

Identification of candidate biomarkers associated with the diagnosis and prognosis of endometrial cancer: a bioinformatics analysis

Changqiang Wei

The First Affiliated Hospital of Guangxi Medical University

Le Huang

Shenzhen Maternity and Child Healthcare Hospital Affiliated to Southern Medical University

Lingjie Deng

The First Affiliated Hospital of Guangxi Medical University

Huisi Lin

The First Affiliated Hospital of Guangxi Medical University

Weicheng Pan

The First Affiliated Hospital of Guangxi Medical University

Lihong Pang (✉ panglihong@stu.gxmu.edu.cn)

The First Affiliated Hospital of Guangxi Medical University <https://orcid.org/0000-0002-5179-8396>

Research article

Keywords: WGCNA, MTHFD2, KIF4A, TPX2, RPS6KA6, SIX1, downregulated genes

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-318398/v1>

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

[Read Full License](#)

Abstract

Background

Endometrial cancer (EC) is a common malignancy tumor that seriously threatens the wellbeing and health of women. This study aimed to map the hub genes and potential pathways that may be involved in EC.

Methods

In our study, three gene expression profiles of GEO and EC data from the TCGA database were analyzed. We performed WGCNA, and Cox regression analyses to screen hub genes and further validated them by using HPA, KM plots, and other databases. We also studied the methylation level by UCSC Xena and mutation of hub genes using Cbioportal .

Results

We identified 363 differentially expressed genes (DEGs) (146 upregulated and 217 downregulated genes). Pathway analysis revealed that hub genes are mainly related to cell cycle, DNA damage response, EMT, hormone ER, RAS/MAPK, and PI3K/AKT. Finally, we identified five hub genes that are related to the progression and prognosis of EC and screened several relevant small-molecule drugs.

Conclusions

MTHFD2, KIF4A, TPX2, RPS6KA6, and SIX1 were identified as candidate biomarkers for further basic and clinical research on endometrial cancer.

Introduction

Endometrial cancer (EC) is a common malignancy tumor in the female reproductive system that seriously threatens the wellbeing and health of women. The age-standardized incidence rate of EC is steadily increasing. A previous study reported that approximately 63,400 new cases and 21,800 deaths from EC occurred in China in 2015[1]. In the United States, about 65,620 new cases and 12,590 deaths from EC were estimated in 2020, representing 3.6% of all new cancer cases[2]. The early treatment of EC involves surgery, followed by radiotherapy and/or chemotherapy, which provide favorable prognosis.

Unfortunately, chemotherapy remains the main treatment option for advanced stages, and most cases are currently incurable[3]. Therefore, effective diagnostic markers, prognostic factors, and therapeutic targets must be identified for the successful treatment of patients with EC. Owing to experimental constraints, bioinformatics analyses have become a valuable tool for early diagnosis and prognosis of cancer. Bioinformatics analyses combined with databases are increasingly utilized to screen biomarkers

for various diseases[4, 5]. WGCNA, which was developed by Langfelder and Horvath, is a method widely used in finding genes with similar expressions[6]. For example, Luo et al.[7] identified nine key genes in renal cell carcinoma by using WGCNA and other analytic tools. In the present study, we performed differential expression analysis, WGCNA, and Cox regression analyses to screen hub genes. We further validated five hub genes by using HPA, KM plots, and other databases. We studied the methylation level and mutation of the five hub genes and screened several relevant small-molecule drugs.

Materials And Methods

Research design and data processing

The EC gene expression profiles were downloaded from the GEO database (<http://ncbi.nlm.nih.gov/geo/>). GSE17025, which was derived from the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), included 91 EC tissue samples and 12 normal endometrial tissue samples. GSE115810 and GSE36389 were based on the GPL96 platform (Affymetrix Human Genome U133A Array). The former included 24 EC tissue samples and 3 normal endometrial tissue samples, whereas the latter included 13 EC tissue samples and 7 normal endometrial tissue samples. The raw data were merged and then adjusted by batches by using the “sva” R package (version 3.6.3). The “limma” package was used to screen the DEGs between the EC and normal tissue samples. P value was adjusted to <0.05, and $|\log_{2}FC$ (fold change)|>1 was set as the screening index. The EC mRNA expression profile data and clinical data were also obtained from the TCGA database (<https://cancergenome.nih.gov/>). The data included 23 normal tissues and 552 cancer tissues. Differential expression analysis was performed by Wilcox test. Genes with P value < 0.05 and $|\log_{2}FC$ (fold change)| > 1 were selected as DEGs.

Functional enrichment analysis

GO terms and KEGG pathway enrichment analysis of the DEGs were conducted using the “clusterProfiler” package[8]. P<0.05 was set as a standard for significant enrichment.

Weighted gene coexpression network analysis

The “WGCNA” package was used to construct a coexpression network. $\beta = 4$ was selected as the appropriate soft threshold (scale $R^2 = 0.92$) to construct a scale-free coexpression network. The matrix was converted into adjacency and TOM matrices. The dissimilarity of the TOM matrix (dissTOM) was calculated, and at least 30 genes were clustered into different modules by using a dynamic tree-cutting method. Eigengene was calculated and abline was set to 0.25 to merge similar modules on the cluster tree. Finally, the clinical information was combined with the modules to find the modules most relevant to EC.

Screening and validation of hub genes

Genes that play important roles in cancer often have abnormal expressions. Therefore, the DEGs in the green module were selected as the candidate hub genes for further analysis. UALCAN (<http://ualcan.path.uab.edu/index.html>) was used to query the mRNA and protein expression of these genes to screened the real DEGs[9]. Univariable and multivariable Cox proportional hazard regression analyses were performed to screen the hub genes and other factors that are related to the prognosis of EC. Immunohistochemical images were obtained from HPA (<http://proteinatlas.org>) to validate the protein expression of the hub genes[10]. Finally, ROC curves were plotted using the “survivalROC” package to evaluate the capacity of the hub genes for distinguishing tumors from normal tissues[11].

Survival analysis

Survival analysis was performed using the Kaplan–Meier plotter (<http://kmplot.com/analysis/>) to estimate the correlations between the hub genes and cancer prognosis[12]. Information about the OS and RFS of patients with EC with gene expression can be easily obtained from the Kaplan–Meier plotter. TISIDB (<http://cis.hku.hk/TISIDB/>) was used to evaluate the prognostic value of the expression of the hub genes in other tumor types, as well as explore the relationship between the hub genes and their OS [13].

Hub gene analysis

GSCA Lite (<http://www.bioinfo.life.hust.edu.cn/web/GSCALite/>) and GSEA were used to investigate further the molecular mechanisms of the hub genes in EC.

GSCA Lite (<http://bioinfo.life.hust.edu.cn/web/GSCALite/>) is a network analysis platform for genomic cancer analysis. It can be used to calculate the score for 10 cancer-related pathways in 32 cancer types on the basis of the RPPA data from TCPA[14].

GSEA is a comprehensive analysis method for genome-wide expression profile chip data [15]. In this study, EC samples from the TCGA database were divided into two groups according to the risk score and then utilized for GSEA using GSEA 4.0.3 software(<http://software.broadinstitute.org/gsea/index.jsp>). `.all.v7.1.symbols.gmt[HARMARKS]`, and `c2.cp.kegg.v7.0.symbols.gmt[KEGG]` was set as the gene set database. Each analysis process was performed 1,000 times by using the permutation test. P values less than 0.05 and FDR less than 0.25 were considered to be statistically significant.

Genetic mutation and methylation of the hub genes

The cBioPortal for Cancer Genomics (<http://cbioportal.org/>) is a web-based cancer genomics database based on TCGA[16]. It can be used to explore, visualize, and analyze multidimensional cancer genomic data. In the present study, it was used to explore genetic mutations connected with the five hub genes and the effects of these mutations on prognosis. mRNA expression z-scores (RNASeq V2 RSEM) with a z-score threshold was set as ± 2.0 . Kaplan–Meier curves were plotted to show the relationship between the mutations of the five genes with the OS and DFS of EC. UCSC Xena (<http://xena.ucsc.edu/>) is an online tool for multi-omic research that integrates data from TCGA, ICGC, GTEx, and other databases. This tool also includes clinical and phenotype data[17]. This database was used to study the methylation of the

five hub genes and analyze their differential methylation sites. The correlation between the mRNA expression of the hub genes and the significantly different methylation sites was explored.

Potential drug prediction

To explore the potential small-molecule drugs related to EC, we uploaded the DEG list to the CMap database for analysis.

CMap (<https://portals.broadinstitute.org/cmap/>) is a database used in studying gene expression and drugs. This database can help researchers use data on gene expression profiles to compare drugs with high correlation with diseases[18].

PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) was used to retrieve the chemical structures of the related small-molecule drugs.

Result

Differentially expressed genes of EC

The flowchart of the study is depicted in Figure 1. We analyzed three integrated GEO expression profiles, including 128 cancer samples and 22 normal samples. A total of 501 DEGs (187 upregulated and 314 downregulated genes) were screened. In TCGA, we screened 7394 DEGs containing 4505 upregulated and 2889 downregulated genes from EC samples. Of the DEGs from GEO and TCGA databases, 146 upregulated genes and 217 downregulated genes were identified. The results of DEGs are shown in the volcano plots in (Figures 2A and 2B).

Gene Ontology and KEGG pathway enrichment

GO analysis revealed that the DEGs were mainly enriched in organelle fission, nuclear division, chromosome segregation, extracellular structure organization, and cell cycle, etc (Figure 2C). KEGG pathway analysis showed that the DEGs were related to microRNAs in cancer, cell cycle, cellular senescence, chemokine signaling pathway, IL-17 signaling pathway, p53 signaling pathway, and some disease-associated pathways (Figure 2D).

WGCNA results

We selected 150 samples and 12,402 genes from GEO for WGCNA and drew a sample dendrogram and trait heat map (Figure 3A). $\beta = 4$ ($R^2 = 0.92$) was selected to construct a scale-free network (Figures 3B and 3C). On the basis of the TOM matrix, the modules were hierarchically clustered, and similar modules on the cluster tree were merged (Figure 4A). Twenty coexpression modules were obtained, and the correlation between these modules with clinical information was calculated (Figure 4B). The green module, which contained 649 genes, was the most related to clinical information and thus selected for further analysis. Correlation analysis between module membership and gene significance was performed for the green module. Results suggested the existence of a close relationship between them (Figure 4C).

Identification and validation of hub genes

A total of 65 upregulated and 36 downregulated genes were included in the green module (Figures 4D and 4E). A total of 64 real DEGs were obtained using UALCAN. These DEGs were evaluated via univariate Cox regression analyses in the TCGA cohort. For the top 15 genes correlated with the OS of EC patients according to the results of univariate analyses, multivariate Cox regression analysis was performed. Cox regression analysis successfully screened MTHFD2, KIF4A, TPX2, RPS6KA6, and SIX1, and these hub genes were used to establish a prediction model for the prognosis of EC. Risk scores based on gene expression and regression coefficient were calculated using the following formula: Risk score = $MTHFD2 \times 0.36349 + KIF4A \times (-0.51299) + TPX2 \times 0.46623 + RPS6KA6 \times 0.54220 + SIX1 \times 0.20801$. According to the risk scores, the patients were divided into high-risk and low-risk groups. Kaplan–Meier curves showed that the high-risk group had significantly poorer OS than the low-risk group (Figure. 5A) The AUC value of the ROC curve was 0.70285 for the 5-year survival, indicating that both the high-risk and low-risk groups have a diagnostic value for the prognosis of patients with EC (Figure 5B). The expression of the five hub genes in different groups is depicted in a heat map (Figure 5C). In UALCAN, the expression of mRNA and protein levels markedly differed (Figures 6A–6K), and their mRNA expression levels not only correlated with sample type but also with cancer stages (Figures 6K–6O). Based on the HPA database, the protein expression levels of MTHFD2, TPX2, SIX1, and KIF4A were significantly higher in tumor tissues than in normal tissues, but no significant difference was found in the protein expression level of RPS6KA6 (Figures 7A–7E).

The ROC curve was analyzed using the dataset from the GEO database. Results indicated that the five hub genes have a good diagnostic value for discriminating tumor and normal tissues (Figures 8A–8E). Cox regression analysis suggested that age, grade, stage, type, and risk score were independent prognostic factors for patients with EC (Table 1).

Survival analysis

K–M curves from the Kaplan–Meier plotter database showed that the expression levels of the five hub genes were negatively associated with the prognosis of patients with EC. High expression levels of these hub genes resulted in low OS and RFS, except for the RFS of SIX1 (Figures 9A–9J). The results of TISIDB are summarized in Supplementary Figure1-5. High mRNA expression levels of KIF4A and TPX2 led to the poor prognosis of UCEC, ACC, KIRP, LGG, LUAD, and MESO (Supplementary Figures 1 and 2). Interestingly, the high expression of RPS6KA6 was not only associated with the poor prognosis of UCEC and CESC but also related to the good prognosis of KIRC and LGG (Supplementary Figure 3). The results of MTHFD2 and SIX1 were similar to those of RPS6KA6; high MTHFD2 expression was negatively correlated with the OS of several tumors, such as UCEC, HNDC, KIRC, and KIRP but positively correlated with GBM and LGG (Supplementary Figure 4). High mRNA expression of SIX1 led to a good prognosis of patients with LUAD and LUSC but a bad prognosis of patients with UCEC, ACC, KIRC, and SARC , etc (Supplementary Figure 5).

Hub gene analysis

In the GSCALite website, we found that these five hub genes are mainly related to the activation of several

pathways, such as cell cycle, DNA damage response, and EMT. They are also involved in inhibiting various pathways, such as hormone ER, RAS/MAPK, and PI3K/AKT (Figures 11A and 11B). A total of 100 significant genes with positive or negative correlation were identified by GSEA. Interestingly, the KEGG pathways enriched in the high-risk group according to GSEA enrichment analysis were similar to those from the KEGG enrichment analysis of DEGs by clusterProfiler package. Overall, the pathways were highly associated with cell cycle, p53 signaling pathway, DNA replication, oocyte meiosis, and several cancer pathways (Figures 12A–12E). The expression profiles of 100 significant genes are shown in the heat map in Figure 12F. GSEA with the hallmark gene sets suggested that the most relevant pathways included mitotic spindle, G2M checkpoint, spermatogenesis, E2F targets, MYC targets V1, MYC targets V2, UV response, cholesterol homeostasis, mtorc1 signaling, and heme metabolism (Figures 12A–12E). The abnormal expression of these genes may have an important relationship to the disorder of these pathways and may influence the occurrence and development of EC.

Mutation and methylation of the hub genes

The genetic mutations of the five hub genes (TCGA and Firehose Legacy) were analyzed using the cBioPortal. Results showed that 61 out of 177 patients had mutations (34%). Among them, TPX2 and TSPYL5 had the highest mutation rates (19% and 12%, respectively) (Figure 13A). The KM curves showed that the patients with the mutations of the five hub genes had a lower overall survival rate and a shorter disease-free survival time than those without the mutations (Figures 13B and 13C).

Using UCSC Xena, we analyzed the methylation and its effects on the expression of the five hub genes in the TCGA cohort. Several sites had a significantly different methylation level in EC tissues compared with in normal tissues. Further analysis suggested that the methylation of these sites were closely related to the gene expression of the hub genes. The methylation levels of MTHFD2, KIF4A, and TPX2 were lower in the tumor samples than in the normal samples, whereas the overall methylation level of RPS6KA6 in the tumor samples was relatively high. The methylation levels were negatively correlated with the mRNA expression of these genes (Supplementary Figures 6–9). However, the methylation of SIX1 was different. In the 21 methylation sites affecting gene expression, 6 of them had a lower methylation status in the tumor tissues, whereas the other 15 showed a higher methylation status. Correlation analysis revealed that these six sites were negatively correlated with SIX1 expression unlike the other 15 sites. These results were consistent with the high expression level of SIX1 (Supplementary Figure 10).

Small-molecule drugs related to EC

We screened a total of 153 small-molecule drugs by using CMap. The top 10 drugs most relevant to EC are shown in Table 2. Among these small-molecule drugs, trichostatin A, resveratrol, and vorinostat showed a strong negative correlation with EC. The chemical structures of these drugs were obtained from the PubChem database (Figure 14).

Discussion

EC is a serious disease that leads to patient mortality. Abnormal vaginal bleeding is the main clinical manifestation of the early stage of EC. It may take a long period from the onset of symptoms to confirm EC. A previous study reported that only 9% of women with postmenopausal bleeding was diagnosed with EC[19]. The 5-year survival rate of patients with localized EC is about 95%, whereas it is 17.3% for patients with distant metastasis[2]. Although considerable progress has been made in EC research, the mortality rate of EC has been increasing worldwide in recent decades. According to the US Surveillance, Epidemiology, and End Results database, the death rate was 4.21% in 1992 and 5.02% in 2017 in the United States. Unfortunately, no new therapeutic approaches for EC have been developed[20]. Consequently, screening candidate genes is warranted for the early diagnosis and treatment of patients with EC.

In the present study, three gene expression datasets from the GEO and mRNA profile data of EC were obtained. A total of 363 DEGs were screened. MTHFD2, KIF4A, TPX2, RPS6KA6, and SIX1 were screened as the candidate biomarkers. The risk scores based on these five hub genes indicated a high predictive value for EC prognosis. The ROC curves showed the ability of these hub genes to distinguish tumor tissues from normal tissues, suggesting that these hub genes play an important role in the development and progression of EC.

MTHFD2 is an enzyme of folate metabolism that is a common target of cancer chemotherapeutics. Even during growth restriction, MTHFD2 overexpression has the capacity to promote cell proliferation[21]. Significant upregulation of MTHFD2 has been reported in various cancers, including CRC, HCC, breast cancer, and NSCLC[22–25]. In CRC, silencing MTHFD2 expression inhibits proliferation, weakens migration ability, blocks cell cycle, and promotes the apoptosis of cancer cells. Meanwhile, high MTHFD2 expression results in bad prognosis[22, 23]. MTHFD2 overexpression is associated with poor prognosis and tumor metastasis in HCC and breast cancer[24, 25]. In vitro and in vivo experiments of NSCLC indicated that knockdown of MTHFD2 can reduce cell growth and tumorigenicity[26]. Kawai et al.[27] reported that DS18561882, an MTHFD2 inhibitor, has a strong antitumor effect on mouse xenograft model. In the present study, the mRNA and protein expressions of MTHFD2 were higher in EC tissues than in normal tissues. Moreover, mRNA expression was associated with the patients' cancer stages. Furthermore, MTHFD2 was an independent prognostic factor for the short OS of patients with EC, indicating that MTHFD2 participates in the progression of EC.

KIF4A overexpression has been recently observed in various types of cancers, including HCC, OSCCs, and CRC[28–30]. Moreover, these studies reported that KIF4A overexpression promotes cell proliferation throughout the entire cell cycle. The level of KIF4A overexpression is negatively related to survival time but significantly correlated with clinical stage, metastasis, and tumor dimension[28]. Yasuyuki et al. [29] reported that cell-cycle is arrested at the G2/M phase and cell proliferation is significantly decreased in KIF4A knockdown cells in OSCCs. A study demonstrated that KIF4A promotes the proliferation of CRC via the p21-mediated cell cycle and results in a poor prognosis[30]. These results are consistent with those of our study and suggest that KIF4A may be a key regulator in the tumor progression of EC.

TPX2, formerly known as P100, is known to be a key factor for mitosis and spindle assembly[31]. A review showed that TPX2 is considered a candidate oncogene because of its overexpression in numerous malignancies that correlate with the progression of diseases and unfavorable prognosis[32]. Yan et al. [33] found that the mRNA and protein levels of TPX are upregulated in bladder carcinoma. Moreover, they proved that TPX2 modulates cell proliferation, migration, invasion, and tumorigenicity via a TPX2-p53-GLIPR1 regulatory circuitry. In OC, TPX2 is upregulated, and TPX knockdown affects cell cycle and decreases proliferation[34]. Tomii et al.[35] reported that high TPX2 expression in GC is associated with a decrease in DFS and RFS. A study reported that miR-491 exerts inhibitory effects on BC cell invasion and migration by targeting TPX2[36]. Zhang et al.[37] proved that TPX2 is regulated by miR-29a-5p overexpression in EC tissues and that it is associated with poor outcome. These results agreed with the findings of the present study that TPX2 plays an important role in EC.

RPS6KA6, also known as RPS4, is considered a tumor suppressor. It shows a decreased expression in several cancers, including ovarian cancer, breast cancer, and GC[38]. A previous study reported that RPS6KA6 expression may have the capacity to limit the oncogenic, invasive, and metastatic potential of breast cancer[39]. In vitro and in vivo experiments suggested that RPS6KA6 has inhibitory effects on GC[40]. Dewdney et al.[41] found that RPS6KA6 frequently shows a higher methylation status and lower expression in EC compared with normal endometrial tissues. Their results were similar with those of the present study. We further explored the mutation, prognosis, and cancer stage of RPS6KA6 in EC and found that RPS6KA6 is an independent risk factor of OS in EC.

A previous review showed that SIX1 is overexpressed and associated with poor prognosis in numerous cancers[42]. SIX1 is the most well-studied hub gene in EC. Li et al.[43] reported that miR-23a regulates cell proliferation, migration, and invasion by targeting SIX1 in EC. Xin et al.[44] found that SIX1 is involved in the growth of EC via the ERK and AKT signaling cascades. Suen et al.[45] showed that the SIX1 oncoprotein is a biomarker of EC. In the present study, we found that SIX1 was not only aberrantly expressed but also associated with the cancer stage and overall survival of patients with EC.

The results KEGG pathway, GSEA, and GSCA Lite analyses showed that cell cycle had a tight association with EC. Kastan and Bartek[46] indicated that the disorder of the cell cycle pathway is a vital mechanism of tumorigenesis. Another study suggested that targeting cell cycle checkpoints may contribute in identifying new therapeutic targets [47]. Therefore, we speculate that the five hub genes may play an important role in the tumorigenesis and progression of EC by regulating the cell cycle. This conjecture provides a new direction to understand further the molecular mechanism of EC.

The results of mutation analysis showed varying degrees of mutation of the five hub genes. These mutations led to the poor prognosis of patients with EC. DNA methylation results can explain the abnormal expression of the five hub genes in EC from an epigenetic perspective. We screened several small-molecule drugs that may have potential therapeutic effects on EC on the basis of the CMap database. However, the present study has several limitations. First, all data were derived from public

databases. We did not explore the molecular mechanism of EC through experiments. Thus, more clinical data and in vitro and in vivo experiments are required to validate our results.

Conclusions

In Conclusion, we identified five hub genes that are related to the progression and prognosis of EC. Among them, MTHFD2 was reported for the first time in EC and has a close relationship with the occurrence and development of EC patients. We hope that more research will be conducted in the future to explore his role in endometrial cancer. We established a risk prediction model that provides a favorable prognostic value on the basis of the five hub genes and identified several candidate drugs that may have potential treatment effects on EC.

Abbreviations

EC: endometrial cancer; WGCNA: weighted gene co-expression network analysis; DEGs: differentially expressed genes; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; HPA: Human Protein Atlas; ROC: receiver operating characteristic; OS: overall survival; RFS: relapse-free survival; DFS: disease-free survival; TISIDB: tumor-immune system interactions database; GSEA: Gene Set Enrichment Analysis; FDR: false discovery rate; CMap: Connectivity map; UCEC: Uterine Corpus Endometrial Carcinoma; ACC: Adrenocortical Carcinoma; KIRP: Kidney renal papillary cell carcinoma; LGG: Brain Lower Grade Glioma; LUAD: Lung adenocarcinoma; MESO: Mesothelioma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; HNDC: Head and Neck squamous cell carcinoma; KIRC: Kidney renal clear cell carcinoma; LUSC: Lung squamous cell carcinoma; SARC: Sarcoma; CRC: colorectal cancer; HCC: hepatocellular carcinoma; NSCLC: non-small cell lung cancer; OSCCs: oral squamous cell carcinomas;

OC: ovarian cancer; GC: gastric cancer.

Declarations

Availability of data and materials

The datasets supporting the conclusion of this article are included within the article.

Acknowledgements

Not applicable

Funding

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81660256 and Nos. 81960281)

Author information

Affiliations

The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China

Changqiang Wei, Lingjie Deng, Huisi Lin, Weicheng Pan, Lihong Pang

Shenzhen Maternity and Child Healthcare Hospital Affiliated to Southern Medical University, Shenzhen, China.

Le Huang

Contributions

Changqiang Wei conceived and designed the study. Le Huang, Lingjie Deng and Huisi Lin performed the statistical analyses. Changqiang Wei and Weicheng Pan contributed in collecting and interpreting data. Lihong Pang revised the final manuscript. All authors approved the final version of the manuscript and are accountable for the accuracy and integrity in all aspects of the study.

Corresponding author

Correspondence to Lihong Pang.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is

not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA: a cancer journal for clinicians*. 2016;66:115-32.
2. Bethesda, MD. SEER Cancer Stat Facts: Uterine Cancer. National Cancer Institute. <https://seer.cancer.gov/statfacts/html/corp.html>
3. Brooks RA, Fleming GF, Lastra RR, Lee NK, Moroney JW, Son CH, Tatebe K, Veneris JL. Current recommendations and recent progress in endometrial cancer. *CA: a cancer journal for clinicians*. 2019;69:258-79.
4. Zhang JN, Qin X, Liu XG, Ma XN, Huang ZP, Zhang W, Zhang CH, Yuan SX, Zhou WP. WITHDARWN: Identification of abnormally expressed genes and their roles in invasion and migration of hepatocellular carcinoma. *Aging*. 2020;12.
5. Kumar SU, Kumar DT, Siva R, Doss CGP, Zayed H. Integrative Bioinformatics Approaches to Map Potential Novel Genes and Pathways Involved in Ovarian Cancer. *Frontiers in bioengineering and biotechnology*. 2019;7:391.
6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9:559.
7. Luo Y, Shen D, Chen L, Wang G, Liu X, Qian K, Xiao Y, Wang X, Ju L. Identification of 9 key genes and small molecule drugs in clear cell renal cell carcinoma. *Aging*. 2019;11:6029-52.
8. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012;16:284-7.
9. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. 2017;19:649-58.
10. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, et al. A subcellular map of the human proteome. *Science (New York, NY)*. 2017;356.
11. Patrick J. Heagerty. survivalROC: Time-dependent ROC curve estimation from censored survival data. 2013.
12. Nagy Á, Lániczky A, Menyhárt O, Győrffy B. Author Correction: Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Scientific reports*. 2018;8:11515.

13. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW, Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* (Oxford, England). 2019;35:4200-2.
14. Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. *Bioinformatics* (Oxford, England). 2018;34:3771-2.
15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:15545-50.
16. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. 2012;2:401-4.
17. Mary Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N. Brooks, Jingchun Zhu, David Haussler. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. 2019:326470.
18. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* (New York, NY). 2006;313:1929-35.
19. Matteson KA, Robison K, Jacoby VL. Opportunities for Early Detection of Endometrial Cancer in Women With Postmenopausal Bleeding. *JAMA internal medicine*. 2018;178:1222-3.
20. MacKay HJ, Freixinos VR, Fleming GF. Therapeutic Targets and Opportunities in Endometrial Cancer: Update on Endocrine Therapy and Nonimmunotherapy Targeted Options. *American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting*. 2020;40:1-11.
21. Gustafsson Sheppard N, Jarl L, Mahadessian D, Strittmatter L, Schmidt A, Madhusudan N, Tegnér J, Lundberg EK, Asplund A, Jain M, Nilsson R. The folate-coupled enzyme MTHFD2 is a nuclear protein and promotes cell proliferation. *Scientific reports*. 2015;5:15029.
22. Wei Y, Liu P, Li Q, Du J, Chen Y, Wang Y, Shi H, Wang Y, Zhang H, Xue W, Gao Y, Li D, Feng Y, et al. The effect of MTHFD2 on the proliferation and migration of colorectal cancer cell lines. *OncoTargets and therapy*. 2019;12:6361-70.
23. Ju HQ, Lu YX, Chen DL, Zuo ZX, Liu ZX, Wu QN, Mo HY, Wang ZX, Wang DS, Pu HY, Zeng ZL, Li B, Xie D, et al. Modulation of Redox Homeostasis by Inhibition of MTHFD2 in Colorectal Cancer: Mechanisms and Therapeutic Implications. *Journal of the National Cancer Institute*. 2019;111:584-96.
24. Liu X, Huang Y, Jiang C, Ou H, Guo B, Liao H, Li X, Yang D. Methylenetetrahydrofolate dehydrogenase 2 overexpression is associated with tumor aggressiveness and poor prognosis in hepatocellular

- carcinoma. *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*. 2016;48:953-60.
25. Liu F, Liu Y, He C, Tao L, He X, Song H, Zhang G. Increased MTHFD2 expression is associated with poor prognosis in breast cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2014;35:8685-90.
 26. Yu C, Yang L, Cai M, Zhou F, Xiao S, Li Y, Wan T, Cheng D, Wang L, Zhao C, Huang X. Down-regulation of MTHFD2 inhibits NSCLC progression by suppressing cycle-related genes. *Journal of cellular and molecular medicine*. 2020;24:1568-77.
 27. Kawai J, Toki T, Ota M, Inoue H, Takata Y, Asahi T, Suzuki M, Shimada T, Ono K, Suzuki K, Takaishi S, Ohki H, Matsui S, et al. Discovery of a Potent, Selective, and Orally Available MTHFD2 Inhibitor (DS18561882) with in Vivo Antitumor Activity. *Journal of medicinal chemistry*. 2019;62:10204-20.
 28. Hou G, Dong C, Dong Z, Liu G, Xu H, Chen L, Liu L, Wang H, Zhou W. Upregulate KIF4A Enhances Proliferation, Invasion of Hepatocellular Carcinoma and Indicates poor prognosis Across Human Cancer Types. *Scientific reports*. 2017;7:4148.
 29. Minakawa Y, Kasamatsu A, Koike H, Higo M, Nakashima D, Kouzu Y, Sakamoto Y, Ogawara K, Shiiba M, Tanzawa H, Uzawa K. Kinesin family member 4A: a potential predictor for progression of human oral cancer. *PloS one*. 2013;8:e85951.
 30. Hou PF, Jiang T, Chen F, Shi PC, Li HQ, Bai J, Song J. KIF4A facilitates cell proliferation via induction of p21-mediated cell cycle progression and promotes metastasis in colorectal cancer. *Cell death & disease*. 2018;9:477.
 31. Heidebrecht HJ, Buck F, Steinmann J, Sprenger R, Wacker HH, Parwaresch R. p100: a novel proliferation-associated nuclear protein specifically restricted to cell cycle phases S, G2, and M. *Blood*. 1997;90:226-33.
 32. Neumayer G, Belzil C, Gruss OJ, Nguyen MD. TPX2: of spindle assembly, DNA damage response, and cancer. *Cellular and molecular life sciences : CMLS*. 2014;71:3027-47.
 33. Yan L, Li Q, Yang J, Qiao B. TPX2-p53-GLIPR1 regulatory circuitry in cell proliferation, invasion, and tumor growth of bladder cancer. *Journal of cellular biochemistry*. 2018;119:1791-803.
 34. Ma S, Rong X, Gao F, Yang Y, Wei L. TPX2 promotes cell proliferation and migration via PLK1 in OC. *Cancer biomarkers : section A of Disease markers*. 2018;22:443-51.
 35. Tomii C, Inokuchi M, Takagi Y, Ishikawa T, Otsuki S, Uetake H, Kojima K, Kawano T. TPX2 expression is associated with poor survival in gastric cancer. *World journal of surgical oncology*. 2017;15:14.
 36. Tan GZ, Li M, Tan X, Shi ML, Mou K. MiR-491 suppresses migration and invasion via directly targeting TPX2 in breast cancer. *European review for medical and pharmacological sciences*. 2019;23:9996-10004.
 37. Jiang T, Sui D, You D, Yao S, Zhang L, Wang Y, Zhao J, Zhang Y. MiR-29a-5p inhibits proliferation and invasion and induces apoptosis in endometrial carcinoma via targeting TPX2. *Cell cycle (Georgetown, Tex)*. 2018;17:1268-78.

38. Arechavaleta-Velasco F, Zeferino-Toquero M, Estrada-Moscoso I, Imani-Razavi FS, Olivares A, Perez-Juarez CE, Diaz-Cueto L. Ribosomal S6 kinase 4 (RSK4) expression in ovarian tumors and its regulation by antineoplastic drugs in ovarian cancer cell lines. *Medical oncology (Northwood, London, England)*. 2016;33:11.
39. Thakur A, Sun Y, Bollig A, Wu J, Biliran H, Banerjee S, Sarkar FH, Liao DJ. Anti-invasive and antimetastatic activities of ribosomal protein S6 kinase 4 in breast cancer cells. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2008;14:4427-36.
40. Hu C, Dai J, Lin X, Meng Y, Liang H. Effect of RSK4 on Biological Characteristics of Gastric Cancer. *Cancer management and research*. 2020;12:611-9.
41. Dewdney SB, Rimel BJ, Thaker PH, Thompson DM, Jr., Schmidt A, Huettner P, Mutch DG, Gao F, Goodfellow PJ. Aberrant methylation of the X-linked ribosomal S6 kinase RPS6KA6 (RSK4) in endometrial cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011;17:2120-9.
42. Blevins MA, Towers CG, Patrick AN, Zhao R, Ford HL. The SIX1-EYA transcriptional complex as a therapeutic target in cancer. *Expert opinion on therapeutic targets*. 2015;19:213-25.
43. Li HL, Sun JJ, Ma H, Liu SJ, Li N, Guo SJ, Shi Y, Xu YY, Qi ZY, Wang YQ, Wang F, Guo RM, Liu D, et al. MicroRNA-23a inhibits endometrial cancer cell development by targeting SIX1. *Oncology letters*. 2019;18:3792-802.
44. Xin X, Li Y, Yang X. SIX1 is overexpressed in endometrial carcinoma and promotes the malignant behavior of cancer cells through ERK and AKT signaling. *Oncology letters*. 2016;12:3435-40.
45. Suen AA, Jefferson WN, Wood CE, Padilla-Banks E, Bae-Jump VL, Williams CJ. SIX1 Oncoprotein as a Biomarker in a Model of Hormonal Carcinogenesis and in Human Endometrial Cancer. *Molecular cancer research : MCR*. 2016;14:849-58.
46. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. *Nature*. 2004;432:316-23.
47. Visconti R, Della Monica R, Grieco D. Cell cycle checkpoint in cancer: a therapeutically targetable double-edged sword. *Journal of experimental & clinical cancer research : CR*. 2016;35:153.

Tables

Due to technical limitations, table 1, 2 is only available as a download in the Supplemental Files section.

Figures

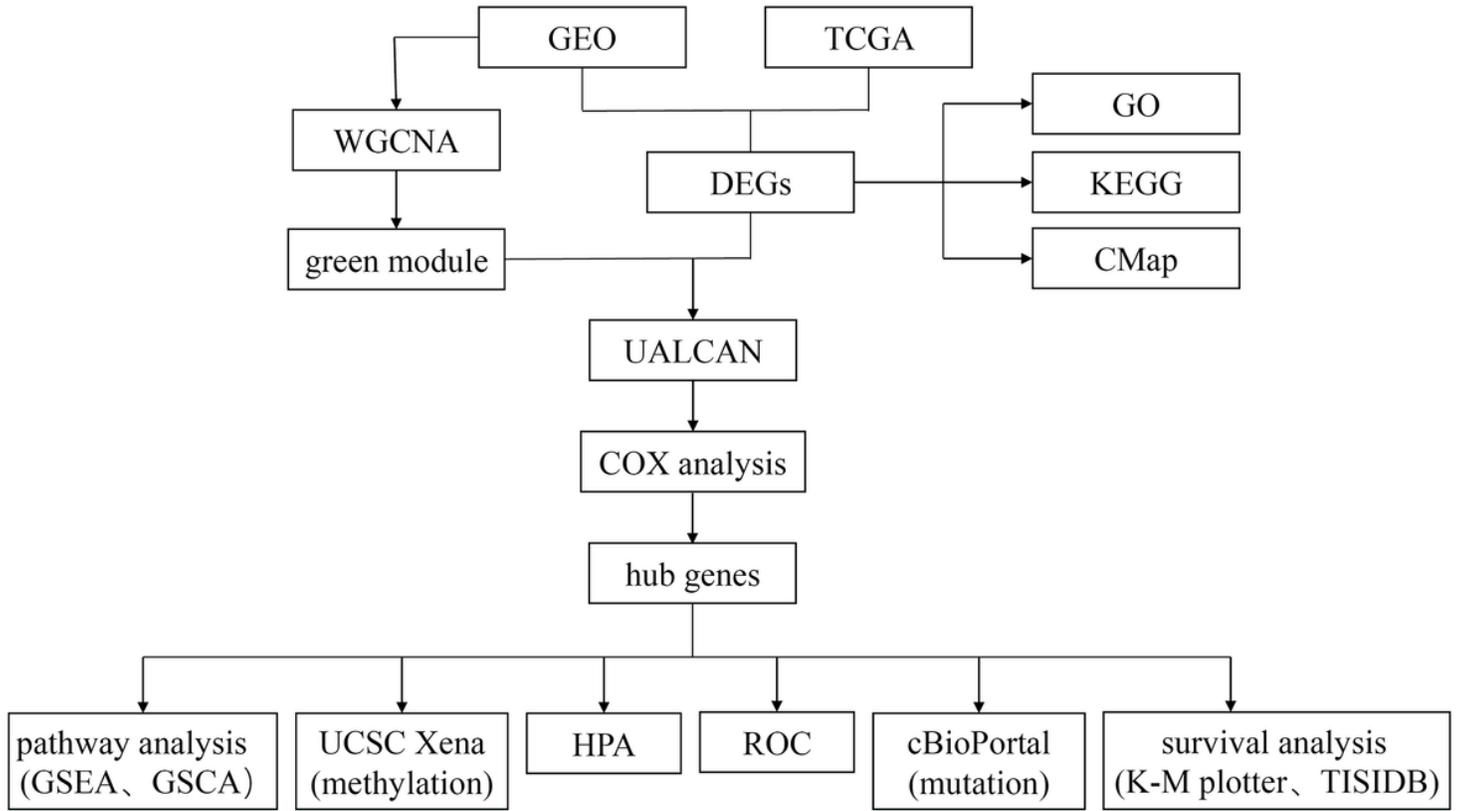


Figure 1

Flow chart of study design

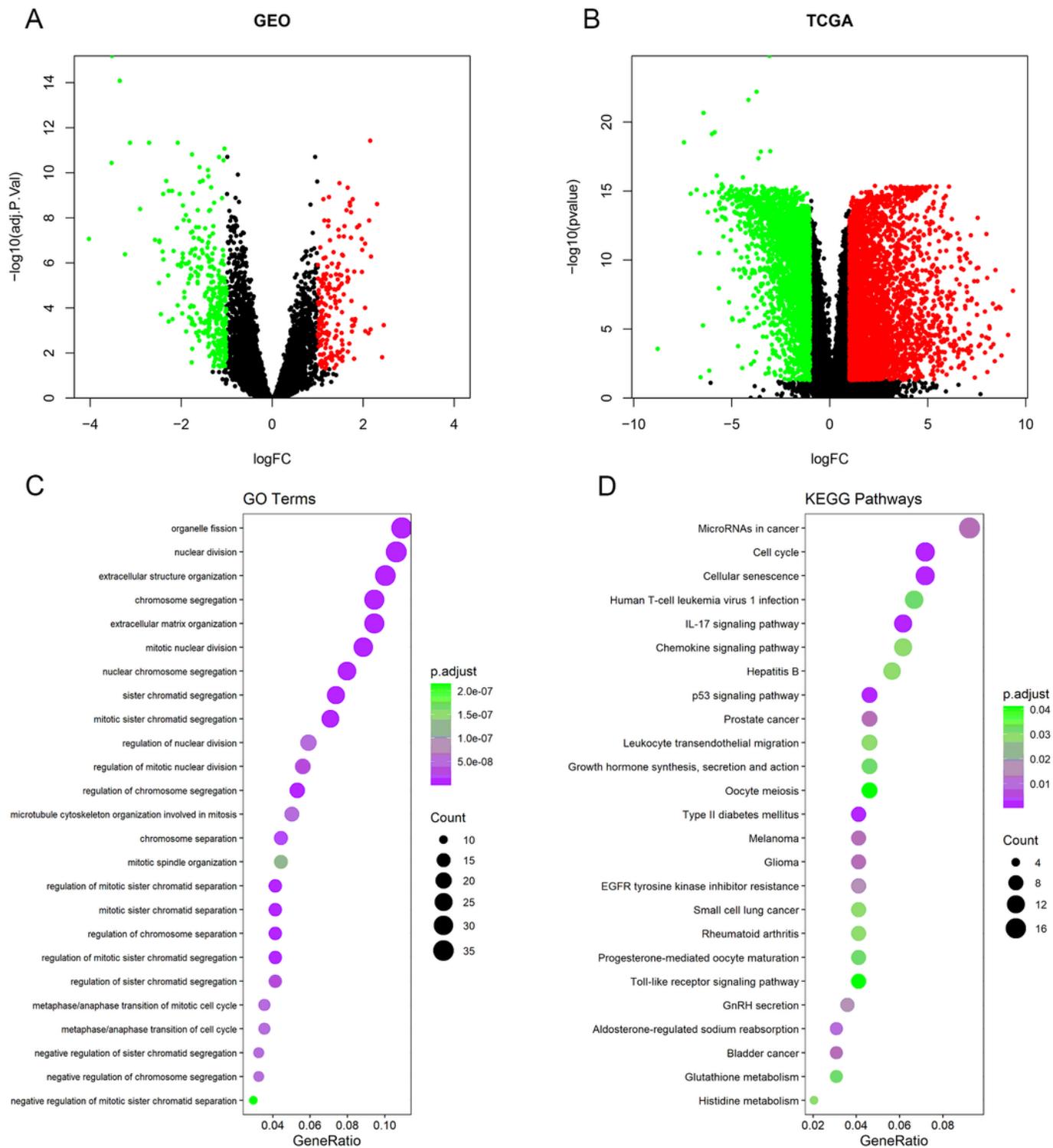


Figure 2

(A,B) The volcano plots visualize the DEGs in GEO, TCGA. Red dots represent upregulated genes, green dots represent downregulated genes. (C,D) Functional enrichment results of DEGs. (C) GO Terms. (D) KEGG Pathways.

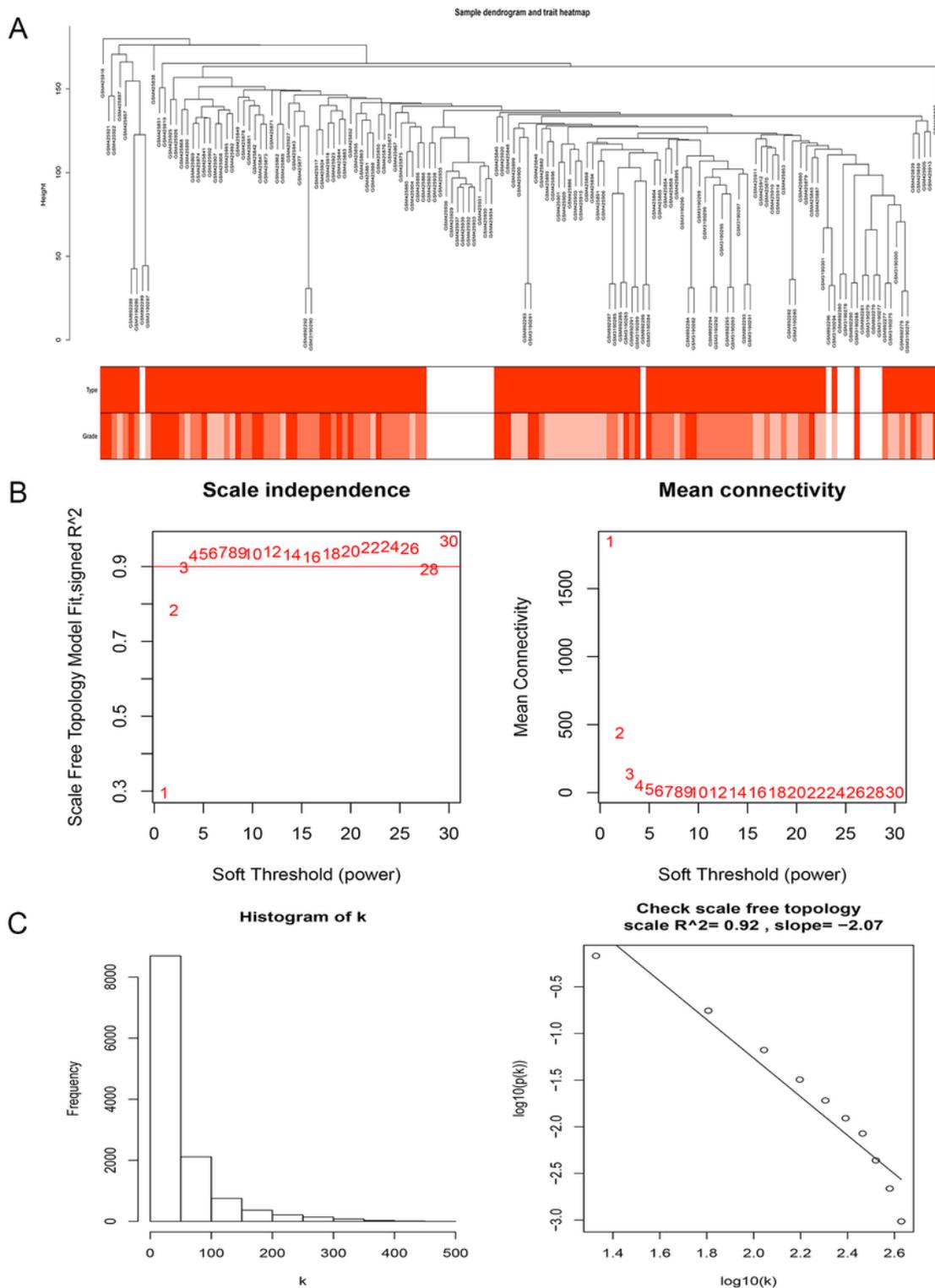


Figure 3

(A) Clustering dendrogram of 150 samples from GEO. (B) Analysis of trie scale-free fit index for various softthresholding powers (β) and Analysis of the mean connectivity for various soft-thresholding powers. (C) Histogram of connectivity distribution when $\beta = 4$ and Checking the scale free topology when $\beta = 4$.

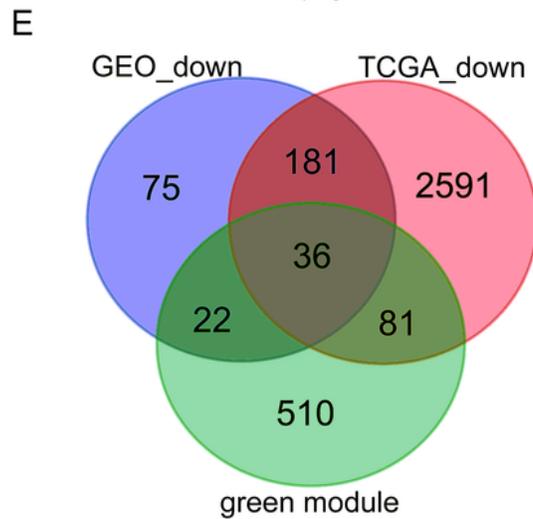
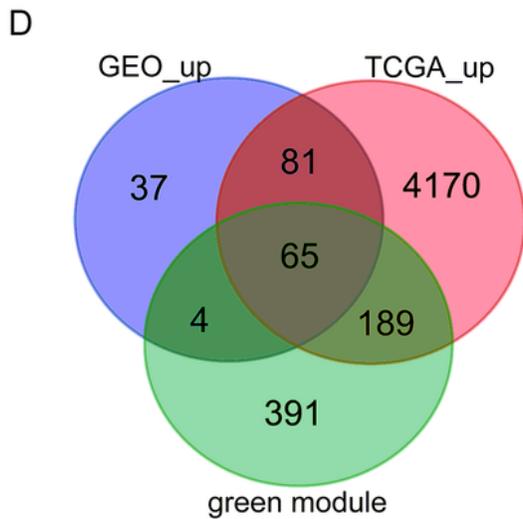
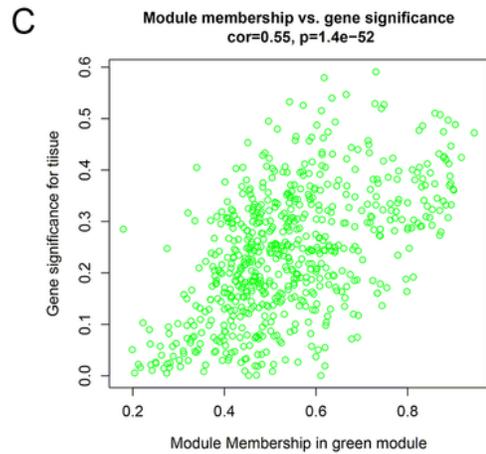
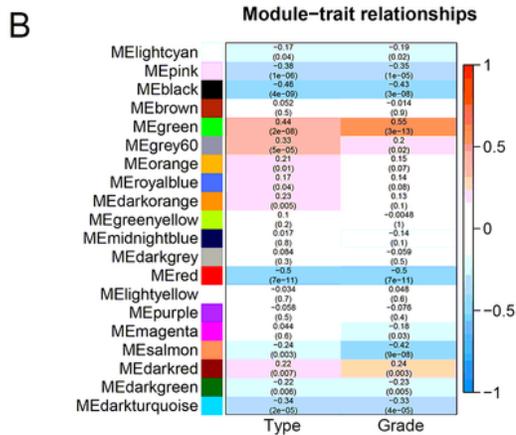
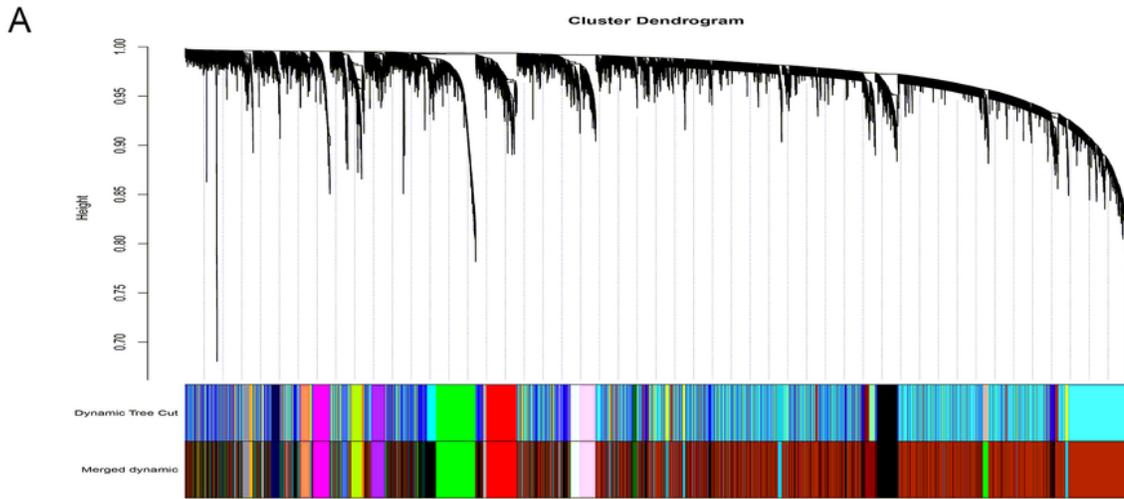


Figure 4

(A) Dendrogram of genes clustered based on a dissimilarity measure (1-TOM). (B) Heatmap of the correlation between MEs and clinical traits. (C) Scatter plot of module eigengenes in green module. (D) Up-regulated genes in green module. (E) Down-regulated genes in green module.

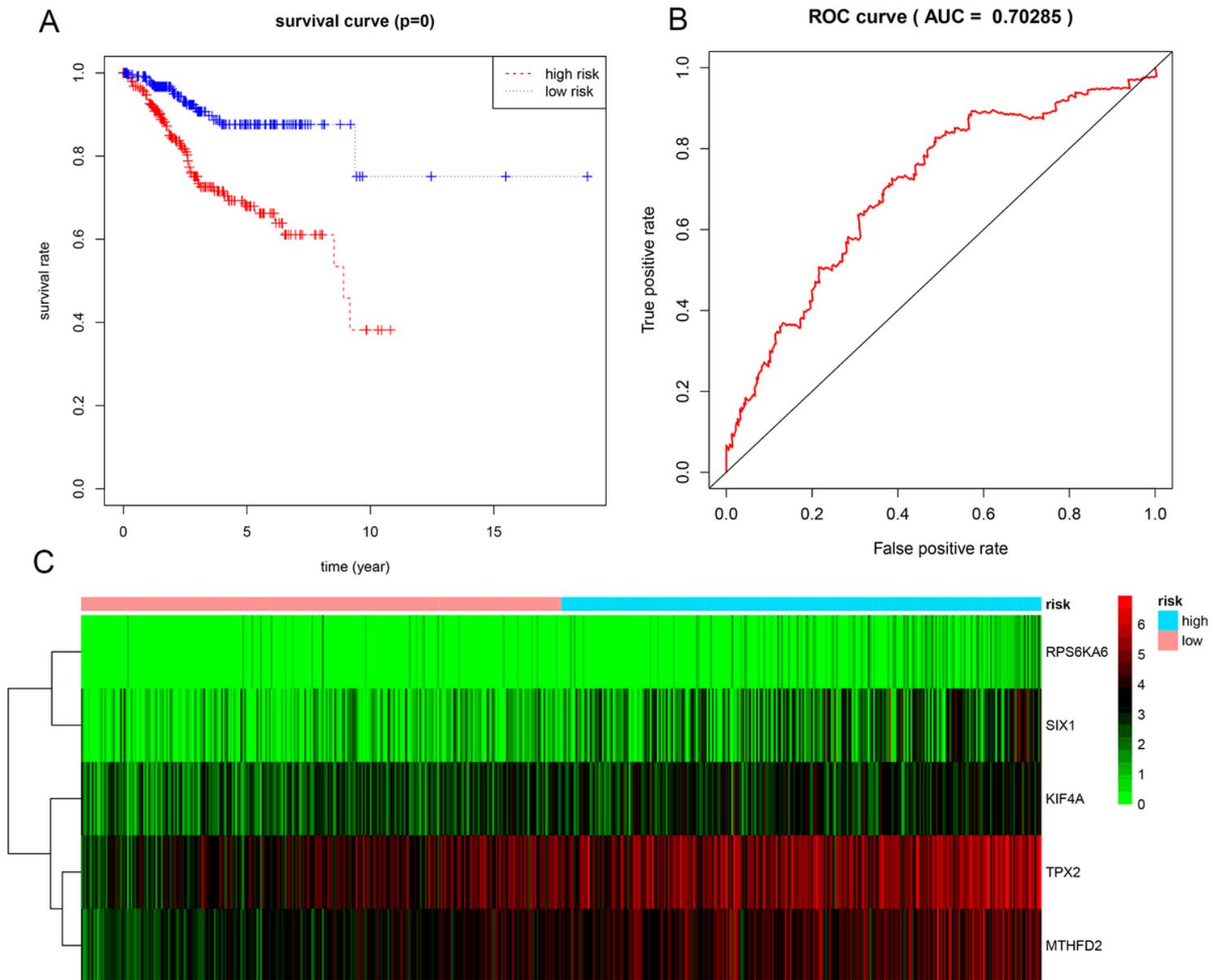


Figure 5

(A) Survival analysis of high-risk group and low-risk group in TCGA cohort. (B) ROC curve of the diagnosis value of 5 year OS for high-risk and low-risk group. (C) Five hub genes expression in different groups, with green to red indicating low expression to high expression.

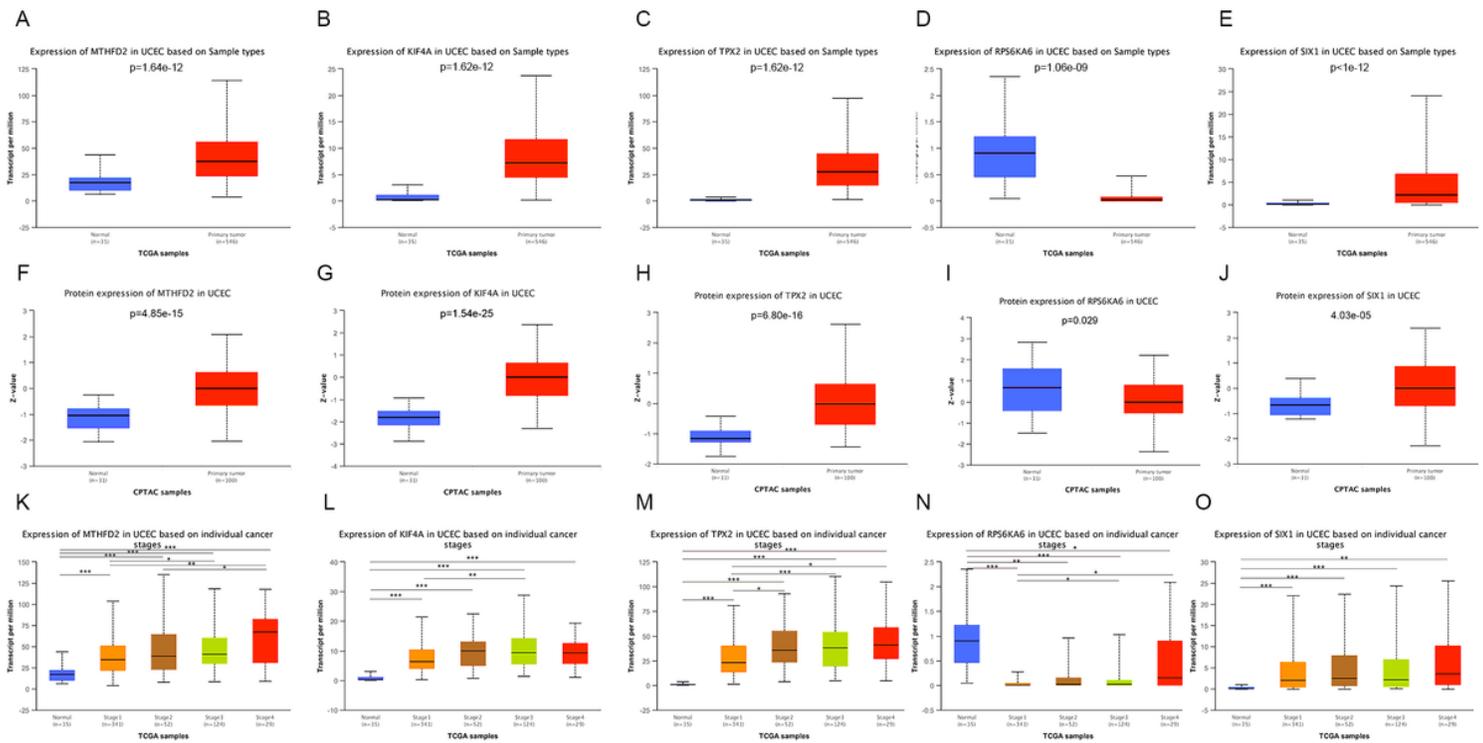


Figure 6

Validation of UALCAN database. (A-E) mRNA expression levels of 5 hub genes in EC and normal tissues. (A) MTHFD2; (B) KIF4A; (C) TPX2; (D) RPS6KA6; (E) SIX1. (F-J) Protein expression levels of 5 hub genes in EC and normal tissues. (F) MTHFD2; (G) KIF4A; (H) TPX2; (I) RPS6KA6; (J) SIX1. (K-O) mRNA expression of 5 hub genes in EC and normal tissues at different cancer stages. (K) MTHFD2; (L) KIF4A; (M) TPX2; (N) RPS6KA6; (O) SIX1. (* :P<0.05; ** :P<0.01; ***: P<0.001).

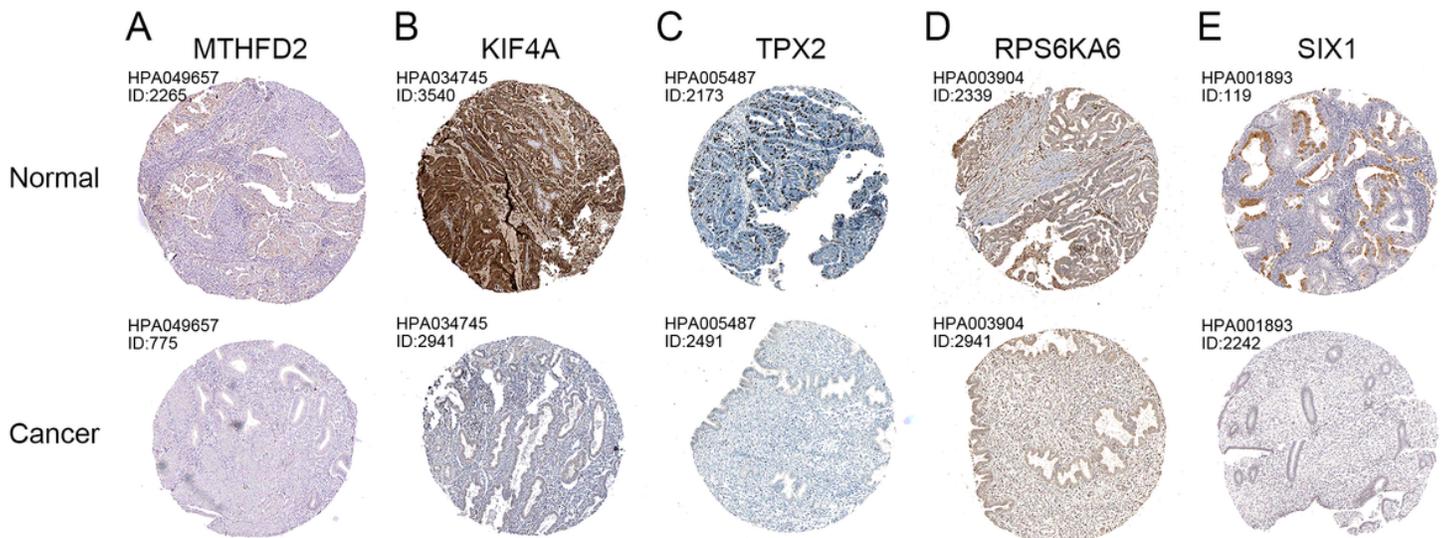


Figure 7

Immunohistochemistry images of 5 hub genes in EC tissues and normal endometrium tissues from the Human Protein Atlas database. (A) MTHFD2; (B) KIF4A; (C) TPX2; (D) RPS6KA6; (E) SIX1.

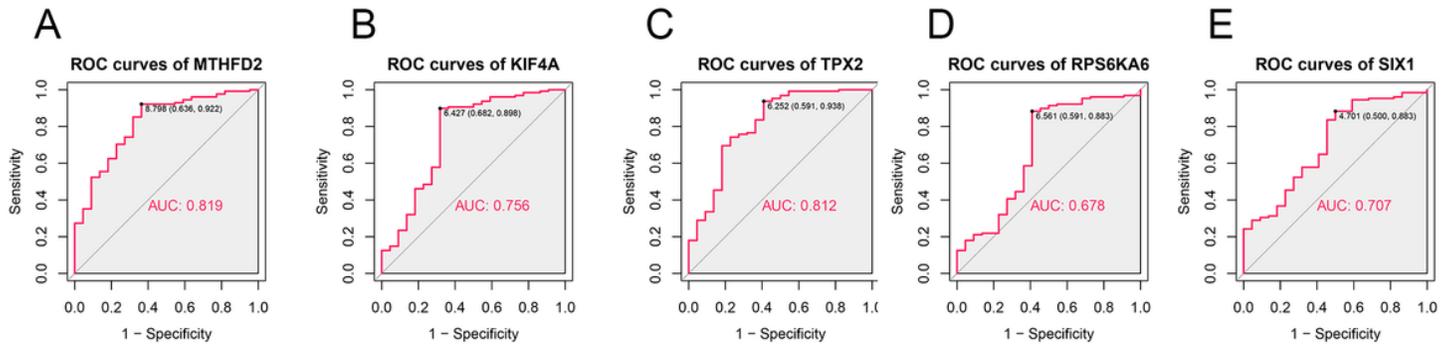


Figure 8

ROC curves of 5 hub genes for distinguishing EC from normal tissues. (A) MTHFD2; (B) KIF4A; (C) TPX2; (D) RPS6KA6; (E) SIX1.

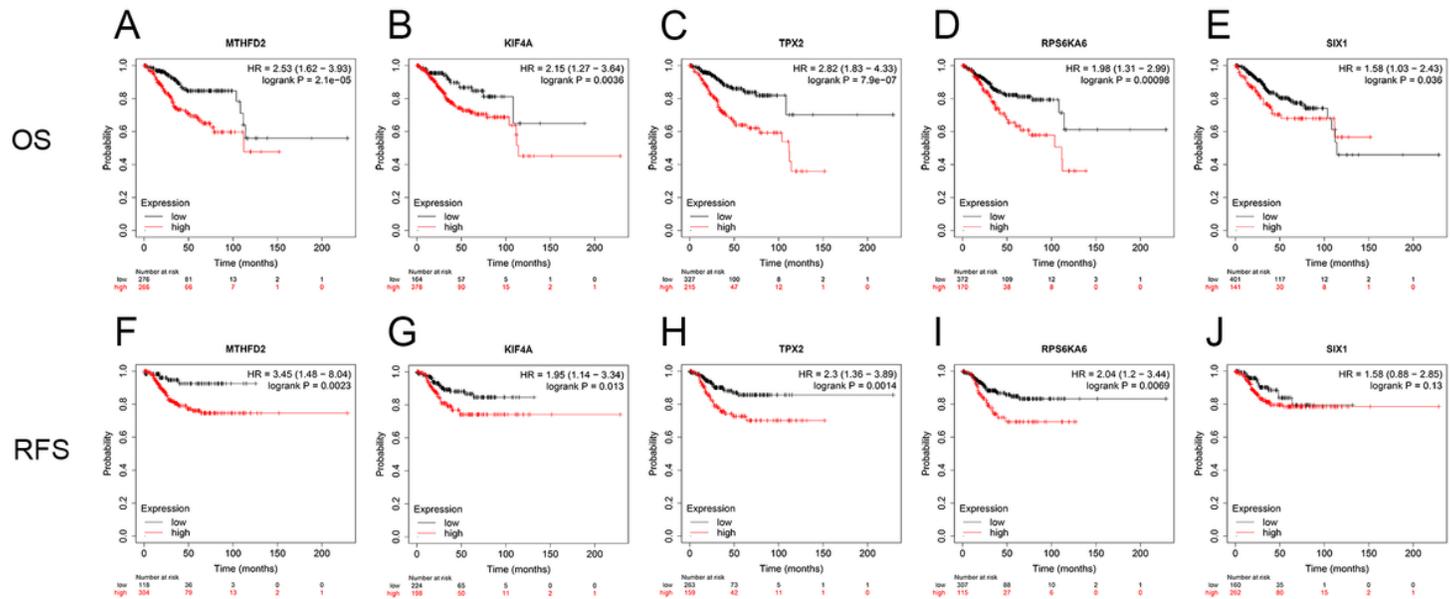


Figure 9

(A-E) Overall survival analysis of 5 hub genes in EC based on Kaplan-Meier plotter. (A) MTHFD2; (B) KIF4A; (C) TPX2; (D) RPS6KA6; (E) SIX1. (F-J) Relapse-free survival analysis of 5 hub genes in EC based on Kaplan-Meier plotter. (F) MTHFD2; (G) KIF4A; (H) TPX2; (I) RPS6KA6; (J) SIX1.

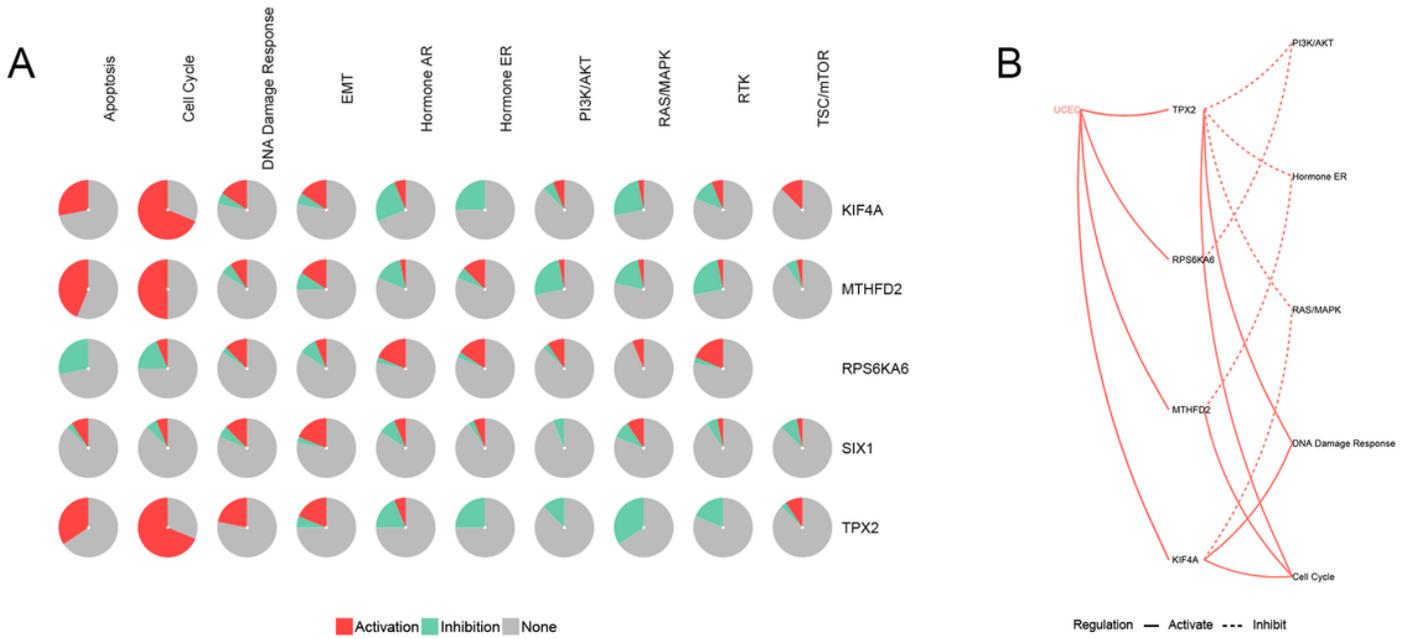


Figure 10

Cancer-related pathways of 5 hub genes based on GSCA Lite. (A) Global percentage; (B). Relation network.

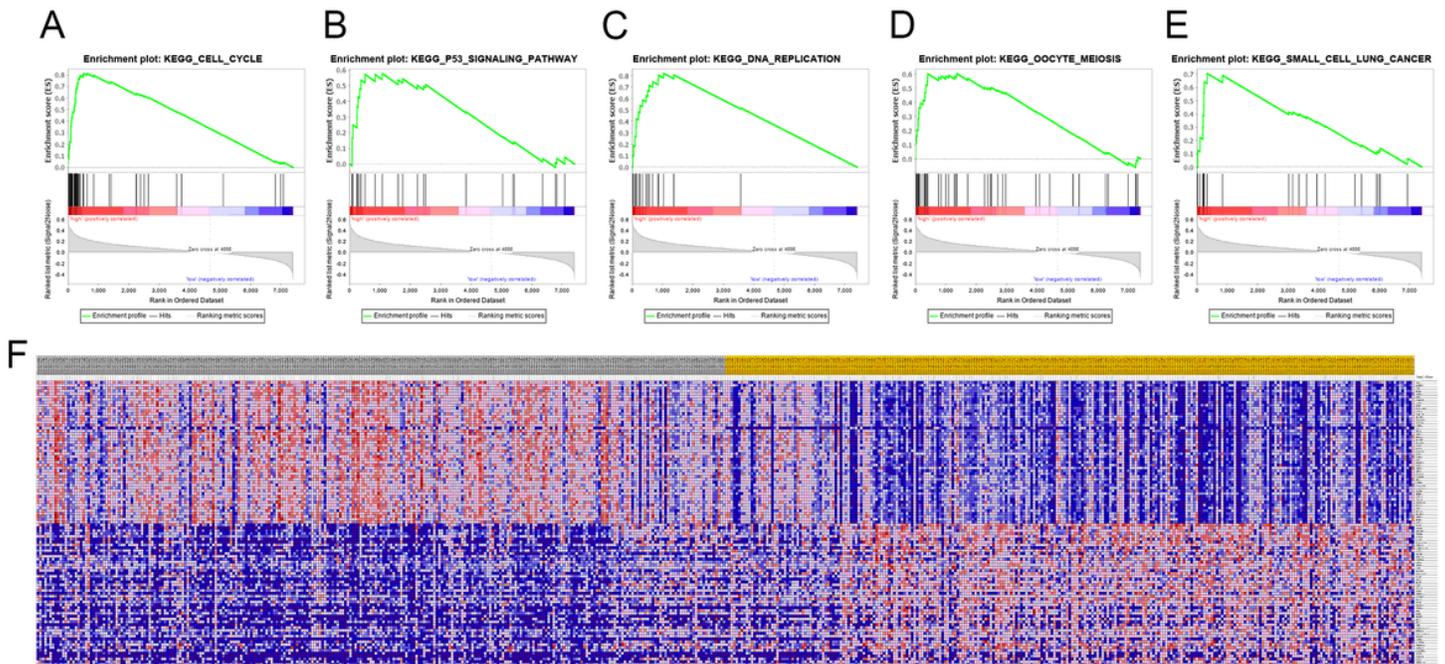


Figure 11

significantly correlated KEGG pathways and genes by GSEA. (A-E) KEGG pathways : (A) cell cycle; (B) p53 signaling pathway; (C) DNA replication; (D) oocyte meiosis; (E) small cell lung cancer. (F) Transcriptional expression profiles of the 100 significant genes were shown in the heat map.

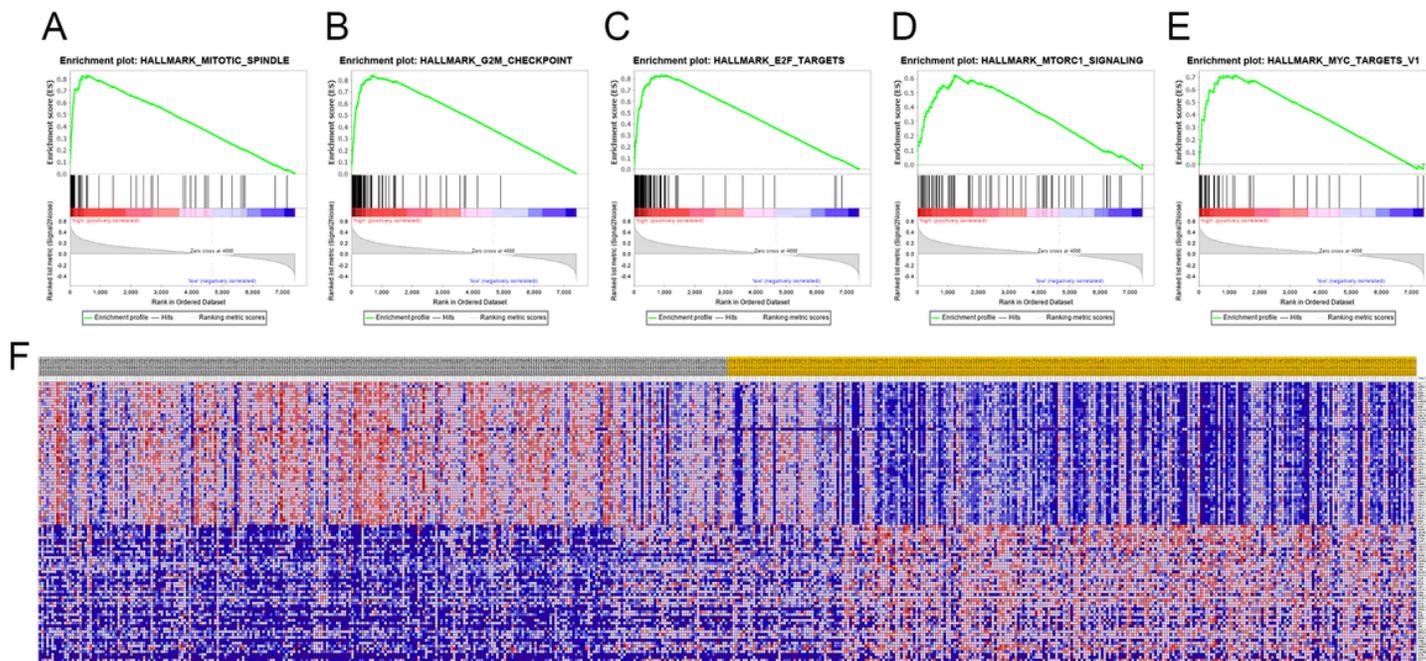


Figure 12

significantly correlated HALLMARK pathways and genes by GSEA. (A-E) HALLMARK pathways: (A) mitotic spindle; (B) G2M checkpoint; (C) E2F targets; (D) MTORC1 signaling; (E) MYC targets V1. (F) Transcriptional expression profiles of the 100 significant genes were shown in the heat map.

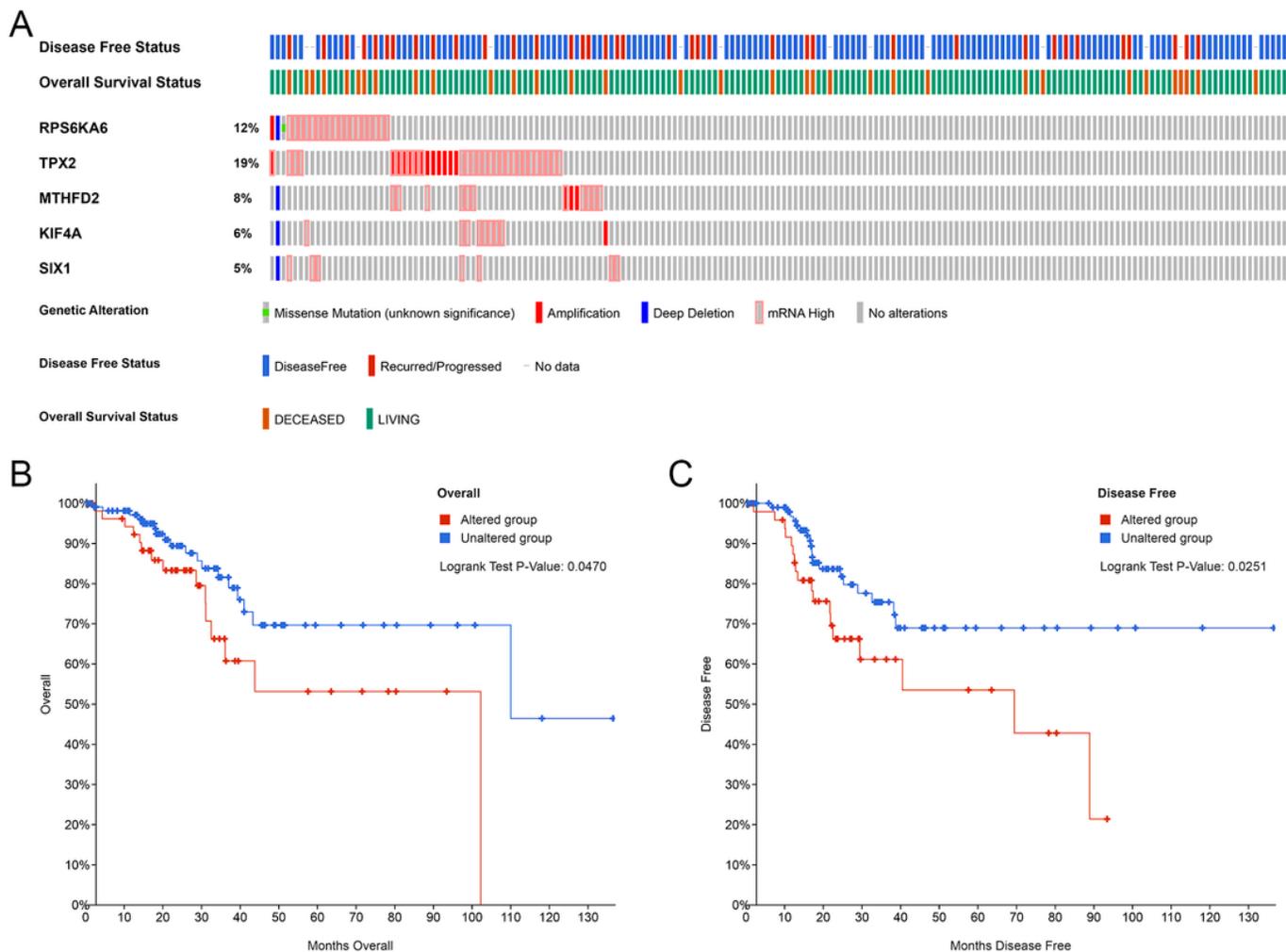


Figure 13

Analyses of genetic variations, OS and DFS of 5 hub genes in EC (cBioPortal). (A) OncoPrint visual summary of variations on a query of 5 hub genes. Genetic alterations in 5 hub genes were associated with shorter OS (B) and DFS (C) of EC patients.

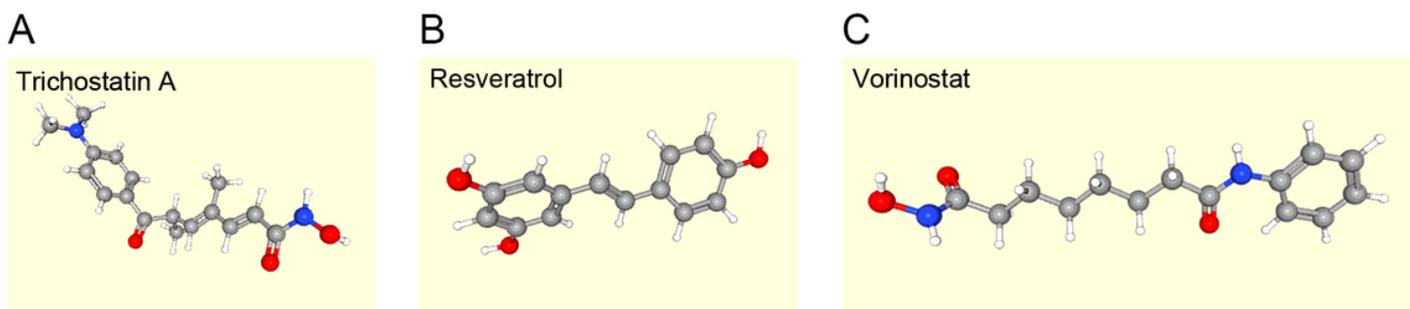


Figure 14

chemical structures of the top 3 molecule drugs. (A) Trichostatin A; (B) Resveratrol; (C) Vorinostat.

Supplementary Files

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

- [SupplementaryFigure1.tif](#)
- [SupplementaryFigure2.tiff](#)
- [SupplementaryFigure3.tiff](#)
- [SupplementaryFigure4.tiff](#)
- [SupplementaryFigure5.tiff](#)
- [SupplementaryFigure6.tiff](#)
- [SupplementaryFigure7.tiff](#)
- [SupplementaryFigure8.tiff](#)
- [SupplementaryFigure9.tiff](#)
- [SupplementaryFigure10.tiff](#)
- [OnlineTable1.png](#)
- [OnlineTable2.png](#)