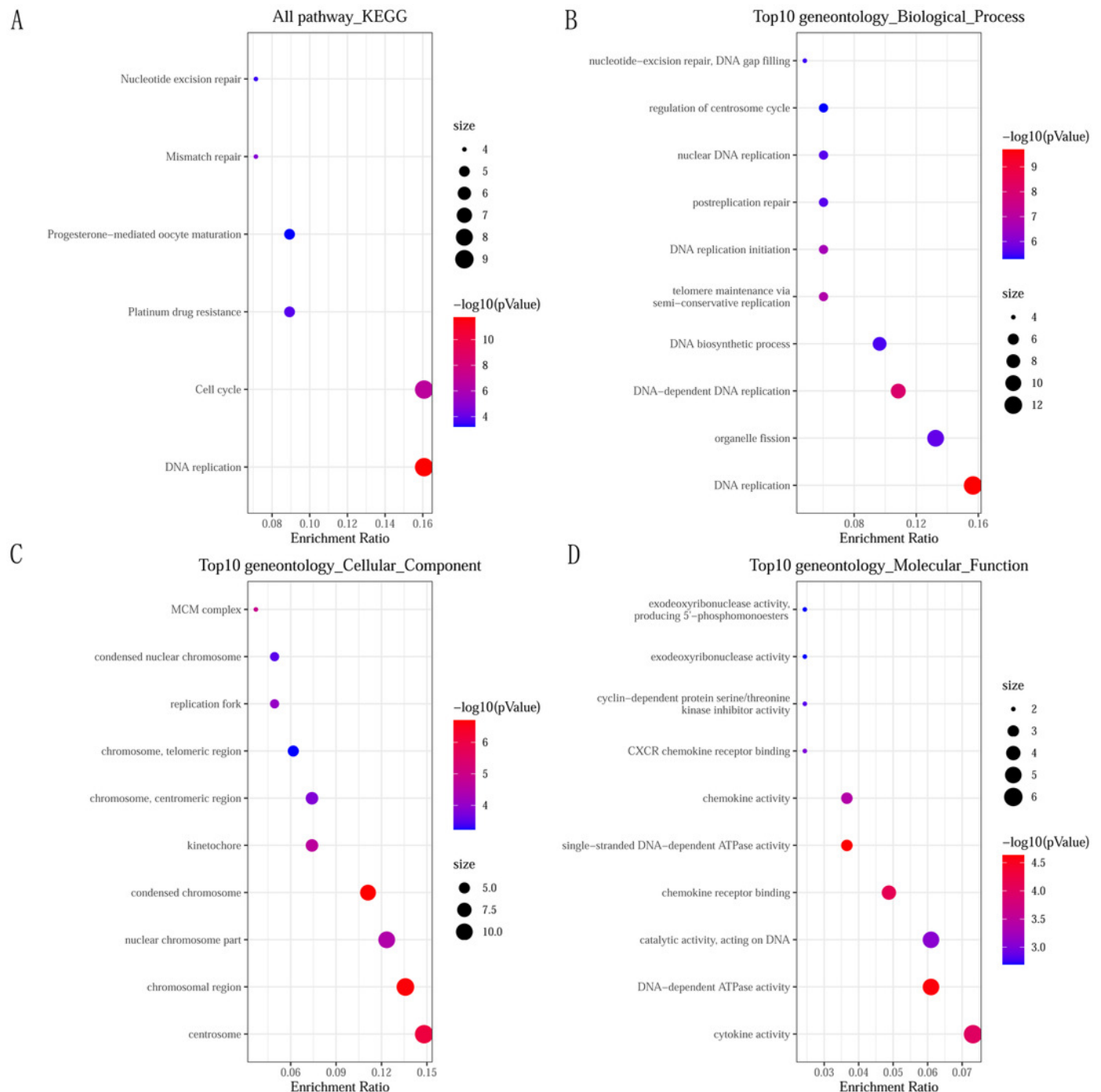



Figure 3

PPI network and the most significant module of DEGs.

(A) The PPI network of DEGs was constructed using Cytoscape. (B) The most significant module was obtained from PPI network with 10 nodes and 200 edges.

Figure 4

Identification of differentially expressed genes in CC based on TCGA and GTEx databases.

(A) Volcano plot of the expression level of differentially expressed mRNAs in CC and adjacent normal tissues. Yellow dots represent a high expression of genes, black dots represent a normal expression of genes and blue dots represent a low expression of genes. (B) Heatmap of expression level of differentially expressed mRNAs between CC and adjacent normal tissues. The abscissa indicates the sample names, and the ordinate shows the gene names. Red represents high expression, and green represents low expression. DEGs were defined with FDR < 0.05 (-log10 FDR > 1.301) and |logFC| >1.

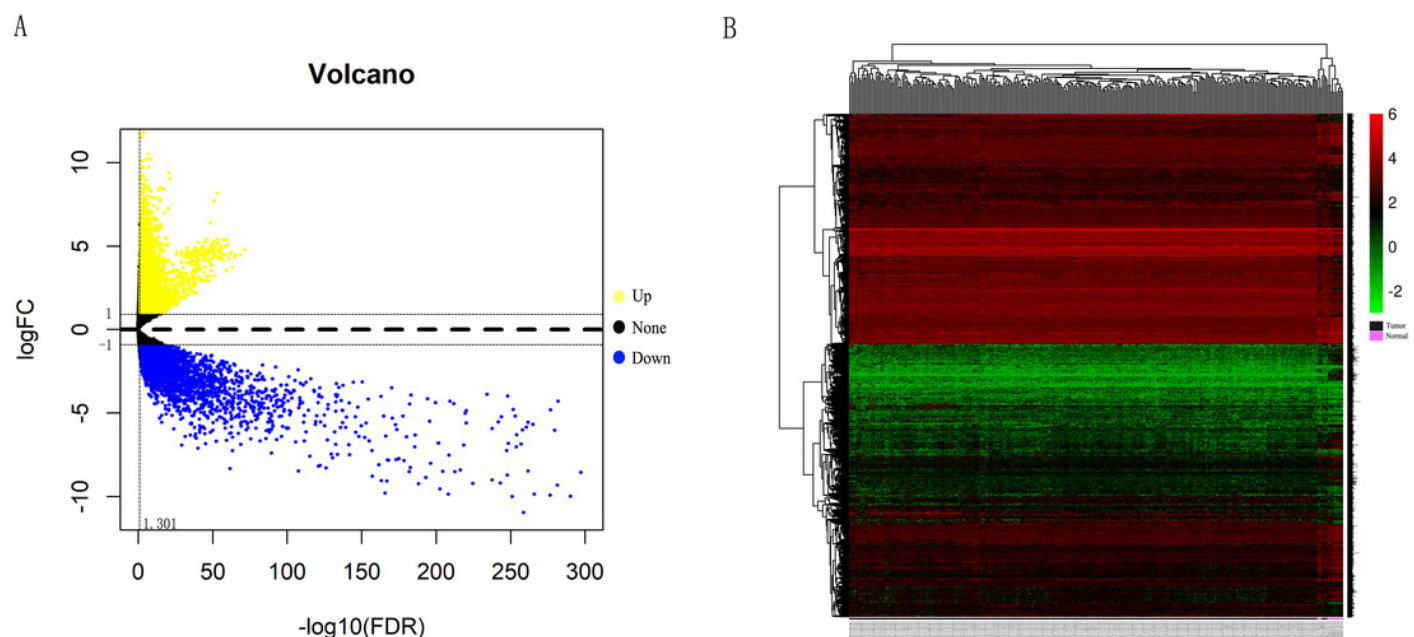



Figure 5

Overall survival analysis and expression of hub genes in normal and cancer tissues.

(A) Survival curves analysis for CDC45, RFC4, and TOP2A. (B) The transcription levels of CDC45, RFC4 and TOP2A in the normal and cancer tissues (Oncomine). (C) Differential expression of CDC45, RFC4 and TOP2A in the normal and cancer tissues (GEPIA).

A

CDC45 · RFC4 · TOP2A

B

CDC45 · RFC4 · TOP2A

C

CDC45 · RFC4 · TOP2A

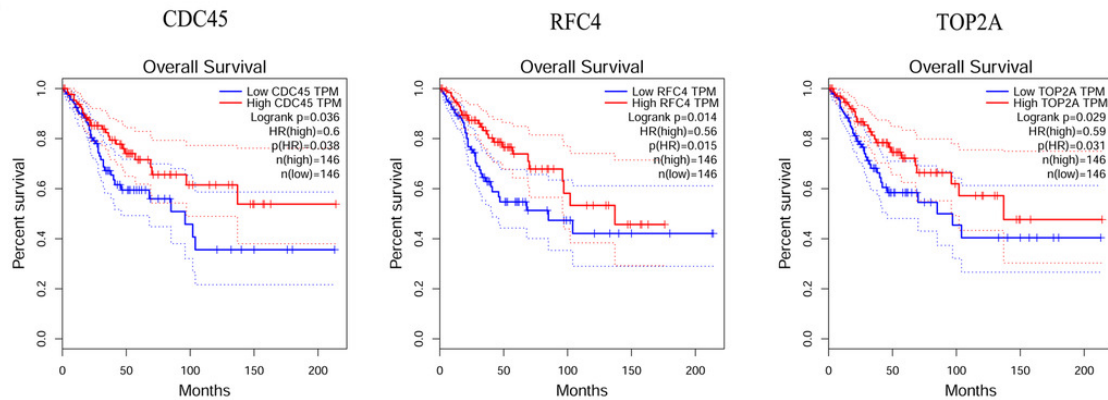
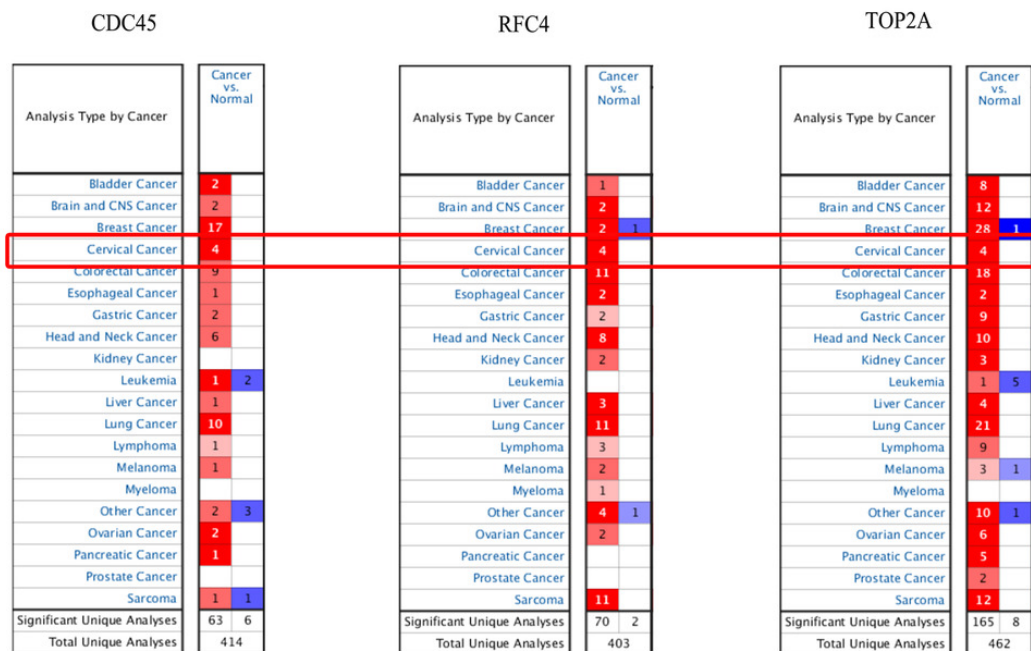
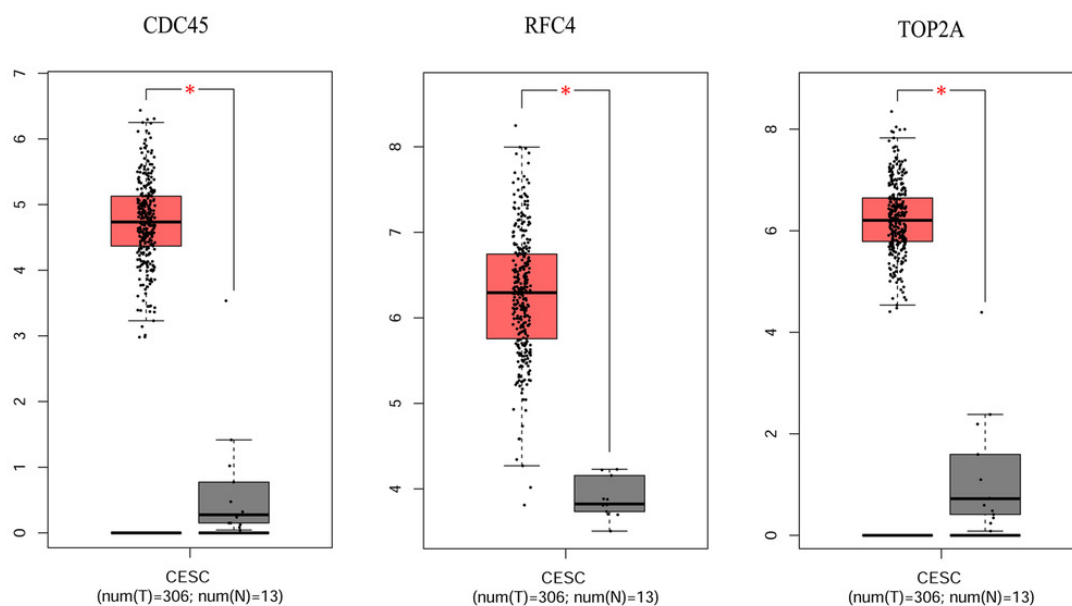



Figure 6

Multivariate Cox analysis of CDC45 expression and clinical stage.

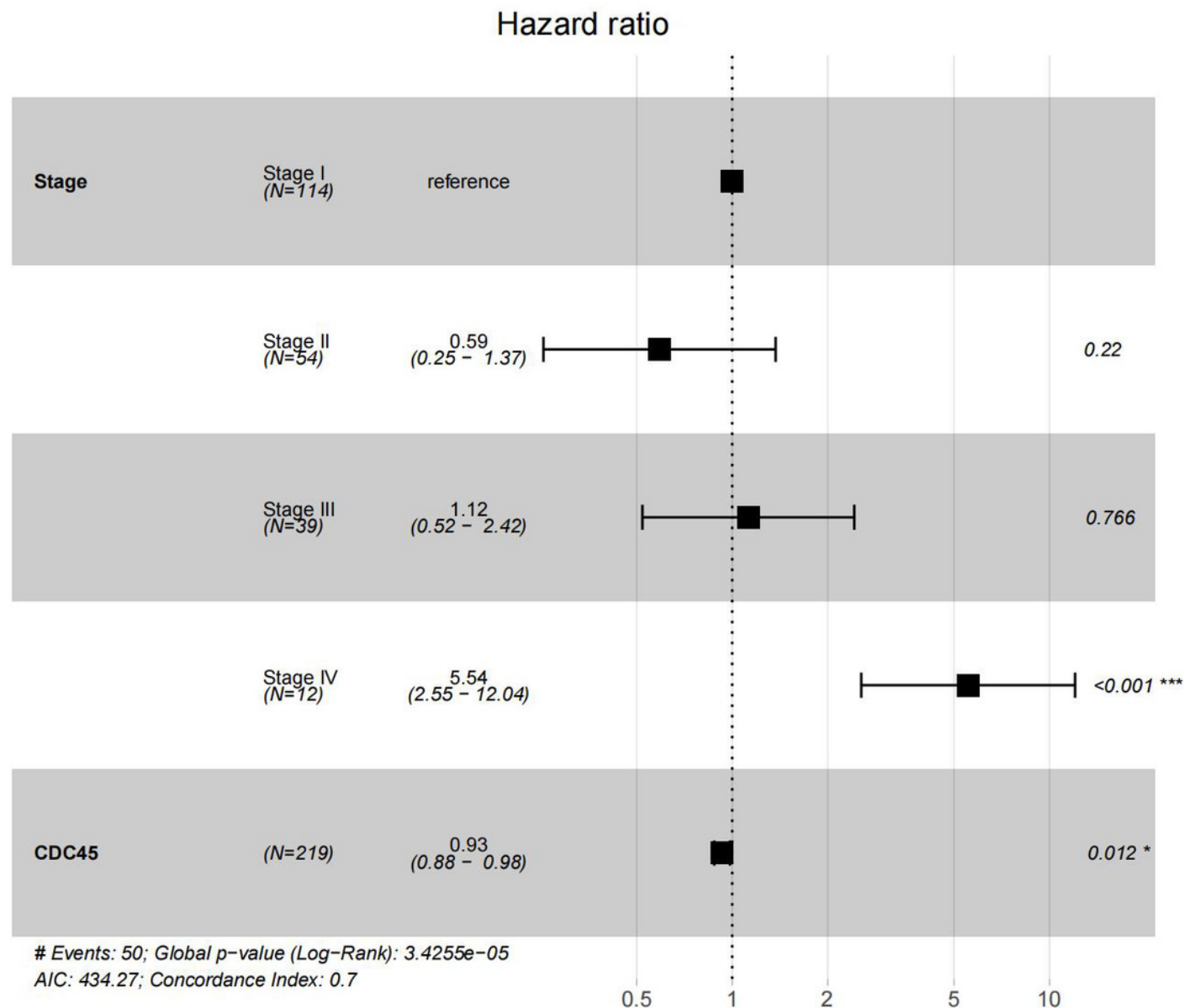



Figure 7

ROC and DCA curve analysis.

(A-C) ROC curve analysis of CDC45 on the GEO database (GSE63514, GSE67522 and GSE39001). (D) The decision curve analysis (DCA) shows the net benefit of CDC45 in the 3-year survival. The abscissa represented the threshold probabilities, and the ordinate measured the net benefit.

A

Sensitivity (TPR) vs 1−Specificity (FPR)

GSE63514
AUC: 0.845
CI: 0.737−0.954

B

Sensitivity (TPR) vs 1−Specificity (FPR)

GSE67522
AUC: 0.945
CI: 0.874−1.000

C

Sensitivity (TPR) vs 1−Specificity (FPR)

GSE39001
AUC: 0.975
CI: 0.941−1.000

D

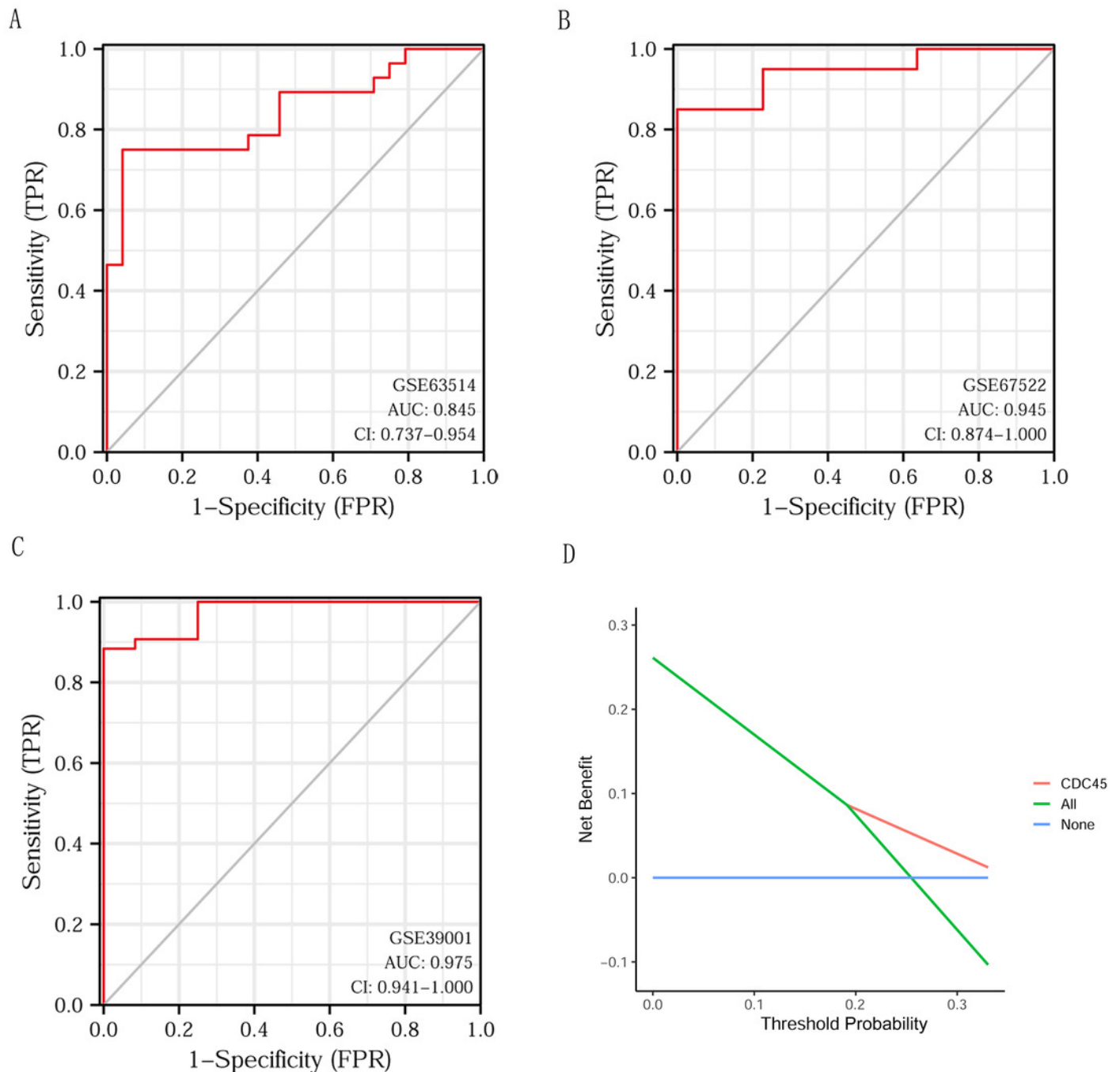
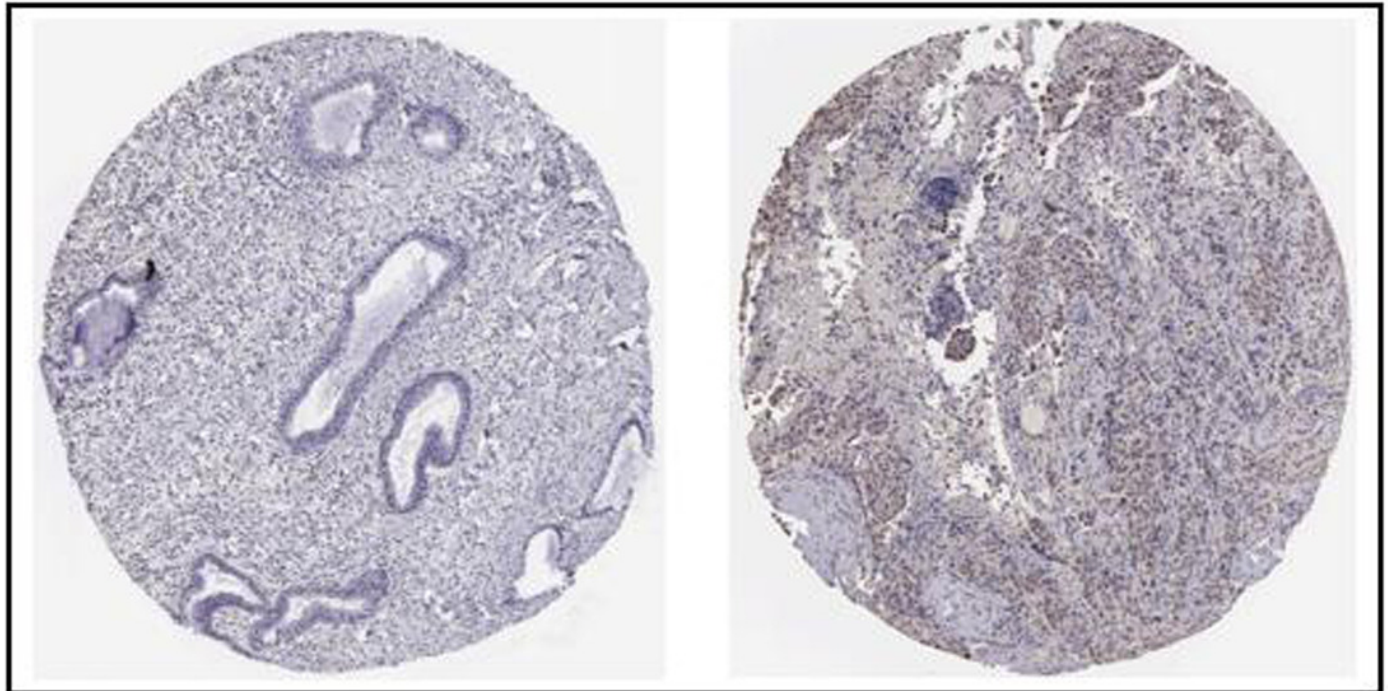
Net Benefit vs Threshold Probability

CDC45
All
None

Figure 8

Analysis of the protein expression level of CDC45 in CC by the Human Protein Atlas (HPA) database.

Cervix, uterine
Patient ID: 1657
Sex: female
Age: 70
Stain: Not detected
Intensity: weak
Quantity: <25%

CC
Patient ID: 926
Sex: female
Age: 54
Stain: Low
Intensity: weak
Quantity: >75%

Figure 9

The relationship between the CDC45 expression and tumor-infiltration immune cells.

(A) Barplot showed the relative content of 22 immune cells in CC samples. (B) Block diagram showed the correlation of 22 immune cells in CC. (C) Violin diagram showed the difference of CDC45 expression in 22 immune cells. High expression groups are indicated in yellow and low expression groups in blue. (D) Scatterplot showed the correlation between CDC45 and immune cells.

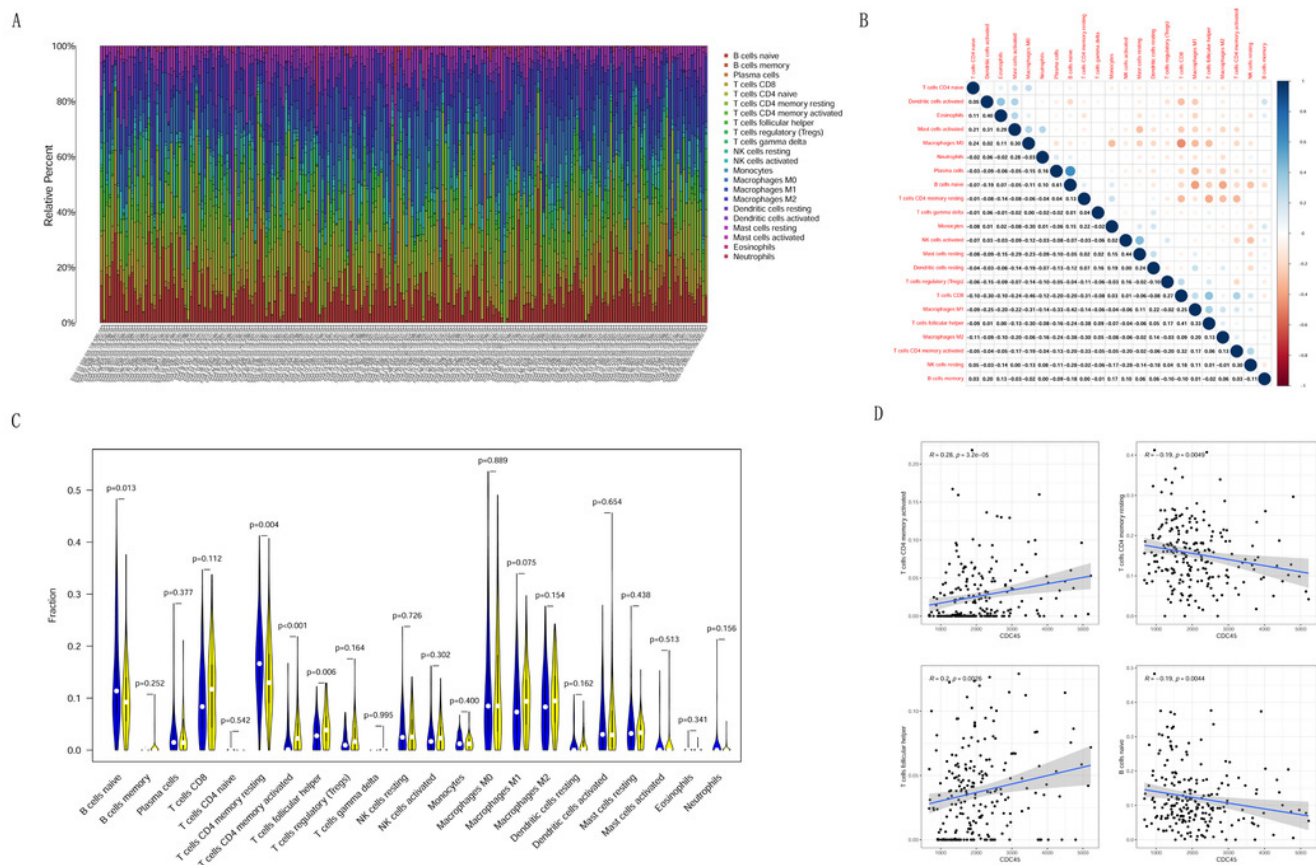
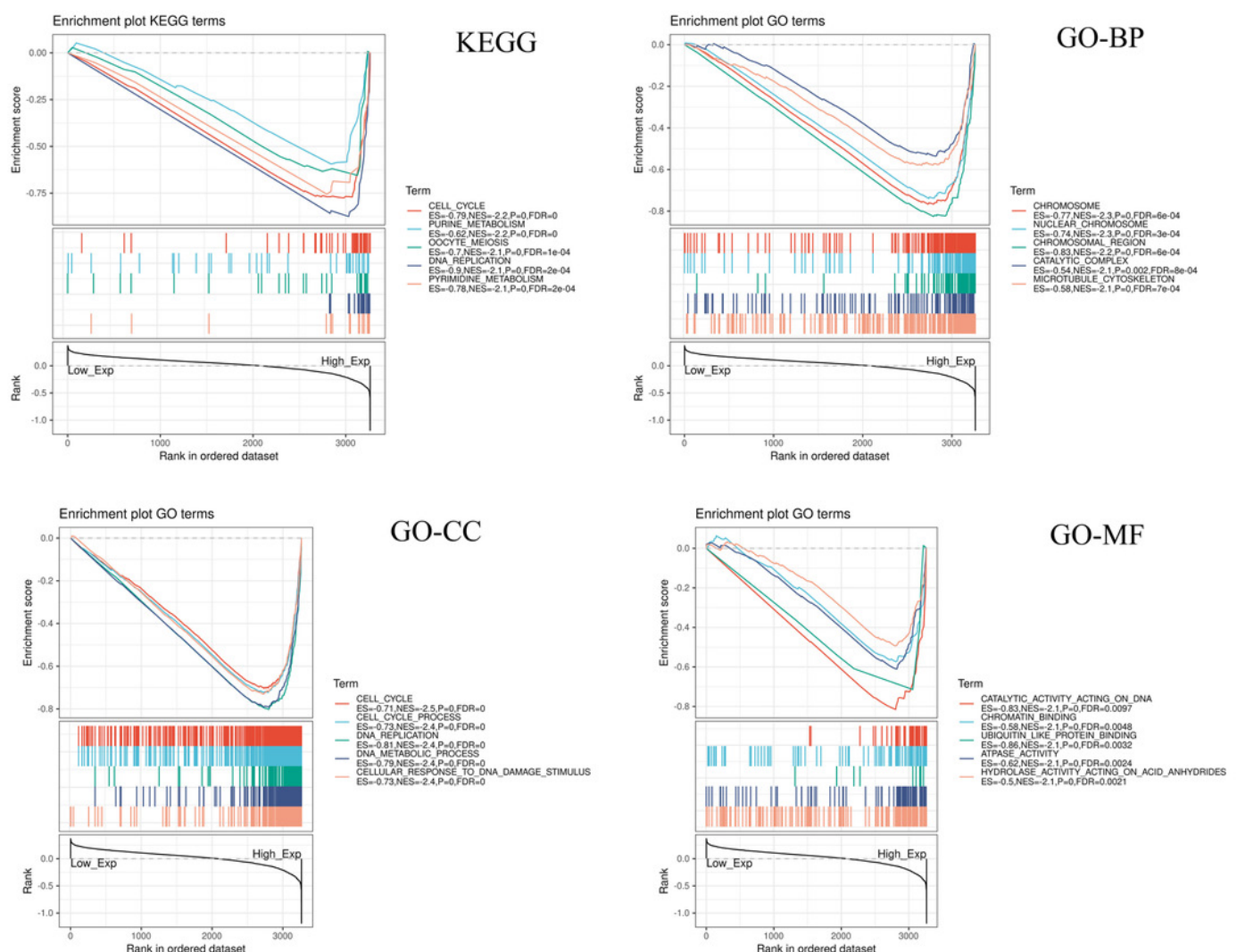



Figure 10

Enrichment plots from gene set enrichment analysis (GSEA).

Differential enrichment of gene in KEGG, GO-BP, GO-CC and GO-MF pathways with high CDC45 expression. (KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene ontology; BP: biological process; CC: cellular component; MF: molecular function.)

Association with overall survival and clinicopathologic characteristic in TCGA patients using Cox regression analysis.

Clinical Characteristics	HR(95%CI)	P-val
UniCOX		
Age	1.02(1.00-1.04)	0.134
Stage I	Ref.	
Stage II	0.66(0.28-1.51)	0.318
StageIII	1.24(0.58-2.66)	0.582
StageIV	6.22(2.87-13.48)	<0.001***
Grade1	Ref.	
Grade2	1.02(0.24-4.34)	0.973
Grade3	1.00(0.23-4.30)	0.997
CDC45	0.93(0.88-0.98)	0.008**
RFC4	0.99(0.96-1.01)	0.203
TOP2A	1.00(0.99-1.02)	0.598
MultiCOX		
Stage I	Ref.	
Stage II	0.59(0.25-1.37)	0.220
StageIII	1.12(0.52-2.42)	0.766
StageIV	5.54(2.55-12.04)	<0.001***
CDC45	0.93(0.88-0.98)	0.012*

2 HR: Hazard Ratio; Ref: Reference group.

Gene sets enriched in phenotype.

Gene set name	NES	NOM	FDR q-val
KEGG			
KEGG_PURINE_METABOLISM	-2.178	0	6.14E-04
KEGG_CELL_CYCLE	-2.158	0	3.07E-04
KEGG_OOCYTE_MEIOSIS	-2.101	0	5.53E-04
KEGG_PYRIMIDINE_METABOLISM	-2.075	0.002	0.001
KEGG_DNA_REPLICATION	-2.067	0	8.28E-04
GO_BP			
GO_CHROMOSOME	-2.321	0	5.57E-04
GO_NUCLEAR_CHROMOSOME	-2.301	0	2.79E-04
GO_CHROMOSOMAL_REGION	-2.184	0	5.57E-04
GO_CATALYTIC_COMPLEX	-2.145	0.002	8.18E-04
GO_MICROTUBULE_CYTOSKELETON	-2.140	0	6.54E-04
GO_CC			
GO_CELL_CYCLE	-2.460	0	0
GO_CELL_CYCLE_PROCESS	-2.409	0	0
GO_DNA_REPLICATION	-2.375	0	0
GO_DNA_METABOLIC_PROCESS	-2.372	0	0
GO_CELLULAR_RESPONSE_TO_DNA_DAMAGE_STIMU	-2.369	0	0
GO_MF			
GO_CATALYTIC_ACTIVITY_ACTING_ON_DNA	-2.097	0	0.010
GO_CHROMATIN_BINDING	-2.093	0	0.005
GO_UBIQUITIN_LIKE_PROTEIN_BINDING	-2.092	0	0.003
GO_ATPASE_ACTIVITY	-2.091	0	0.002
GO_HYDROLASE_ACTIVITY_ACTING_ON_ACID_ANHY	-2.085	0	0.002