

# CCT6A, a Novel Prognostic Biomarker for Ewing's Sarcoma

**Jie Jiang**

Guangxi Medical University

**Chong Liu**

Guangxi Medical University First Affiliated Hospital

**Guoyong Xu**

Guangxi Medical University

**Tuo Liang**

Guangxi Medical University

**Chaojie Yu**

Guangxi Medical University

**Zhaojie Qin**

Guangxi Medical University

**Shian Liao**

Guangxi Medical University

**Jiang Xue**

Guangxi Medical University

**Haopeng Zeng**

Guangxi Medical University

**Zide Zhang**

Guangxi Medical University

**Zhaojun Lu**

Guangxi Medical University

**Zequn Wang**

Guangxi Medical University

**Jiarui Chen**

Guangxi Medical University

**Tianyou Chen**

Guangxi Medical University

**Hao Li**

Guangxi Medical University

**Xinli Zhan** (✉ [zhanxinli\\_215@163.com](mailto:zhanxinli_215@163.com))

Guangxi Medical University <https://orcid.org/0000-0002-4599-1544>

## Primary research

**Keywords:** Ewing's sarcoma, biomarker, CCT6A gene, Overall survival, Event-free survival

**Posted Date:** August 11th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-52492/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Ewing's sarcoma (ES) is the second most prevalent malignancy among bone tissue tumors, and there is no adequate prognosis biomarker. The protein encoded by CCT6A is a molecular chaperone. Early studies have suggested that CCT6A is involved in the development of many cancers, however, there is no clear evidence of a role for CCT6A in ES.

**Methods:** In this study, we performed a bioinformatics analysis of 32 Ewing sarcoma specimens from the GSE17618 dataset for differences in gene expression and overall survival, event-free survival, and gene expression in different subgroups.

**Results:** After three screenings, we identified CCT6A as highly correlated with Ewing's sarcoma prognosis. Survival analysis showed low overall survival (OS) for CCT6A high expression ( $P=0.024$ ). On the other hand, Cox regression analysis showed that CCT6A expression, event-free survival (EFS), and age were strongly associated with the prognosis of Ewing sarcoma, identified as independent poor prognostic biomarkers. (CCT6A:  $P=0.015$ ; Age:  $P\text{-value}=0.026$ ; EFS:  $P\text{-value}=0.001$ ).

**Conclusion:** The expression level of CCT6A is strongly associated with the prognosis of Ewing's sarcoma. High expression of the CCT6A gene may serve as a biomarker for poor prognosis in patients with Ewing's sarcoma.

## Introduction

Ewing's sarcoma (ES) is a primary, highly malignant bone tumor originating in the bone marrow, most often occurring in children and adolescents, and is second only to osteosarcoma in its incidence in the adolescent population. ES is characterized by a poor prognosis, with a five-year survival rate of only 65–75%, and an even lower five-year survival rate ( $< 30\%$ ) in patients who have already developed metastases<sup>[1]</sup>. Notably, the annual incidence of ES is less than 1% of patients diagnosed with cancer<sup>[2]</sup>. Recent studies have shown that tumor prognosis can be predicted by biomarkers<sup>[3],[4]</sup>. However, there is still a gap in the field of prognosis-related biomarker research for ES.

The protein encoded by the CCT6A gene is a molecular chaperone, which is a member of the chaperone containing the TCP1 complex (CCT). It has been shown that CCT6A protein expression levels are elevated in tumor tissues of patients with hepatocellular carcinoma and that the overall survival of these patients is relatively shorter<sup>[5]</sup>. Previous studies have also confirmed that the number of mutant peptides has potential prognostic value, with somatic variants showing increased expression of CCT6A, which has clinical significance<sup>[6]</sup>. CCT6A is expression-inducible and amplification-inducible in glioblastoma, and shows a significant negative correlation with survival<sup>[7]</sup>. In addition, CCT6A has been found in breast cancer and it may play an important role in breast cancer progression<sup>[8]</sup>. Thus, CCT6A is involved in the progression of several cancers. However, the prognostic role of the expression level of CCT6A in ES remains largely unknown.

The search for biomarkers to predict the prognosis of a tumor through bioinformatics is accepted and recognized by a wide range of scholars. Biomarkers have been used to differentiate between patients at increased risk for nonalcoholic steatohepatitis (NASH) and advanced fibrosis, and to predict potential adverse clinical utility outcomes in patients at high risk<sup>[9]</sup>. It has excellent clinical relevance and value in guiding prognosis. Therefore, this study goes through a bioinformatics approach to search for biomarkers of Ewing's sarcoma, aiming to guide clinical diagnosis and treatment as well as to predict patient prognosis.

## Materials And Methods

### 1. Data download and preliminary filtering

We downloaded gene expression data and clinical information data for Ewing sarcoma patients from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). We analyzed gene expression data and clinical information data from the dataset GSE17618<sup>[10]</sup>. Data set GSE17618 belongs to platform GPL570, gene expression data was generated from Affymetrix Human Genome U133 Plus 2.0 Array on a microarray platform. The dataset GSE17618 consists of 43 Ewing sarcoma patient samples and 11 Ewing sarcoma cell lines. The exclusion criteria for this study were 1) a sample with a diagnosis of non-Ewing's sarcoma, 2) samples from Ewing sarcoma cell lines, and 3) specimens with imperfect clinical information. We extracted the following information from all clinical information for subsequent analyses: gender, age, tumor status, overall survival time(OS), event-free survival (EFS), and survival status. This study was based on data from an open database. Ethics and patient consent are not applicable.

### 2. Data filtering and genetic screening

All statistical analysis operations were performed using R software<sup>[11]</sup> (version 4.0.0). We used R package(Impute) to fill in and correct the data for the expression matrix, and we used the Limma package to standardize the gene expression data<sup>[12]</sup>. Gene expression matrix data were log<sub>2</sub>-scale transformed. The first step, survival filtering. We used the R package (survival package) to perform a survival analysis of patient survival status and survival time using the Kaplan-Meier method for all genes in the gene expression data. On the other hand, a survival analysis of the patient's survival status and survival time was also performed using Cox regression analysis of all genes in the gene expression data using the R package (survival package). Genes that also meet the following conditions will proceed to the next round of filtering: 1) P-value < 0.05 obtained using the Kaplan-Meier method; 2) P-value < 0.05 obtained using the Cox regression method; 3) genes with a >10% difference in the five-year survival of patients. In the second step, the independent prognostic analysis was filtered. We used univariate Cox regression analysis and multivariate Cox regression analysis to compare the genes obtained in the first step with survival time and survival status, respectively, and the genes with P-value less than 0.05 obtained by both methods were saved by us for subsequent filtering. The third step is clinical relevance filtering. We grouped patients according to gender, age, tumor status, and EFS separately. Age, sex, and EFS were each

divided into two subgroups, and we used the Wilcox method to analyze these subgroups in relation to survival status survival time, P-value<0.05 was considered statistically significant. The tumor status was divided into three subgroups, which we analyzed with survival status and survival time using the Kruskal-Wallis method, with a p-value < 0.05 considered significant. Genes that satisfied P-value < 0.05 in both methods were eventually preserved by us.

### **3. Data analysis and visualization**

We used the R package(limma) package for differential analysis of gene expression data for all samples with the cut off value set to: adjusted P-value<0.05,|logFC|>0.5. The R package (pheatmap) was used to construct a heat map of the differential genes, highlighting the regions where the differential genes were mainly concentrated, and to construct a Volcano map of the differential genes. Subsequently, Genome Ontology (GO) enrichment analysis<sup>[13]</sup> and KEGG pathway enrichment analysis<sup>[14]</sup> and visualization of differential genes were performed based on the R package (clusterProfile<sup>[15]</sup>, rg.Hs.eg.db, enrichplot, and ggplot2<sup>[16]</sup> respectively). The value of cut off is set to p-value<0.05. We imported differentially expressed genes into the STRING database (<https://string-db.org/>)<sup>[17]</sup> to obtain protein interaction networks, which were imported into cytoscape (version 3.8.0)<sup>[18]</sup> for visualization. We plotted the correlation coefficient heat map between genes based on the R package (corrplot).

## **Results**

### **1. The relationship between high expression levels of CCT6A and survival in ES patients.**

After initial screening, a total of 32 samples in the dataset GSE17618 met our screening criteria. Subsequently, after our three-step screening process, only the CCT6A gene finally met all of our screening criteria. We used the Kaplan-Meier method for survival analysis of patients and divided all samples into two groups using the median value of gene expression of CCT6A as a reference. Median gene expression values greater than CCT6A are classified as high expression group; median gene expression values less than or equal to CCT6A are classified as low expression group. Survival curves were plotted based on the survival time of patients in the high and low expression groups( Figure 1). We can derive from the survival curves that the probability of survival of patients with high expression of the CCT6A gene was significantly lower than the probability of survival of patients with low expression of the CCT6A gene (P=0.024), and the difference was statistically significant.

### **2. Univariate and multivariate Cox regression analysis for CCT6A**

We used univariate Cox regression analysis and multivariate Cox regression analysis to analyze CCT6A in relation to survival status and survival time, respectively. From the univariate Cox regression analysis plot (Figure 2.A), we found that CCT6A was significantly associated with survival status and survival time (P < 0.001, HR=7.953), and HR>1 indicated that CCT6A was a high-risk factor, and the higher the CCT6A expression value, the poorer the prognosis of this patient. Also, CCT6A can be considered as an

independent prognostic biomarker and we can predict the patient's prognosis by detecting the gene expression level of CCT6A. From Figure 2.A we can also see that EFS is also strongly associated with survival status and survival time ( $P < 0.001$ ,  $HR = 0.218$ ), with  $EFS < 1$  indicating that EFS is a low-risk factor. From the forest plot obtained from multivariate Cox analysis (Figure. 2.B), we found that CCT6A was significantly associated with survival ( $P < 0.001$ ,  $HR = 9.513$ ) and was a high-risk factor for ES prognosis. Age was also strongly associated with survival, ( $P = 0.002$ ,  $HR = 0.269$ ) and was a low-risk factor for ES prognosis. EFS was strongly associated with survival ( $P < 0.001$ ,  $HR = 0.196$ ) and was a low-risk factor for ES prognosis.

### **3. Expression of CCT6A in different subgroups**

We analyzed the gene expression of CCT6A in the different groups and found substantial differences. Gene expression of CCT6A in both age groups: the median value of gene expression in the  $>20$ -year-old group was higher than the median value of gene expression in the  $\leq 20$ -year-old group, but the difference was not statistically significant ( $P > 0.05$ ), see Figure. 3.A for details. The median value of CCT6A expression was significantly different in the group of patients with  $EFS \leq 5$  years compared to the group of patients with  $EFS > 5$  years ( $P = 0.01$ ), and gene expression was higher in the  $EFS \leq 5$  group compared to the  $EFS > 5$  group, as detailed in Figure. 3.B. From this subgroup of tumor origin, it was observed that patients in the recurrent group had the highest median value of CCT6A gene expression, followed by those in the metastatic group and last in the primary group. Among them, the expression of CCT6A in the patients in the primary group was significantly different from that in the relapsed group ( $P < 0.01$ ), as detailed in Figure 3.C. From the expression of the CCT6A gene in different sexes (Figure. 3.D), we observed that the median value of CCT6A expression in females was significantly higher than that in males, and the difference was statistically significant ( $P < 0.05$ ).

### **4. Screening of differentially expressed genes, enrichment analysis, and visualization, and construction of protein interaction networks.**

We analyzed differentially expressed genes and performed GO enrichment analysis and KEGG enrichment analysis and visualization of differentially expressed genes; finally, we constructed protein reciprocal networks through databases. Screening of differentially expressed genes according to the screening criteria yielded 188 differentially expressed genes, of which 106 were up-regulated and 82 were down-regulated. Heat maps and volcanoes of differentially expressed genes are shown in Figure 4.A and Figure 4.B. A total of 106 genes were positively correlated with CCT6A and 82 genes were negatively correlated with CCT6A, see Figure 5. Subsequently, we performed GO enrichment analysis and visualization of differentially expressed genes (Figure. 6.A), showing the biological process (BP), cell component (CC), and molecular function (MF) located in the top 10, respectively. The differentially expressed genes were analyzed and visualized for KEGG enrichment (Figure. 6.B), demonstrating the KEGG-enriched pathways. Finally, we imported the differentially expressed genes into the STRING database to obtain the protein interaction network, which was imported into Cytoscape for visualization (Figure. 7).

## Discussion

As early as 2006, CCT6A was shown to play an important regulatory role in cancer<sup>[19]</sup>. Meng Zhu in 2017 found that low-frequency missense variants in the chaperone protein accompanying CCT6A were significantly associated with survival in patients with non-small cell lung cancer<sup>[20]</sup>. Klimczak M used the TCGA database to identify the overexpression of CCT6A in breast cancer as a significant contributor to poor prognosis<sup>[21]</sup>. In a study on colon adenocarcinoma, it was pointed out that high expression of the CCT6A promoted the growth of colon adenocarcinoma cells on the one hand, and was associated with low survival rate of colon adenocarcinoma on the other<sup>[22]</sup>. CCT6A also showed a strong association with survival and prognosis in renal cancer, with patients with altered CCT6A mRNA (stage I-IV) having significantly shorter overall survival compared to healthy controls<sup>[23]</sup>. In summary, the CCT6A gene plays a key role in the development of many tumors and affects prognosis. These are consistent with our research. Notably, in the present study, patients with high CCT6A expression had a lower survivorship probability than those with low CCT6A gene expression. In addition, most patients with EFS  $\leq$  5 years showed high expression of the CCT6A gene. These results suggest that patients with ES accompanied by elevated CCT6A gene expression levels may have a strong association with low survival probability and EFS  $\leq$  5 years. Another noteworthy point is that in our results on CCT6A gene expression under different subgroups, we found that the median gene expression level of CCT6A gene expression in patients in the metastatic group was much higher than the median gene expression of patients in the primary tumor group and the median gene expression of patients in the recurrent group. This suggests that there is a relationship between high expression of the CCT6A gene and tumor metastasis. It is well known that in most cases, metastasis of the tumor is a vital factor in the death of the patient<sup>[24]</sup>. Therefore, high expression of CCT6A may be accompanied by tumor metastasis and EFS  $\leq$  5 years. In other words, high expression of the CCT6A gene could serve as an important biomarker for the predictive prognosis of ES. If high expression of the CCT6A gene is detected in a given ES patient, it predicts a much lower SURVIVAL probability than in a patient without high CCT6A expression, as well as a high probability that the patient will have an EFS  $\leq$  5 years.

At the same time, the results of KEGG enrichment analysis and GO enrichment analysis corroborate our view. In this study, the KEGG enrichment analysis of differentially expressed genes was mainly distributed in the Cell cycle, DNA replication, Mismatch repair, cellular senescence and other pathways. As early as 2001, it was shown that the cell cycle in cancer, dysregulation of cell proliferation and a number of other factors together constitute the minimum required site for tumor development<sup>[25]</sup>. Research in DNA replication is also progressing, and its relationship to cancer development is becoming clearer. Dysregulation of DNA replication can lead to genomic instability, and an important feature of cancer formation is this instability<sup>[26]</sup>. Cancer may be related to errors in DNA at the replication stage and the combined effects of changes caused by environmental factors and mutations in genetic information<sup>[27]</sup>. The Mismatch repair pathway plays an important role in the study of many tumors<sup>[28],[29],[30]</sup>. Cellular senescence, which refers to a permanent state of cell cycle arrest, can lead to a decrease in the regenerative potential and tissue function of an aging organism, which can lead to tumorigenesis<sup>[31]</sup>.

Another way in which aging promotes cancer progression is by promoting pathological cell proliferation<sup>[32]</sup>. The aforementioned KEGG pathway enrichment results indicate that the pathways enriched by the differential genes in this study are strongly associated with tumor cell initiation, development, and formation. In the GO enrichment analysis, CCT6A was mainly enriched in the DNA biosynthetic process pathway. It has been noted that acquired mutations or inheritance in major factors regulating DNA methylation can be observed in cancer<sup>[33]</sup>. We conjecture that the CCT6A gene likely promotes ES by regulating the DNA biosynthetic process. There are a total of 13 genes in the protein reciprocal network that are closely associated with CCT6A, and all 13 genes are very closely related to cancer. Among them, the CSE1L gene has been found to show a positive correlation between its high expression and high cancer stage and poor prognosis in cancer patients<sup>[34]</sup>. Meanwhile, the CCT2 gene has also been shown to reduce the survival rate of colorectal cancer patients with higher levels of gene expression<sup>[35]</sup>. Elevated expression levels of The CCNE2 gene coincided with shorter overall survival in patients with hepatocellular carcinoma<sup>[36]</sup>. Thus, the results presented by the network of protein interactions suggest that all of these genes linked to CCT6A are associated with the development of cancer.

The results of this study clearly show that the expression level of CCT6A presents high expression in ES patients, and high expression of CCT6A in ES patients is associated with poor prognosis. Differentially expressed genes in GO enrichment analysis are mainly distributed during DNA biosynthesis, DNA conformational changes, chromosomal regions, catalytic activity, and action on DNA. KEGG Pathway enrichment analysis was mainly enriched in Cell cycle, DNA replication, Mismatch repair, and cellular senescence pathways. The protein interaction network also suggests that all of these genes linked to CCT6A are associated with cancer development. In the survival curve, patients accompanied by high expression levels of the CCT6A gene had a much lower SURVIVAL probability than patients not accompanied by high expression levels of CCT6A. Patients with EFS  $\leq$  5 years all had high expression of the CCT6A gene. High expression levels of CCT6A have also been linked to tumor metastasis in ES patients. All of these results suggest that CCT6A can serve as a biomarker for ES and is an important gene for predicting prognosis.

This study analyzed gene table data and clinical data from a bioinformatics perspective only to derive biomarkers that correlate Ewing sarcoma with prognosis. There is also a limitation to our study: there is no experimental way to verify that the genes we arrive at are also consistent in the experiment.

## Conclusion

CCT6A's can be an effective biomarker for ES prediction prognosis. High expression of CCT6A in ES predicted a lower 5-year survival rate for patients.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** All co-authors were consulted and gave their consent for publication.

**Availability of data and materials** The datasets supporting the conclusions of this article are available in the GEO repository, <https://www.ncbi.nlm.nih.gov/geo/>.

GSE17618: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17618>

**Competing interests** There are no conflicts of interest among any of the co-authors.

**Funding** Not applicable.

**Author contributions** Jie Jiang, Chong Liu, and Xinli Zhan designed the study. Zhaojie Qin, Chaojie Yu, Tuo Liang, Shian Liao, Jiang Xue, Haopeng Zeng, Guoyong Xu, Zide Zhang, Zhaojun Lu, Zequn Wang, Jiarui Chen, Tianyou Chen and Hao LI analyze the data. Jie Jiang wrote and revised the manuscript. Chong Liu and Xinli Zhan revised the manuscript. All authors read and approved the final manuscript.

**Acknowledgements** Thanks to all co-authors who contributed to this research.

## References

1. Gaspar N, Hawkins DS, Dirksen U, Lewis IJ, Ferrari S, Le Deley M-C, Kovar H, Grimer R, Whelan J, Claude L *et al*: **Ewing Sarcoma: Current Management and Future Approaches Through Collaboration**. *J Clin Oncol* 2015, **33**(27):3036-3046.
2. Ferguson JL, Turner SP: **Bone Cancer: Diagnosis and Treatment Principles**. *Am Fam Physician* 2018, **98**(4):205-213.
3. Montoya C, Rey L, Rodriguez J, Fernandez MJ, Troncoso D, Canas A, Moreno O, Henriquez B, Rojas A: **Epigenetic control of the EWSFLI1 promoter in Ewing's sarcoma**. *Oncol Rep* 2020, **43**(4):1199-1207.
4. Orth MF, Holting TLB, Dallmayer M, Wehweck FS, Paul T, Musa J, Baldauf MC, Surdez D, Delattre O, Knott MML *et al*: **High Specificity of BCL11B and GLG1 for EWSR1-FLI1 and EWSR1-ERG Positive Ewing Sarcoma**. *Cancers (Basel)* 2020, **12**(3).
5. Zeng G, Wang J, Huang Y, Lian Y, Chen D, Wei H, Lin C, Huang Y: **Overexpressing CCT6A Contributes To Cancer Cell Growth By Affecting The G1-To-S Phase Transition And Predicts A Negative Prognosis In Hepatocellular Carcinoma**. *OncoTargets and therapy* 2019, **12**:10427-10439.
6. Ma Y-S, Huang T, Zhong X-M, Zhang H-W, Cong X-L, Xu H, Lu G-X, Yu F, Xue S-B, Lv Z-W *et al*: **Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis**. *Mol Cancer* 2018, **17**(1):139.
7. Hallal S, Russell BP, Wei H, Lee MYT, Toon CW, Sy J, Shivalingam B, Buckland ME, Kaufman KL: **Extracellular Vesicles from Neurosurgical Aspirates Identifies Chaperonin Containing TCP1 Subunit 6A as a Potential Glioblastoma Biomarker with Prognostic Significance**. *Proteomics* 2019, **19**(1-2):e1800157.

8. Huang K, Zeng Y, Xie Y, Huang L, Wu Y: **Bioinformatics analysis of the prognostic value of CCT6A and associated signalling pathways in breast cancer.** *Molecular medicine reports* 2019, **19**(5):4344-4352.
9. Vilar-Gomez E, Chalasani N: **Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers.** *J Hepatol* 2018, **68**(2):305-315.
10. Savola S, Klami A, Myllykangas S, Manara C, Scotlandi K, Picci P, Knuutila S, Vakkila J: **High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients.** *ISRN Oncol* 2011, **2011**:168712.
11. Mair P, Hofmann E, Gruber K, Hatzinger R, Zeileis A, Hornik K: **Motivation, values, and work design as drivers of participation in the R open source project for statistical computing.** *Proceedings of the National Academy of Sciences of the United States of America* 2015, **112**(48):14788-14792.
12. ME R, B P, D W, Y H, CW L, W S, research SGJNa: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** 2015, **43**(7):e47.
13. Gaudet P, Dessimoz C: **Gene Ontology: Pitfalls, Biases, and Remedies.** *Methods Mol Biol* 2017, **1446**:189-205.
14. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic Acids Res* 2017, **45**(D1):D353-D361.
15. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *OMICS* 2012, **16**(5):284-287.
16. Walter W, Sánchez-Cabo F, Ricote M: **GOplot: an R package for visually combining expression data with functional analysis.** *Bioinformatics* 2015, **31**(17):2912-2914.
17. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.** *Nucleic Acids Res* 2017, **45**(D1):D362-D368.
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498-2504.
19. Tanic N, Brkic G, Dimitrijevic B, Dedovic-Tanic N, Gefen N, Benharroch D, Gopas J: **Identification of differentially expressed mRNA transcripts in drug-resistant versus parental human melanoma cell lines.** *Anticancer Res* 2006, **26**(3A):2137-2142.
20. Zhu M, Geng L, Shen W, Wang Y, Liu J, Cheng Y, Wang C, Dai J, Jin G, Hu Z *et al*: **Exome-Wide Association Study Identifies Low-Frequency Coding Variants in 2p23.2 and 7p11.2 Associated with Survival of Non-Small Cell Lung Cancer Patients.** *J Thorac Oncol* 2017, **12**(4):644-656.
21. Klimczak M, Biecek P, Zylicz A, Zylicz M: **Heat shock proteins create a signature to predict the clinical outcome in breast cancer.** *Sci Rep* 2019, **9**(1):7507.
22. Hu M, Fu X, Si Z, Li C, Sun J, Du X, Zhang H: **Identification of Differently Expressed Genes Associated With Prognosis and Growth in Colon Adenocarcinoma Based on Integrated Bioinformatics Analysis.** *Front Genet* 2019, **10**:1245.

23. Di Meo A, Batruch I, Brown MD, Yang C, Finelli A, Jewett MA, Diamandis EP, Yousef GM: **Searching for prognostic biomarkers for small renal masses in the urinary proteome.** *International journal of cancer* 2020, **146**(8):2315-2325.
24. Massagué J, Obenauf AC: **Metastatic colonization by circulating tumour cells.** *Nature* 2016, **529**(7586):298-306.
25. Evan GI, Vousden KH: **Proliferation, cell cycle and apoptosis in cancer.** *Nature* 2001, **411**(6835):342-348.
26. Kitao H, Iimori M, Kataoka Y, Wakasa T, Tokunaga E, Saeki H, Oki E, Maehara Y: **DNA replication stress and cancer chemotherapy.** *Cancer science* 2018, **109**(2):264-271.
27. Tomasetti C, Li L, Vogelstein B: **Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention.** *Science* 2017, **355**(6331):1330-1334.
28. Arend RC, Jones BA, Martinez A, Goodfellow P: **Endometrial cancer: Molecular markers and management of advanced stage disease.** *Gynecol Oncol* 2018, **150**(3):569-580.
29. Baretta M, Le DT: **DNA mismatch repair in cancer.** *Pharmacol Ther* 2018, **189**:45-62.
30. Fishel R: **Mismatch repair.** *J Biol Chem* 2015, **290**(44):26395-26403.
31. Hernandez-Segura A, Nehme J, Demaria M: **Hallmarks of Cellular Senescence.** *Trends Cell Biol* 2018, **28**(6):436-453.
32. Campisi J: **Aging, cellular senescence, and cancer.** *Annu Rev Physiol* 2013, **75**:685-705.
33. Michalak EM, Burr ML, Bannister AJ, Dawson MA: **The roles of DNA, RNA and histone methylation in ageing and cancer.** *Nat Rev Mol Cell Biol* 2019, **20**(10):573-589.
34. Tai C-J, Hsu C-H, Shen S-C, Lee W-R, Jiang M-C: **Cellular apoptosis susceptibility (CSE1L/CAS) protein in cancer metastasis and chemotherapeutic drug-induced apoptosis.** *J Exp Clin Cancer Res* 2010, **29**:110.
35. Park SH, Jeong S, Kim BR, Jeong YA, Kim JL, Na YJ, Jo MJ, Yun HK, Kim DY, Kim BG *et al.* **Activating CCT2 triggers Gli-1 activation during hypoxic condition in colorectal cancer.** *Oncogene* 2020, **39**(1):136-150.
36. Sonntag R, Giebeler N, Nevzorova YA, Bangen J-M, Fahrenkamp D, Lambertz D, Haas U, Hu W, Gassler N, Cubero FJ *et al.* **Cyclin E1 and cyclin-dependent kinase 2 are critical for initiation, but not for progression of hepatocellular carcinoma.** *Proceedings of the National Academy of Sciences of the United States of America* 2018, **115**(37):9282-9287.

## Tables

**Table 1. univariate Cox regression analysis table.**

Variables	HR	HR.95L	HR.95H	P-value
CCT6A	4.319148	1.242716	15.01151	0.021344
From	1.016609	0.581309	1.777874	0.953937
Age	1.487818	0.61522	3.598066	0.377879
Gender	0.79254	0.313968	2.000583	0.622606
EFS	0.22656	0.083856	0.612111	0.003413

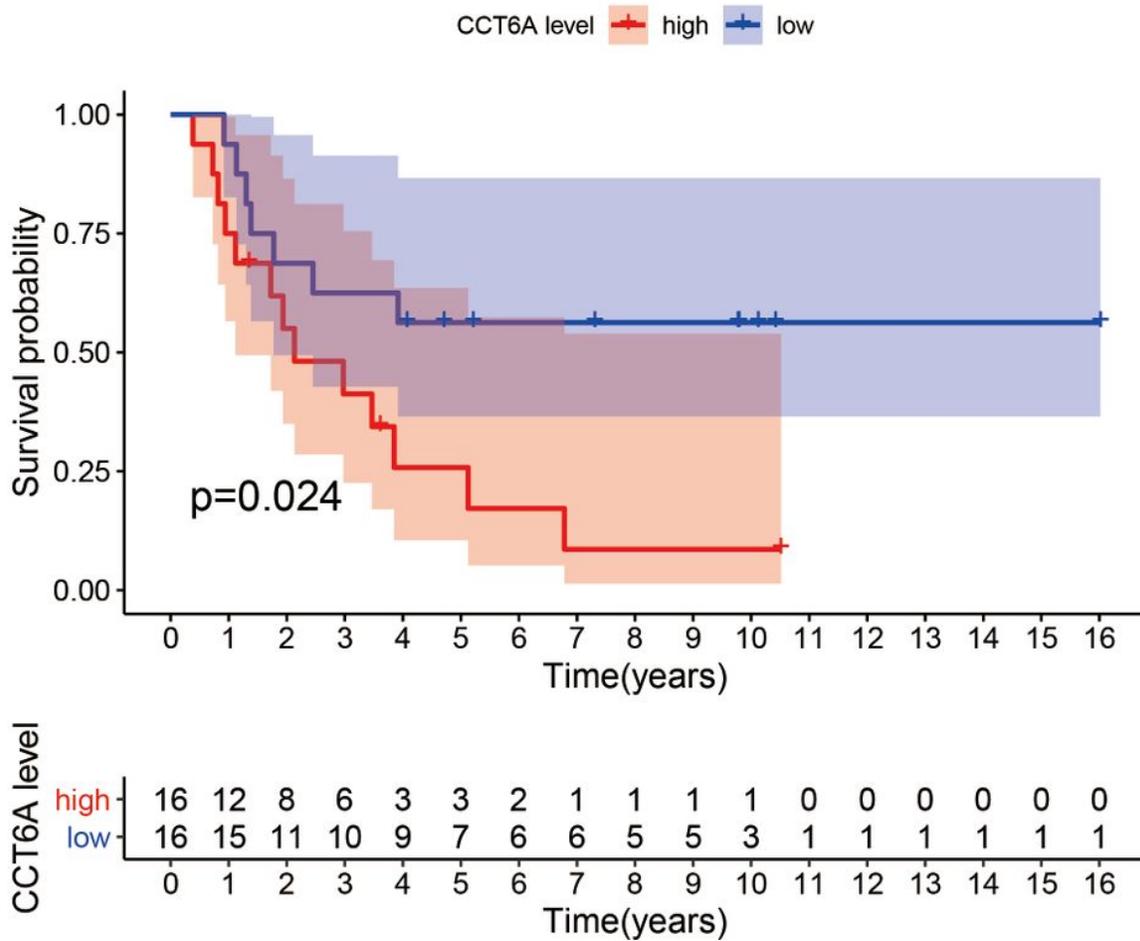
HR represents Hazard ratio, HR.95L represents the lower limit of the 95% confidence interval, and HR.95H represents the upper limit of the 95% confidence interval.

Table 2  
Multivariate Cox regression analysis Table.

Variables	HR	HR.95L	HR.95H	P-value
CCT6A	4.901907	1.354073	17.74549	0.015446
From	0.669709	0.31221	1.436567	0.303189
Age	0.294578	0.100638	0.862259	0.02572
Gender	0.419259	0.147659	1.190434	0.102558
EFS	0.19969	0.073384	0.54339	0.00161

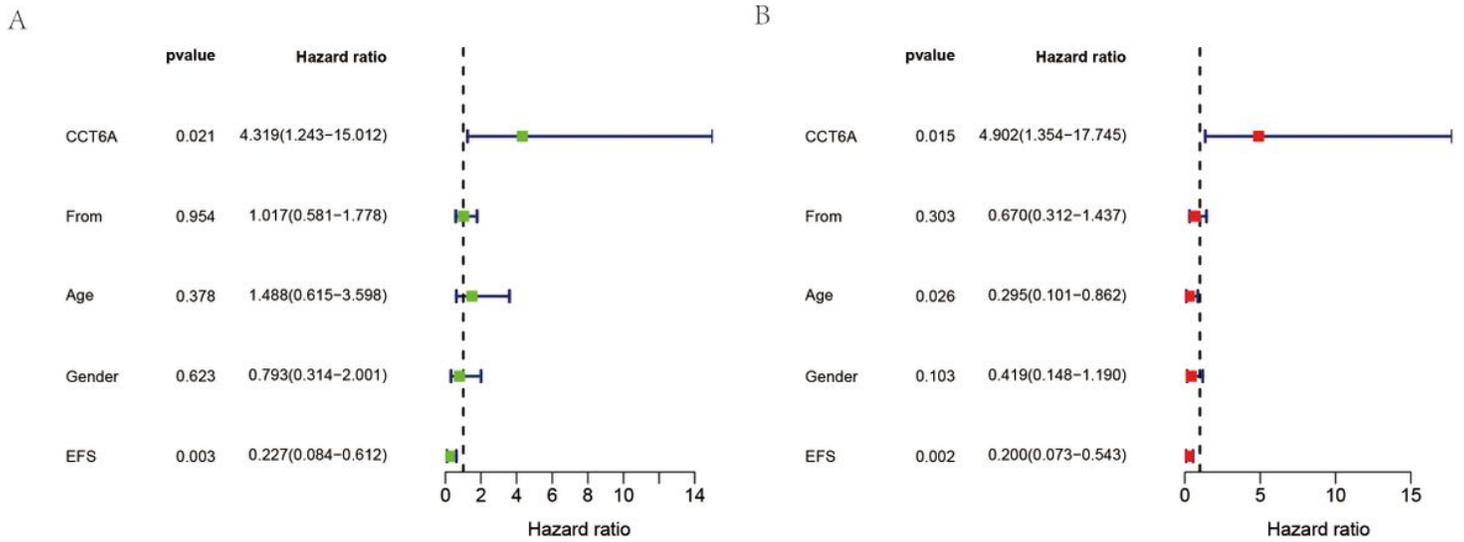
HR represents Hazard ratio, HR.95L represents the lower limit of the 95% confidence interval, and HR.95H represents the upper limit of the 95% confidence interval.

## Figures



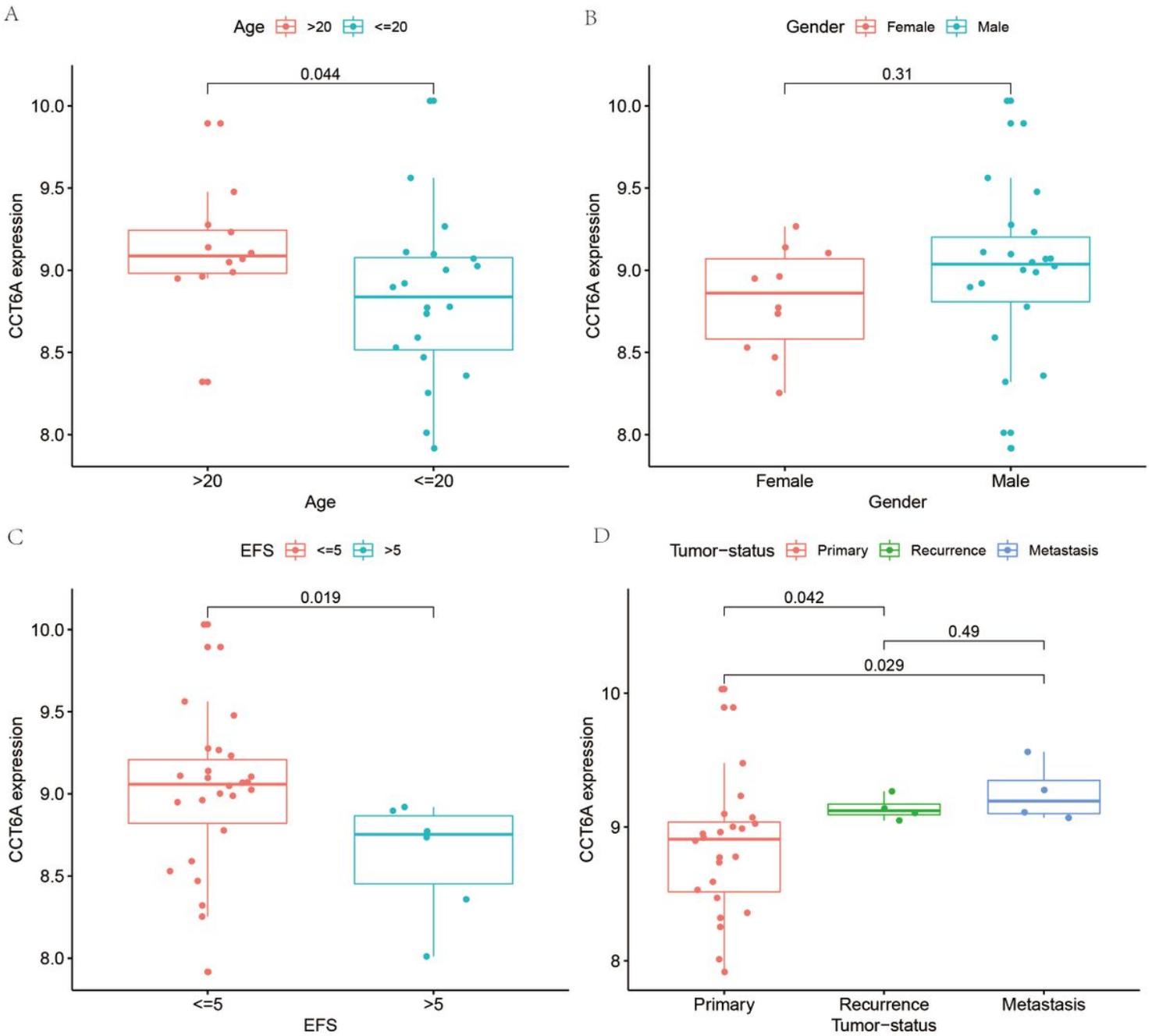
**Figure 1**

CCT6A gene high and low expression group survival curves. Survival curves of patients in the CCT6A high expression group (red curves) and the survival curves of patients in the CCT6A low expression group (blue curves). The x-axis indicates survival time (years) and the y-axis indicates survival probability. Kaplan-Meier survival curves showed that patients in the CCT6A high expression group had worse endpoints for predicting overall survival, whereas patients in the CCT6A low expression group had better outcomes for predicting overall survival compared to high expression, and the difference was statistically significant (P=0.024).



**Figure 2**

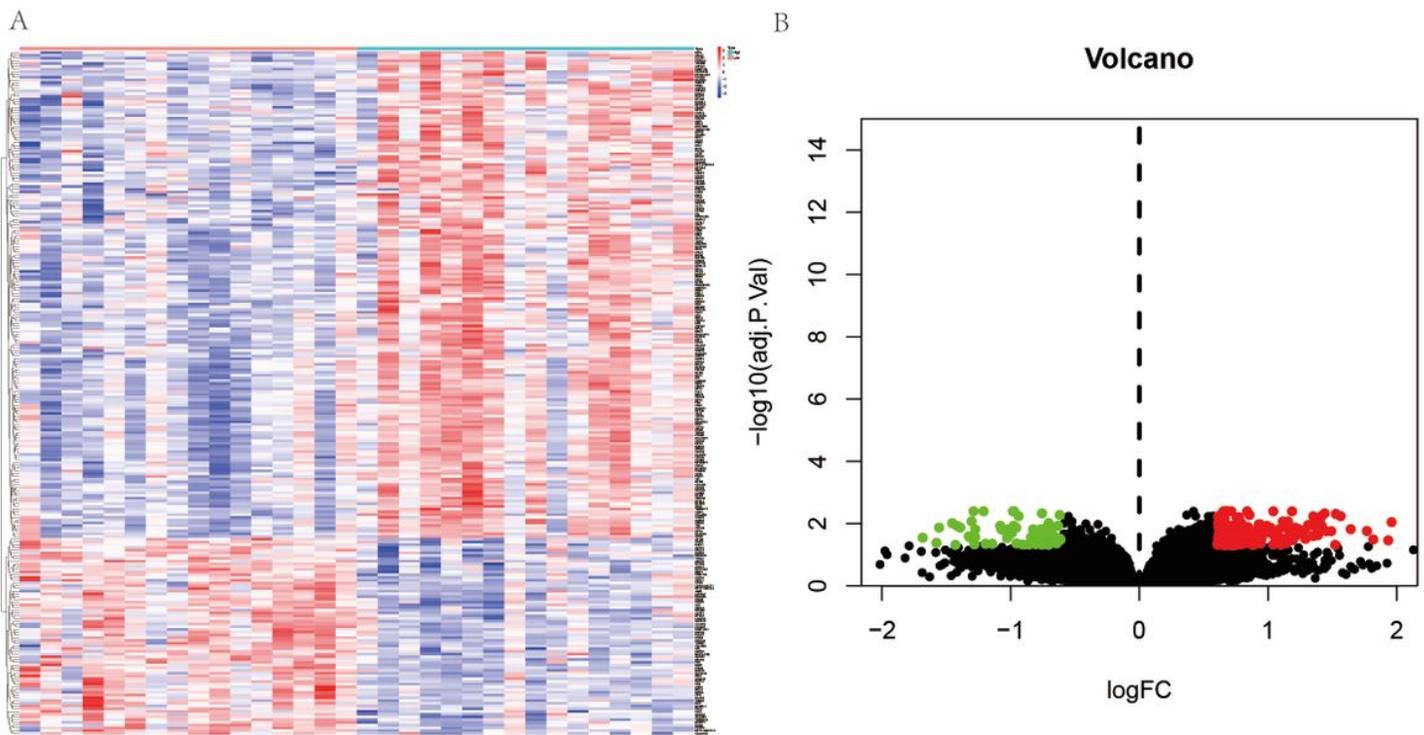
Survival-related Cox regression analysis plots. A and B plots show univariate Cox regression analysis and multivariate Cox regression analysis, respectively. The horizontal coordinates of both graphs represent hazard ratio (HR) and the vertical coordinates are different variables. HR > 1 indicates a high-risk factor and HR < 1 indicates a low-risk factor. In the univariate Cox regression analysis in Figure A, CCT6A and EFS are statistically significant (p-value < 0.05). In the multifactorial Cox regression analysis in Figure B, CCT6A, Age and EFS are statistically significant (p-value < 0.05).



**Figure 3**

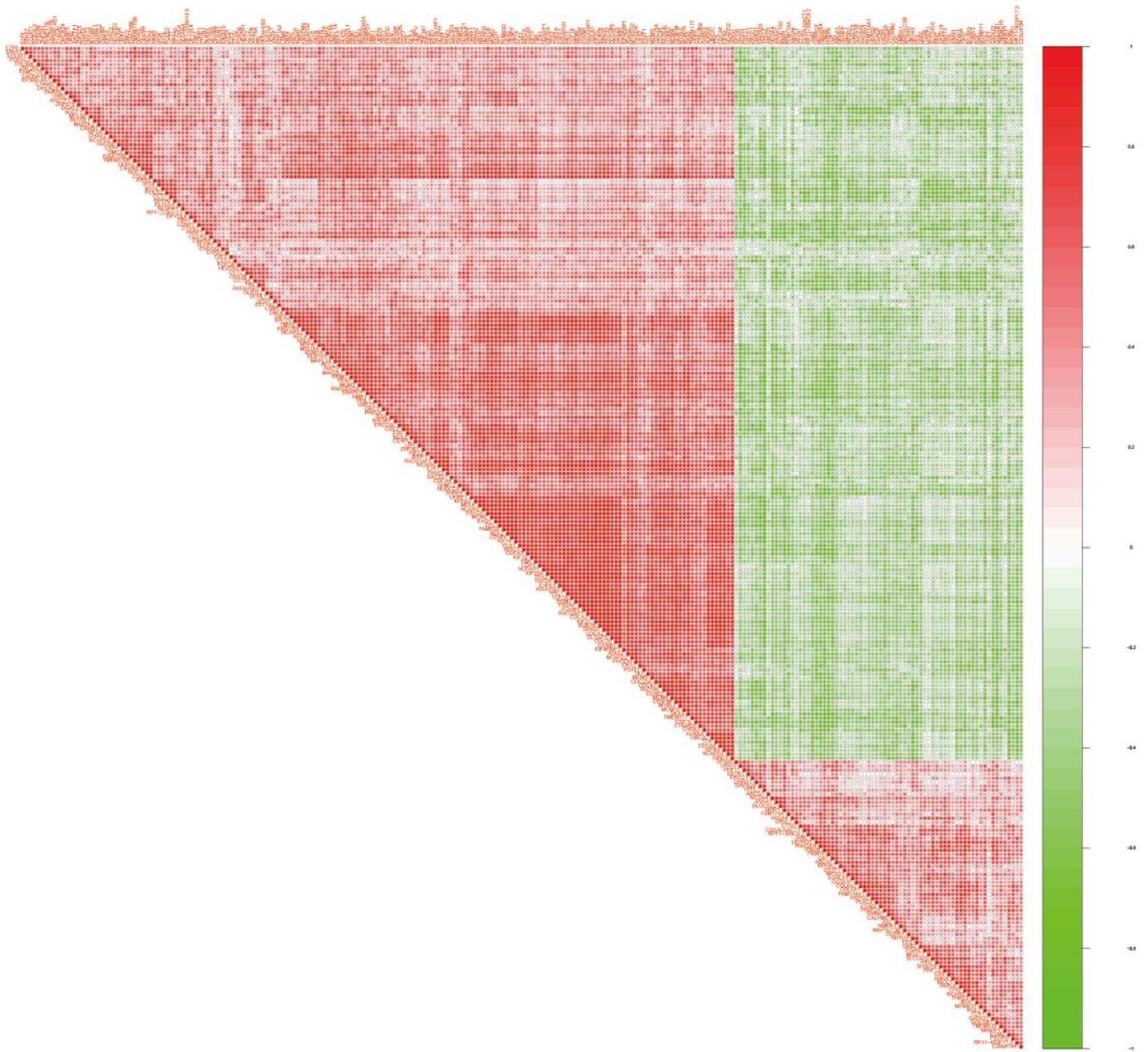
CCT6A gene expression in different subgroups. X-axis all indicate different groupings and Y-axis all indicate the gene expression level of CCT6A. Figure A shows the expression levels of CCT6A in different age groups. The median value of gene expression of CCT6A in the >20-year-old patient group was lower than in the <=20-year-old patient group, and the difference was statistically significant (P-value=0.044). Figure B shows the gene expression of CCT6A in different Genders. The median CCT6A gene expression was greater in the Male group than in the Female group, and the difference was not statistically significant (P-value > 0.05). Figure C shows the expression of CCT6A in different EFS groupings. The median gene expression value of CCT6A was much higher in the EFS <= 5-year group than in the >5-year group, and the difference was statistically significant (P-value=0.019). Figure D shows the gene

expression levels of CCT6A in different tumor sources. The difference between the Primary and Recurrence groups was statistically significant ( $p$ -value=0.042), and the difference between the Primary and Metastasis groups was statistically significant ( $p$ -value=0.029).



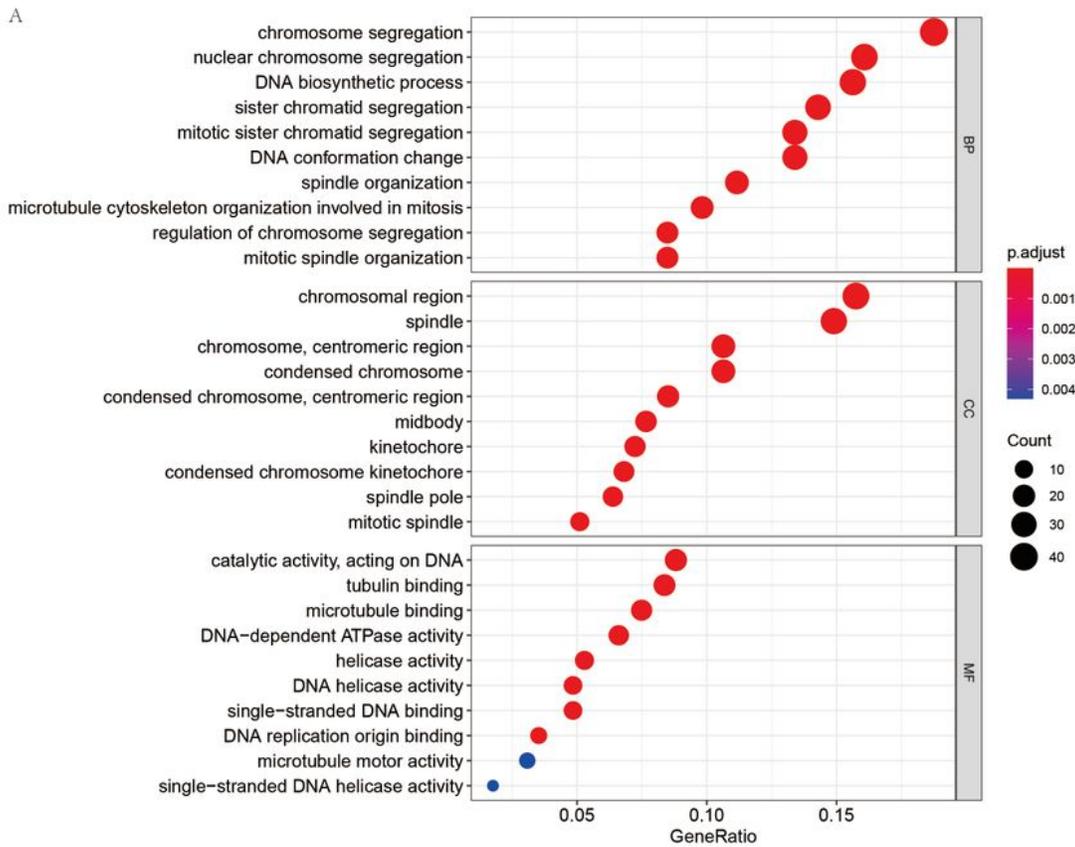
**Figure 4**

Heat map and volcano diagram of differential genes. Figure A shows the heat map of differential genes. Red represents high expression, blue represents low expression, and white represents intermediate expression. There are 178 genes with elevated expression values in the high expression group and 72 genes with decreased expression in the low expression group. Figure B shows the volcano of differential genes. x-axis indicates  $\log_{FC}$  value, y-axis indicates  $-\log_{10}P$ -value. red dots indicate up-regulated genes, green dots indicate down-regulated genes, and black dots indicate genes with insignificant differences.

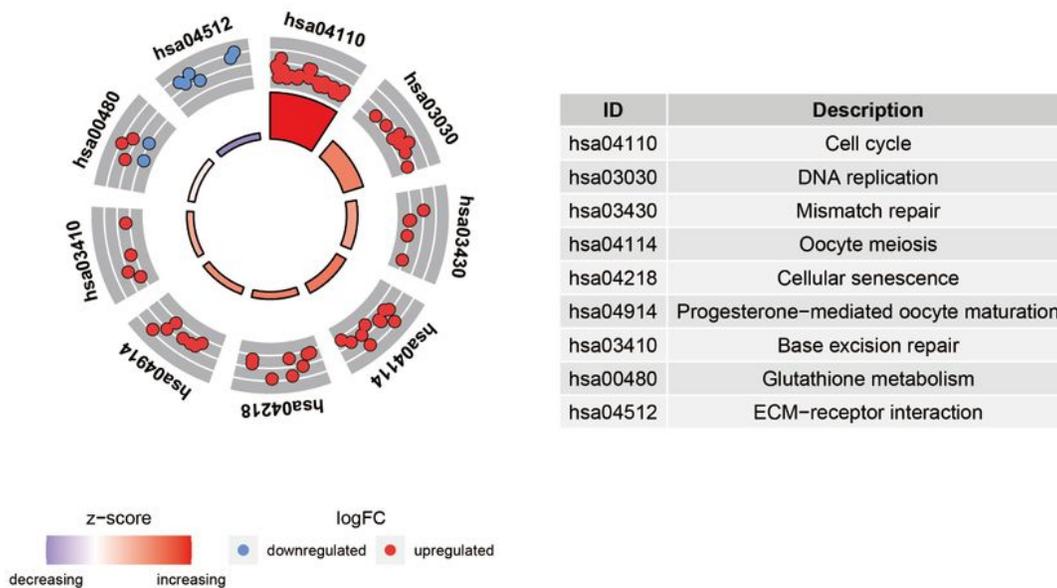


**Figure 5**

Correlation coefficient heat map. Both the right-angled and diagonal sides of the triangle represent genes. The redder the color between the two genes indicates a stronger positive correlation between the two genes; the greener the color between the two genes indicates a stronger negative correlation between the two genes.

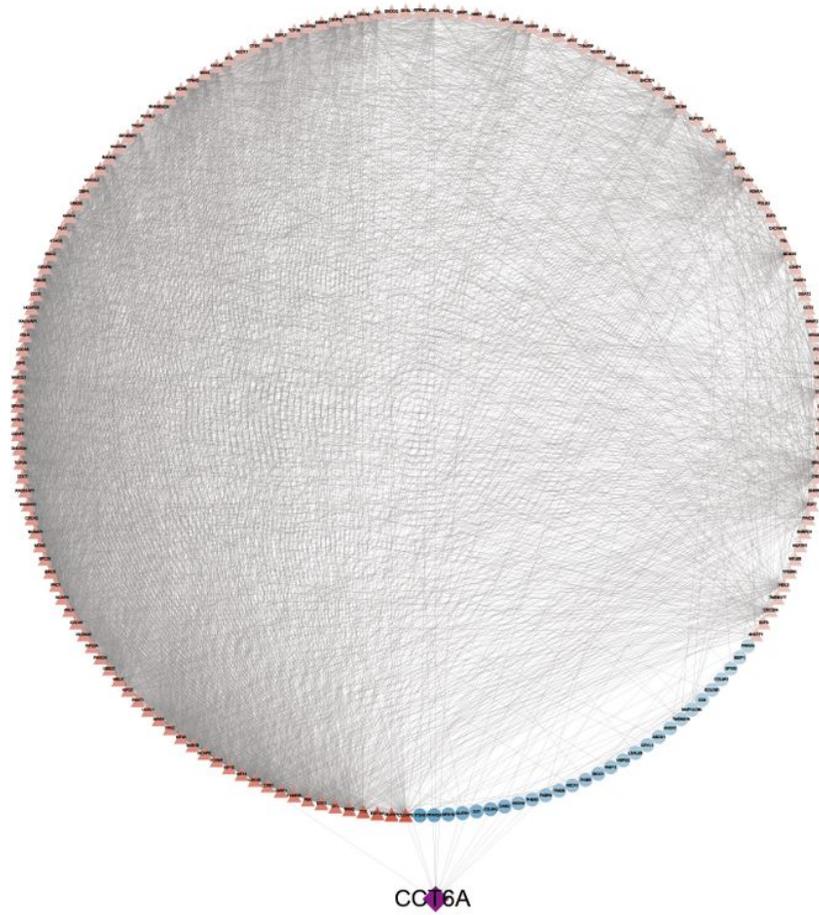


B



**Figure 6**

GO enrichment analysis of differentially expressed genes and KEGG enrichment analysis. Figure A shows the top 30 entries of BP, CC, and MF for GO enrichment analysis of differentially expressed genes. The X-axis shows the gene ratio of each entry, Y-axis shows the GO entries. Figure B shows the enrichment pathway of KEGG. Inner circles indicate z-scores, outer circles indicate gene enrichment pathways, red dots indicate up-regulated genes, and blue dots indicate down-regulated genes.



**Figure 7**

Protein interaction network diagram. Red triangles indicate up-regulated genes, blue circles indicate down-regulated genes, and purple diamonds indicate CCT6A genes.