

# Identification of Hub Genes Associated with Melanoma Development by Comprehensive Bioinformatics Analysis

**Jie Jiang**

Guangxi Medical University

**Guoyong Xu**

Guangxi Medical University

**Tuo Liang**

Guangxi Medical University

**Chaojie Yu**

Guangxi Medical University

**Shian Liao**

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

**Chong Liu**

Guangxi Medical University First Affiliated Hospital

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

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

---

## Research article

**Keywords:** Melanoma, Weighted gene co-expression network analysis, Differential Expression Gene Analysis, Biomarker, Predict prognosis

**Posted Date:** October 9th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-87468/v1>

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

[Read Full License](#)

---

# Abstract

**Introduction:** The aim of this study was to identify important genes associated with melanoma for further development of new target gene therapies and to analyze their significance in relation to prognosis.

**Materials and methods:** Gene expression data for melanoma and normal tissue were downloaded from three databases. Differentially co-expressed genes were identified by WGCNA and DEGs analysis. These genes were subjected to GO and KEGG enrichment analysis as well as construction of the PPI, visualized with Cytoscape, and screened for the top 10 Hub genes using CytoHubba. We verified the protein levels of the Hub gene using the HPA database.

**Results:** A total of 435 differentially co-expressed genes were obtained. Survival curves showed that high expression of FOXM1, EXO1, KIF20A, TPX2 and CDC20 in melanoma patients with 5 of the top 10 hub genes was associated with reduced overall survival (OS). The HPA database results showed that FOXM1, KIF20A, TPX2 and CDC20 showed upregulated expression in melanoma compared to normal tissues.

**Conclusion:** FOXM1, KIF20A, TPX2, and CDC20 are prognosis-associated core genes of melanoma, and their high expression correlates with low prognosis of melanoma patients, and can be used as biomarkers for melanoma diagnosis, treatment, and prognosis prediction.

## Introduction

Melanoma, a highly malignant tumor originating from melanocytes, most commonly occurs in the skin and can evolve from a congenital benign cell nevus or develop from a dysplastic nevus. According to a 2014 article on the epidemiology of melanoma: men are about 1.5 times more likely to develop melanoma than women; the development of malignant melanoma may be associated with changes in external environmental factors (e.g., UV exposure)[1]. In the past few years, the main means of treating melanoma have included: surgery, medication, and radiation therapy[2]. A review of the literature on microsurgery versus extensive local excision for melanoma showed that after controlling for potential confounding variables, patients treated for melanoma with microsurgery were more likely to be alive at five years than those treated for melanoma with extensive local excision[3]. Despite the many treatment options for melanoma, patient survival is still limited.

In recent years, with the rapid development of bioinformatics technology, bioinformatics has become increasingly popular for studying the molecular mechanisms of diseases and discovering disease-specific biomarkers that are increasingly being used to diagnose and treat diseases accurately[4–6]. Weighted gene co-expression network analysis (WGCNA), one of the bioinformatics analysis methods, describes co-expression patterns between genes in microarray samples, providing new insights for predicting the function of co-expressed genes and the development of diseases[7]. Differentially expressed genes (DEGs) are widely used in cancer research and are an excellent method for cancer research[8–10]. It provides a genomics-based method for discovering changes in gene expression levels between experimental and control groups[11]. We could look for potential disease-related biomarkers in

genes with significant differences in gene expression. Thus, we have combined the two approaches by combining genes from relevant modules obtained by the WGCNA method with differentially expressed genes to enhance the discrimination of genes that are expected to be candidate markers of disease.

In this study, to obtain co-expression genes for melanoma, we analyzed melanoma mRNA expression data downloaded from the UCSC database and the GEO database by the WGCNA method and the differential expression gene method. We further identified and validated potential melanoma biomarkers by protein-protein functional interaction network (PPI), survival analysis, and HPA database immunohistochemical staining validation.

## Materials And Methods

### 1. Data Download

Gene expression data for melanoma were downloaded from UCSC Xena (<http://xena.ucsc.edu/>) and the GEO database (<https://www.ncbi.nlm.nih.gov/gds/>). Furthermore, we downloaded all the data and corresponding clinical information on melanoma from the UCSC Xena database free of charge. Data downloaded from the UCSC Xena database contained 471 melanoma samples; gene expression data for 813 normal skin samples were downloaded from the GTEx database (<https://www.gtexportal.org/home/>); the number of melanoma cases for which all clinical information was complete was 322. In total, 55,188 genes were included in our acceptance of our subsequent analysis. On the other hand, we downloaded melanoma sample GSE3189 from the GEO database, a dataset containing 45 melanoma samples, 18 nevus samples, and 7 normal skin samples[12]. We removed 18 moles from the sample. This dataset was studied using the platform GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. Based on the manufacturer's annotation files, we converted the probe files into gene symbols and removed duplicate probes, all expression data were transformed by  $\log_2$ , and the data were standardized. Eventually, a total of 12,549 genes were included for our subsequent analysis.

### 2. Using WGCNA to identify key co-expression modules

WGCNA is a framework for establishing and analyzing weighted gene co-expression networks and is a widely used bioinformatics analysis method that uses the interrelationship between two variables to study biological networks. In this study, all statistical analysis operations were performed based on R (x64 version 4.0.2). We used the R package (WGCNA) to construct a gene co-expression network from the gene expression data of melanoma from UCSC Xena and the gene expression data of melanoma data of GSE3189 from the GEO database[7]. To construct the scale-free network, we use the R command: `softPower = sft$powerEstimate`, which commands R to automatically select the optimal power value, which ends up with a soft threshold of 12 for the UCSC Xena database and 5 for the GEO database. We then construct the adjacency matrix by the following formula:  $a_{ij} = \text{power}(S_{ij}, \beta) = |S_{ij}|^\beta$  ( $a_{ij}$  denotes the adjacency matrix between gene  $i$  and gene  $j$ ;  $S_{ij}$  denotes a similarity matrix completed by Pearson correlations for all gene pairs.;  $\beta$  denotes the soft threshold). Subsequently, we calculated the degree of

dissimilarity between the nodes and converted the adjacency matrix into a TOM matrix. After that, we identified gene networks/modules using a dynamic shear tree algorithm. We correlated previously computed module features with clinical features to further investigate the co-expression network's functional modules. As a result, the modules that were subsequently selected were closely correlated with clinical features and were selected by us for subsequent analyses. More more detailed methods have been elucidated in detail by previous researchers[7].

### **3. Identification of differentially expressed genes and selection of significantly expressed modules**

We used the R package (Limma) to perform differentially expressed genes (DEGs) analysis on gene expression data from both databases[13]. To identify DEGs between melanoma and normal tissues, we performed the differential analysis of gene expression matrices from the two databases using the limma package, respectively. cut off value was set to  $|\logFC| \geq 1$ , adjusted P-value  $< 0.05$ . We then used the R package (pheatmap) to construct heat maps of the DEGs filtered from the two databases; and the R package (ggplot2) to plot the DEGs volcanoes. Next, co-expression genes extracted from the co-expression network overlapped with DEGs were used to identify potential prognostic genes. The overlapping portions of genes were used for subsequent analysis using the R package (VennDiagram) [14] to visualize and plot the genes' overlapping portions into Venn diagrams.

### **4. GO enrichment analysis and KEGG pathway enrichment analysis of target genes**

To explore the overlapping parts of Gene Ontology (GO)[15] and KEGG pathway enrichment analysis[16], we used the R package (clusterProfiler, org.Hs.eg.db, enrichplot, ggplot2)[17] for analysis and visualization. The cut off value is set to p-value  $< 0.05$ . The GO enrichment analysis consists of three main panels, namely Cellular component (CC), Biological process (BP), and Molecular function (MF). KEGG pathway enrichment analysis was mainly enriched in Melanoma, Transcriptional misregulation in cancer.

### **5. Construction of protein-protein interaction networks and screening of Hub genes.**

We used the STRING (<https://string-db.org/>)[18] online database to construct protein-protein interaction (PPI) networks for overlapping partial genes. Subsequently, the resulting PPI network is imported into Cytoscape (v3.8.0) to visualize the PPI. We used a plugin in Cytoscape (CytoHubba)[19] to find Hub genes in these genes. We used the MCC algorithm, and the top 10 genes obtained to the MCC algorithm were used as the central genes for our study.

### **6. The relationship between Hub genes and prognosis, and their expression in various subgroups**

In this study, to verify Hub genes' reliability, we used information from patients whose clinical information was complete for survival analysis. All patients were divided into two groups based on Hub genes' median expression value, with patients greater than or equal to the median value assigned to the high expression group and patients less than the median value assigned to the low expression group. We plotted the Kaplan-Meier univariate survival analysis of overall survival (OS) using the R package (survival, survminer) for the top 10 hub genes. Subsequently, we analyzed the top 10 Hub genes concerning disease-free survival using the online database GEPIA2 (<http://gepia.cancer-pku.cn/>). After

that, we explored the differential expression of Hub genes in cancer and normal tissues, plotting each Hub genes' expression levels in different subgroups between cancer and normal tissues as a box line graph.

## **7. The HPA database validates the protein expression of Hub genes associated with survival**

We used the immunohistochemical staining (IHC) section of The Human Protein Atlas (HPA, <https://www.proteinatlas.org/>) database to identify genes' protein expression between melanoma and normal tissues.

# **Results**

## **1. Construction of weighted gene co-expression modules**

To identify modules associated with prognosis in patients with melanoma, we performed gene co-expression network analysis of gene expression matrices from the UCSC Xena database and GSE3189 using the WGCNA package. There are six co-expression modules constructed from UCSC Xena database expression data (Fig. 1.A) and seven co-expression modules constructed from GSE3189 (Fig. 2.A), not including the grey module cluster to which it is assigned. We developed a heat map module features' relationship, which was used to assess each co-expression module's relationship with two clinical features (normal and cancer). The two modules' features are shown in Fig. 1.B and Fig. 2.B. From the pictures, we can find that the highest correlation between the magenta module in UCSC Xena and the blue module in GSE3189 and normal organization (magenta module:  $r=-0.98$ ,  $P\text{-value} < 1e-200$ ; blue module:  $r=-0.96$ ,  $P\text{-value} < 1e-200$ ).

## **2. Identification of DEGs and gene identification with co-expression modules**

We set the cut-off value for DEGs as  $|\log_{2}FC| \geq 1$ , adjusted  $P\text{-value} < 0.05$ , and a total of 6609 DEGs were identified in the UCSC Xena dataset; heat maps and volcanoes of DEGs are shown in Fig. 3; 6223 DEGs were identified in the GSE3198 dataset; heat maps and volcanoes of differential genes See Fig. 4. Genes from the magenta module in UCSC Xena and genes from the blue module in GSE3189, as well as DEGs from both databases, yielded a total of 435 overlapping genes extracted for subsequent analysis (Fig. 5).

## **3. GO enrichment analysis of 435 genes and KEGG pathway enrichment analysis**

We performed GO enrichment analysis, and KEGG pathway enrichment analysis further explores the 435 genes' potential functions in the overlapping sections. We learned from the GO enrichment analysis that BP was primarily enriched in neutrophil degranulation and neutrophil activation involved in the immune response. CC was mainly enriched in vacuolar and lysosomal membranes. MF was mainly enriched in histone deacetylase binding and integrin-binding (Fig. 6.A). KEGG pathway enrichment analysis was mainly distributed in the melanoma, Transcriptional misregulation in cancer, and Mismatch repair pathways (Fig. 6.B).

## **4. PPI construction and Hub gene identification**

We imported 435 genes from the overlapping parts into the STRING online database to obtain the PPI network. Subsequently, the PPI network was imported into Cytoscape software to visualize the PPI (Fig. 6.A) using the MCC algorithm in the CytoHuba plugin to filter the top 10 Hub genes from the PPI network (Fig. 6.B). We ranked the top 10 Hub genes based on their MCC algorithm scores, which were Aurora Kinase B(AURKB), Exonuclease 1 (EXO1), Kinesin Family Member 20A (KIF20A), TPX2 Microtubule Nucleation Factor (TPX2), Assembly Factor For Spindle Microtubules (ASPM), Mitotic Arrest Deficient 2 Like 1 (MAD2L1), Forkhead Box M1 (FOXM1), Cell Division Cycle 20 (CDC20), Non-SMC Condensin I Complex Subunit H (NCAPH), Baculoviral IAP Repeat Containing 5 (BIRC5).

## 5. Prognostic value of Hub genes and validation of protein expression

We further analyzed the top 10 Hub genes (AURKB, EXO1, KIF20A, TPX2, ASPM, MAD2L1, FOXM1, CDC20, NCAPH, BIRC5) that were screened by the CytoHuba plugin. We analyzed the relationship between the top 10 Hub genes and prognosis separately and plotted univariate survival curves for overall survival using the R package based on the Kaplan-Meier method. Kaplan-Meier analysis showed (Fig. 8) that 5 of the top 10 Hub genes (FOXM1, EXO1, KIF20A, TPX2, CDC20) were significantly correlated with prognosis ( $p < 0.01$ ), and that the relationship between all five hub genes and the prognosis of melanoma patients was such that high expression of the gene was accompanied by low patient Prognosis. Subsequently, we plotted the disease-free survival of the top 10 Hub genes using the GEPIA2 database (Fig. 9) and found that none of the differences were statistically significant ( $P\text{-value} > 0.05$ ). In addition, we validated the four genes with the most significant differences in survival analysis (FOXM1, KIF20A, TPX2, CDC20) according to the HPA database, and the protein levels of FOXM1, KIF20A, TPX2, and CDC20 were increased in melanoma tissues compared with normal tissues (Fig. 11). Unfortunately, EXO1 was unable to find corresponding evidence in the HPA database to identify protein expression of genes between melanoma and normal tissues. Therefore, we confirmed by the above observations that high expression of FOXM1, EXO1, KIF20A, TPX2, and CDC20 is associated with poor prognosis and reduced overall survival in melanoma patients (Fig. 8.F, G, J, D). Subsequently, we verified the expression levels of Hub genes in the UCSC Xena database (Fig. 10).

## Discussion

Melanoma is associated with many factors, is more common in light-skinned races, and has a family history of occurrence. Although melanoma treatment has improved from before, the prognosis for melanoma patients is low due to the lack of precise molecular markers. Therefore, there is an urgent need to identify better and more accurate biomarkers to utilize in the prognosis, diagnosis, and treatment of melanoma. In our study, we used integrated bioinformatics to analyze a total of 435 important genes with co-expression trends identified in the UCSC Xena, GTEx database and the GSE3189 database. These genes were subjected to GO enrichment analysis based on the R package (clusterProfiler), mainly enriched in neutrophil activation involved in immune response, stem cell division, and melanosome. As early as 2011, it was noted that there might be a close relationship between neutrophil activation and cancer [20]. Moreover, the relationship between the stem cell division and melanosome and cancer has also been reported in the literature [21, 22]. Similarly, these genes were subjected to KEGG pathway enrichment

analysis based on the R package (clusterProfiler), and the enrichment results showed that these genes were associated with various cancer pathways, including melanoma, endometrial cancer, prostate cancer, etc.; they were also enriched in transcriptional dysregulation pathways in cancer. In humans, dysregulation of genes such as cofactors and chromatin can lead to many diseases[23]. These genes are even enriched in the melanoma pathway, suggesting that these genes are strongly associated with melanoma. Besides, we screened for the top 10 hub genes associated with melanoma based on how the MCC score was calculated for the CytoHubba plugin in Cytoscape. We also found by analyzing the survival of melanoma patients corresponding to high and low expression of these genes that five of the top 10 hub genes were strongly associated with survival, and all showed that high expression of the genes was associated with a low prognosis in melanoma patients. Finally, we performed immunohistochemical analysis using the HPA database and showed that all four genes showed increased expression in melanoma tumor tissues, whereas their expression was not evident in normal tissues.

FOXM1, also known as Forkhead Box M1, is a gene that encodes a protein that is a transcriptional activator involved in cell proliferation. FOXM1 acts downstream of the PI3K-AKT pathway, the Ras-ERK pathway, the JNK/p38MAPK signaling cascade, and is essential for cell proliferation, differentiation, senescence, DNA damage and repair, and control of the cell cycle[24]. It has also been reported that FOXM1 was overexpressed in a variety of human cancers and that the oncogenic potential of this gene is based on its ability to reactivate target genes involved in different stages of cancer development[25]. It has been shown that the positive feedback of FOXM1 promotes the growth and invasion of gastric cancer and that FOXM1 promotes gastric cancer progression by interacting with PVT1[26]. FOXM1 has also been reported in non-serious epithelial ovarian carcinoma: FOXM1 was upregulated in all epithelial ovarian cancers[27]. In addition, it has been shown that the FOXM1-PSMB4 axis can play a catalytic role in the proliferation and development of cervical cancer[28]. More surprisingly, FOXM1 plays a vital role in many other cancers[29–32]. In our study, FOXM1 was upregulated in tumor tissues compared to normal tissues, suggesting a significant correlation with melanoma. Previous studies have shown that higher levels of FOXM1 in tumor tissues have been strongly associated with poor prognosis in melanoma patients, consistent with our study[33–35]. Kinesin Family Member 20A (KIF20A) is also a protein-encoding gene, and what is known about the diseases associated with this gene is mainly familial restriction Familial Isolated Restrictive Cardiomyopathy and Charcot- Marie-Tooth Disease, Type 4C. Research has also been conducted on the role of this gene in cancer. It has been reported that patients with bladder cancer with high expression of KIF20A have poorer tumor stages and that KIF20A promotes metastasis and proliferation of bladder cancer cells[36]. Also, it has been shown that skin tumor thickness in KIF20A-positive patients with primary melanoma is significantly greater than skin tumor thickness in patients negative for this gene and that KIF20A-positive patients are more likely to relapse earlier[37]. It is well known that while tumor recurrence has a very significant relationship with patient prognosis, this indirectly suggests that KIF20A is associated with survival in melanoma patients, which is consistent with the results of our study. The TPX2 Microtubule Nucleation Factor (TPX2) is a protein-encoding gene. The main diseases known to be associated with the TPX2 gene include Capillary Leak

Syndrome and Colorectal Cancer. It has been shown that activation of TPX2 expression increases the invasion and proliferation of cervical cancer, promoting cancer development[38]. A study of TPX2 in esophageal cancer showed that the 5-year survival rate of esophageal cancer patients with concomitant high TPX2 expression levels was significantly lower than that of esophageal cancer patients with low TPX2 expression levels[39]. Interestingly, in our study, patients with high TPX2 expression of melanoma had a relatively shorter overall survival than patients with low expression. Cell Division Cycle 20 (CDC20) is also a protein-coding gene. The main diseases known to be associated with this gene are Ceroid Lipofuscinosis, Neuronal, 2. Back in 2015, there were reports that CDC20 could be used as a novel cancer treatment modality[40]. In hepatocellular carcinoma, the upregulation of CDC20 expression predicted a decline in overall survival and disease-free survival[41]. In our study, melanoma patients with high expression of CDC20 had a lower survival time than patients with low expression. In summary, the expression of the four hub genes we studied were all strongly associated with cancer, and in our study, high levels of expression of these genes were accompanied by shorter survival times for melanoma patients.

Our study, like others, has limitations regarding the different tumor types. Although we identified potential prognostic genes between melanoma and normal tissue using three different sources of databases with two different bioinformatics analyses, it was less accurate for each of the different subtypes of melanoma patients. Furthermore, EXO1 failed to find corresponding evidence in the HPA database to identify protein expression of genes between melanoma and normal tissues. The Hub genes associated with survival that can influence melanoma patients' prognosis should be further validated by a series of experiments on molecular mechanisms.

In conclusion, by combining the WGCNA analysis method with differentially expressed gene analysis, our study identified the genes FOXM1, KIF20A, TPX2, and CDC20, which are highly correlated with survival melanoma patients and have the potential to serve as a prognostic biomarker in melanoma.

## Conclusions

FOXM1, KIF20A, TPX2, CDC20 are core melanoma prognostic genes, and their high expression is strongly associated with low prognosis in melanoma patients. FOXM1, KIF20A, TPX2, CDC20 can be used as biomarkers for melanoma diagnosis, treatment, and prognosis prediction.

## List Of Abbreviations

Weighted gene co-expression network analysis: WGCNA

Differentially expressed genes: DEGs

protein-protein functional interaction network: PPI

Gene Ontology: GO

Cellular component: CC

Biological process: BP

Molecular function: MF

overall survival: OS

## Declarations

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

**Consent for publication:** All co-authors of this study 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/>), UCSC Xena (<http://xena.ucsc.edu/>) and GTEx database (<https://www.gtexportal.org/home/>).

**Competing interests:** There were no conflicts of interest between all co-authors involved in this study.

**Funding:** Not applicable.

**Authors' contributions:** Jie Jiang, Chong Liu and Xinli Zhan designed the study. Guoyong Xu, Tuo Liang, Shian Liao and Chaojie Yu analyze the data. Zide Zhang, Zhaojun Lu, Zequn wang, Jiarui Chen, Tianyou Chen and Hao Li digital visualization. Jie Jiang wrote and revised the manuscript. Chong Liu and Xinli Zhan revised manuscript. All authors read and approved the final manuscript.

**Acknowledgments:** Thanks to all the co-authors and staff who contributed to this research for their hard work.

## References

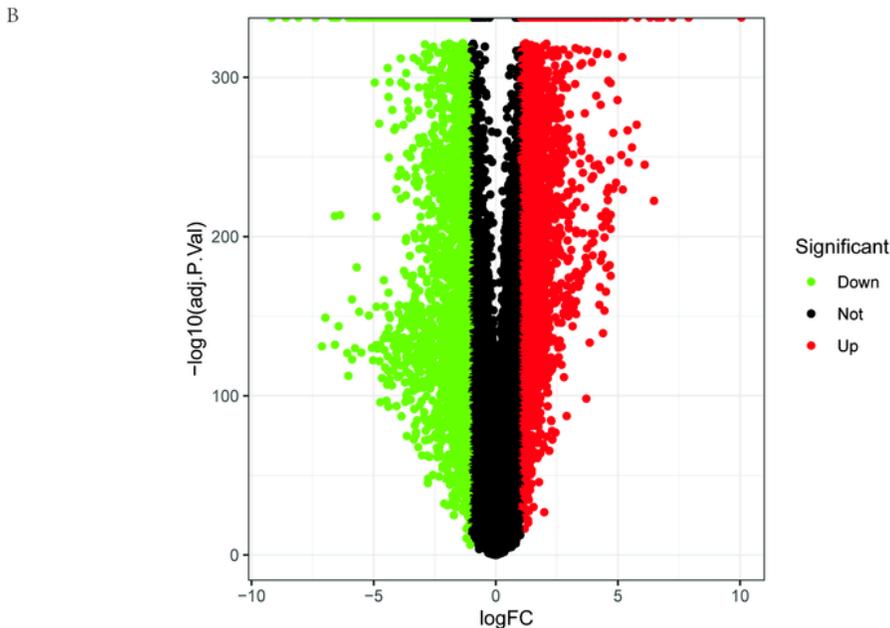
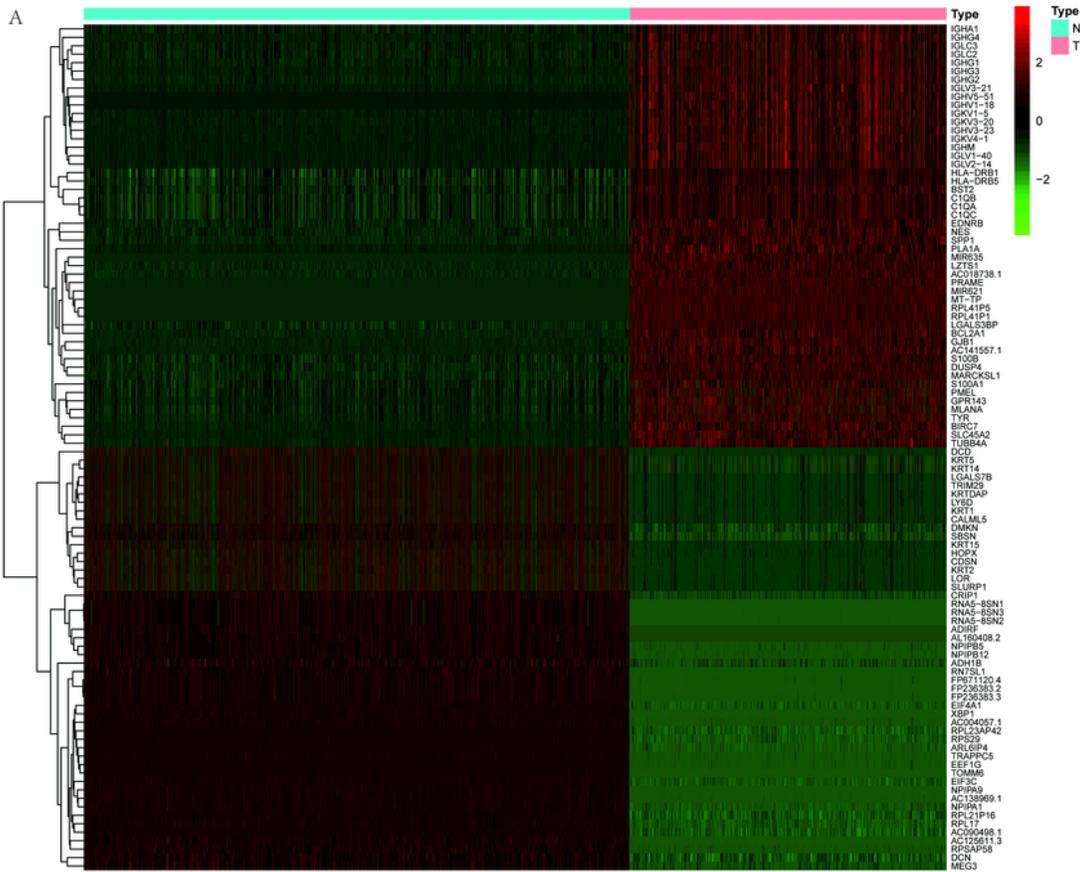
1. Rastrelli M, Tropea S, Rossi CR, Alaibac M: **Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification.** *In Vivo* 2014, **28**(6):1005-1011.
2. Pavri SN, Clune J, Ariyan S, Narayan D: **Malignant Melanoma: Beyond the Basics.** *Plast Reconstr Surg* 2016, **138**(2):330e-340e.
3. Hanson J, Demer A, Liszewski W, Foman N, Maher I: **Improved overall survival of melanoma of the head and neck treated with Mohs micrographic surgery versus wide local excision.** *J Am Acad Dermatol* 2020, **82**(1):149-155.
4. Best MG, Wesseling P, Wurdinger T: **Tumor-Educated Platelets as a Noninvasive Biomarker Source for Cancer Detection and Progression Monitoring.** *Cancer Res* 2018, **78**(13):3407-3412.
5. Chu H-W, Chang K-P, Hsu C-W, Chang IY-F, Liu H-P, Chen Y-T, Wu C-C: **Identification of Salivary Biomarkers for Oral Cancer Detection with Untargeted and Targeted Quantitative Proteomics**

- Approaches.** *Mol Cell Proteomics* 2019, **18**(9):1796-1806.
6. Shan C, Zhang Y, Hao X, Gao J, Chen X, Wang K: **Biogenesis, functions and clinical significance of circRNAs in gastric cancer.** *Mol Cancer* 2019, **18**(1):136.
  7. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC bioinformatics* 2008, **9**:559.
  8. Unfried JP, Serrano G, Suárez B, Sangro P, Ferretti V, Prior C, Boix L, Bruix J, Sangro B, Segura V *et al*: **Identification of Coding and Long Noncoding RNAs Differentially Expressed in Tumors and Preferentially Expressed in Healthy Tissues.** *Cancer Res* 2019, **79**(20):5167-5180.
  9. Ma Q, Xu Y, Liao H, Cai Y, Xu L, Xiao D, Liu C, Pu W, Zhong X, Guo X: **Identification and validation of key genes associated with non-small-cell lung cancer.** *Journal of cellular physiology* 2019, **234**(12):22742-22752.
  10. Grossi E, Raimondi I, Goñi E, González J, Marchese FP, Chapaprieta V, Martín-Subero JI, Guo S, Huarte M: **A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions.** *Nat Commun* 2020, **11**(1):936.
  11. Segundo-Val IS, Sanz-Lozano CS: **Introduction to the Gene Expression Analysis.** *Methods Mol Biol* 2016, **1434**:29-43.
  12. Talantov D, Mazumder A, Yu JX, Briggs T, Jiang Y, Backus J, Atkins D, Wang Y: **Novel genes associated with malignant melanoma but not benign melanocytic lesions.** *Clin Cancer Res* 2005, **11**(20):7234-7242.
  13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**(7):e47.
  14. Chen H, Boutros PC: **VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R.** *BMC bioinformatics* 2011, **12**:35.
  15. Gaudet P, Dessimoz C: **Gene Ontology: Pitfalls, Biases, and Remedies.** *Methods Mol Biol* 2017, **1446**:189-205.
  16. 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.
  17. 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.
  18. 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.
  19. 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.
  20. Mantovani A, Cassatella MA, Costantini C, Jaillon S: **Neutrophils in the activation and regulation of innate and adaptive immunity.** *Nat Rev Immunol* 2011, **11**(8):519-531.

21. Biton M, Haber AL, Rogel N, Burgin G, Beyaz S, Schnell A, Ashenberg O, Su CW, Smillie C, Shekhar K *et al*: **T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation**. *Cell* 2018, **175**(5):1307-1320 e1322.
22. Dror S, Sander L, Schwartz H, Sheinboim D, Barzilai A, Dishon Y, Apcher S, Golan T, Greenberger S, Barshack I *et al*: **Melanoma miRNA trafficking controls tumour primary niche formation**. 2016, **18**(9):1006-1017.
23. Lee TI, Young RA: **Transcriptional regulation and its misregulation in disease**. *Cell* 2013, **152**(6):1237-1251.
24. Yao S, Fan LY, Lam EW: **The FOXO3-FOXM1 axis: A key cancer drug target and a modulator of cancer drug resistance**. *Semin Cancer Biol* 2018, **50**:77-89.
25. Gartel AL: **FOXM1 in Cancer: Interactions and Vulnerabilities**. *Cancer Res* 2017, **77**(12):3135-3139.
26. Xu M, Wang Y, Weng W, Wei P, Qi P, Zhang Q, Tan C, Ni S, Dong L, Yang Y *et al*: **PVT1A Positive Feedback Loop of lncRNA- and FOXM1 Facilitates Gastric Cancer Growth and Invasion**. 2017, **23**(8):2071-2080.
27. Tassi R, Todeschini P, Siegel E, Calza S, Cappella P, Ardighieri L, Cadei M, Bugatti M, Romani C, Bandiera E *et al*: **FOXM1 expression is significantly associated with chemotherapy resistance and adverse prognosis in non-serous epithelial ovarian cancer patients**. 2017, **36**(1):63.
28. Zhou D, Liu J, Liu F, Luo G, Li H, Zhang R, Chen B, Hua WJB, communications br: **A novel FoxM1-PSMB4 axis contributes to proliferation and progression of cervical cancer**. 2020, **521**(3):746-752.
29. Li X, Wu H, Mao X, Jiang L, Wang YJB, communications br: **USP5 promotes tumorigenesis and progression of pancreatic cancer by stabilizing FoxM1 protein**. 2017, **492**(1):48-54.
30. Cui J, Shi M, Xie D, Wei D, Jia Z, Zheng S, Gao Y, Huang S, Xie KJCcraojotAAfCR: **FOXM1 promotes the warburg effect and pancreatic cancer progression via transactivation of LDHA expression**. 2014, **20**(10):2595-2606.
31. Zhou Z, Chen H, Xie R, Wang H, Li S, Xu Q, Xu N, Cheng Q, Qian Y, Huang R *et al*: **Epigenetically modulated FOXM1 suppresses dendritic cell maturation in pancreatic cancer and colon cancer**. 2019, **13**(4):873-893.
32. Arceci A, Bonacci T, Wang X, Stewart K, Damrauer J, Hoadley K, Emanuele MJCr: **FOXM1 Deubiquitination by USP21 Regulates Cell Cycle Progression and Paclitaxel Sensitivity in Basal-like Breast Cancer**. 2019, **26**(11):3076-3086.e3076.
33. Puig-Butille J, Vinyals A, Ferreres J, Aguilera P, Cabré E, Tell-Martí G, Marcoval J, Mateo F, Palomero L, Badenas C *et al*: **AURKA Overexpression Is Driven by FOXM1 and MAPK/ERK Activation in Melanoma Cells Harboring BRAF or NRAS Mutations: Impact on Melanoma Prognosis and Therapy**. 2017, **137**(6):1297-1310.
34. Kruiswijk F, Hasenfuss S, Sivapatham R, Baar M, Putavet D, Naipal K, van den Broek N, Kruit W, van der Spek P, van Gent D *et al*: **Targeted inhibition of metastatic melanoma through interference with Pin1-FOXM1 signaling**. 2016, **35**(17):2166-2177.

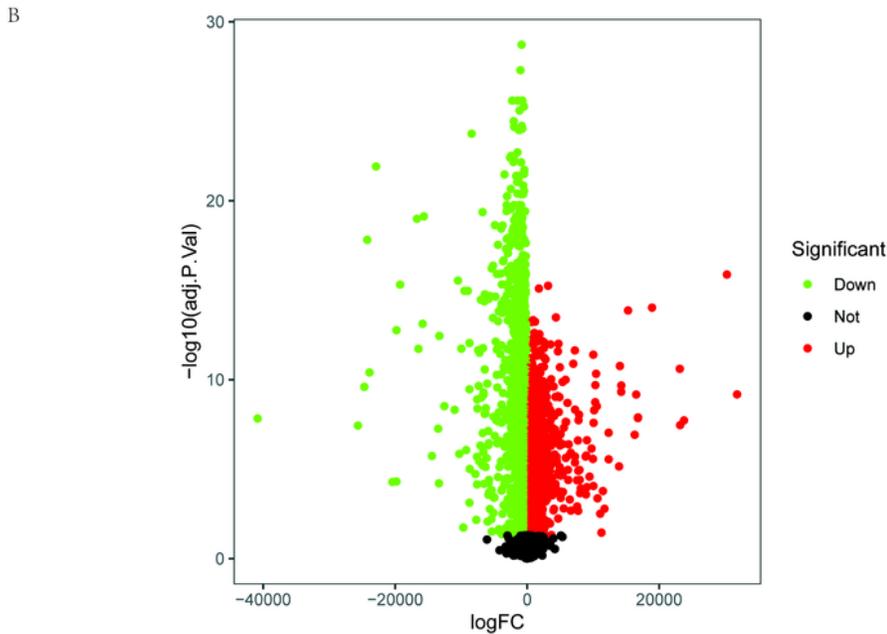
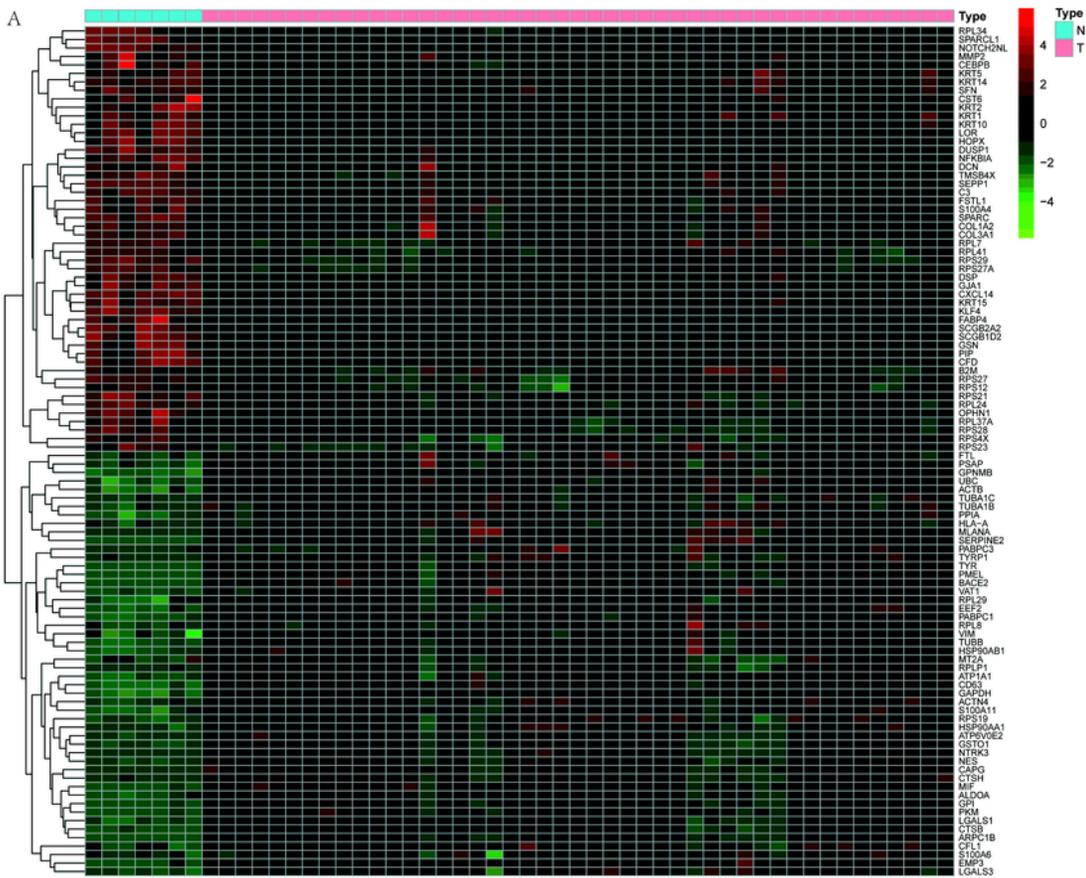
35. Kuzu O, Gowda R, Sharma A, Noory M, Kardos G, Madhunapantula S, Drabick J, Robertson GJCb, therapy: **Identification of WEE1 as a target to make AKT inhibition more effective in melanoma.** 2018, **19(1):53-62.**
36. Shen T, Yang L, Zhang Z, Yu J, Dai L, Gao M, Shang Z, Niu YJdM: **KIF20A Affects the Prognosis of Bladder Cancer by Promoting the Proliferation and Metastasis of Bladder Cancer Cells.** 2019, **2019:4863182.**
37. Yamashita J, Fukushima S, Jinnin M, Honda N, Makino K, Sakai K, Masuguchi S, Inoue Y, Ihn HJAd-v: **Kinesin family member 20A is a novel melanoma-associated antigen.** 2012, **92(6):593-597.**
38. Song T, Xu A, Zhang Z, Gao F, Zhao L, Chen X, Gao J, Kong XJJocp: **CircRNA hsa\_circRNA\_101996 increases cervical cancer proliferation and invasion through activating TPX2 expression by restraining miR-8075.** 2019, **234(8):14296-14305.**
39. Sui C, Song Z, Yu H, Wang HJol: **Prognostic significance of TPX2 and NIBP in esophageal cancer.** 2019, **18(4):4221-4229.**
40. Wang L, Zhang J, Wan L, Zhou X, Wang Z, Wei WJP, therapeutics: **Targeting Cdc20 as a novel cancer therapeutic strategy.** 2015, **151:141-151.**
41. Zhuang L, Yang Z, Meng ZJBri: **Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in Tumor Tissues Predicted Worse Overall Survival and Disease-Free Survival in Hepatocellular Carcinoma Patients.** 2018, **2018:7897346.**

## Figures



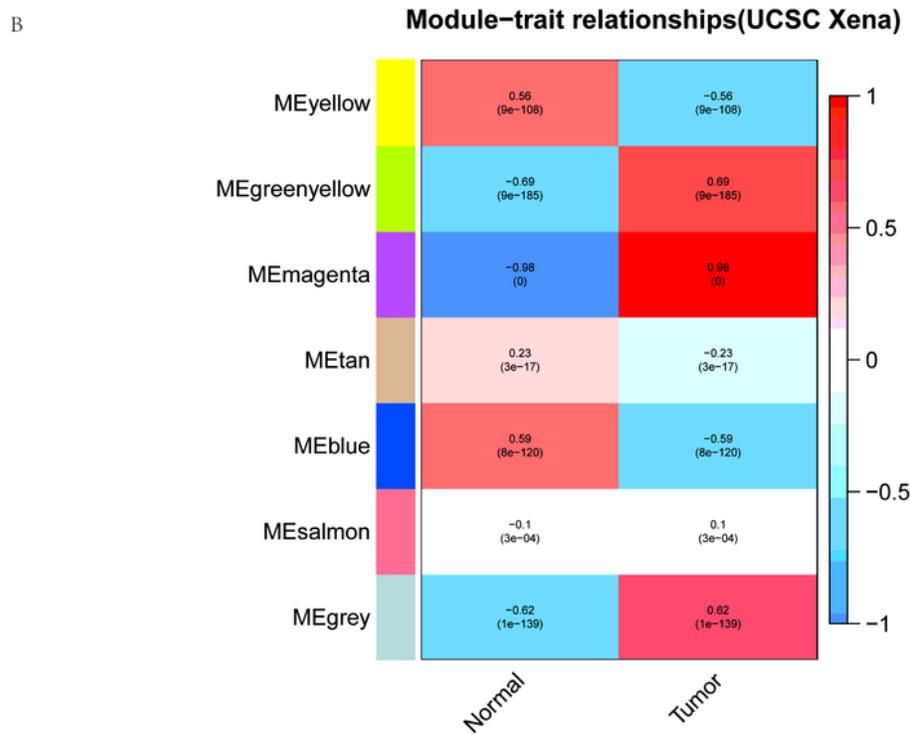
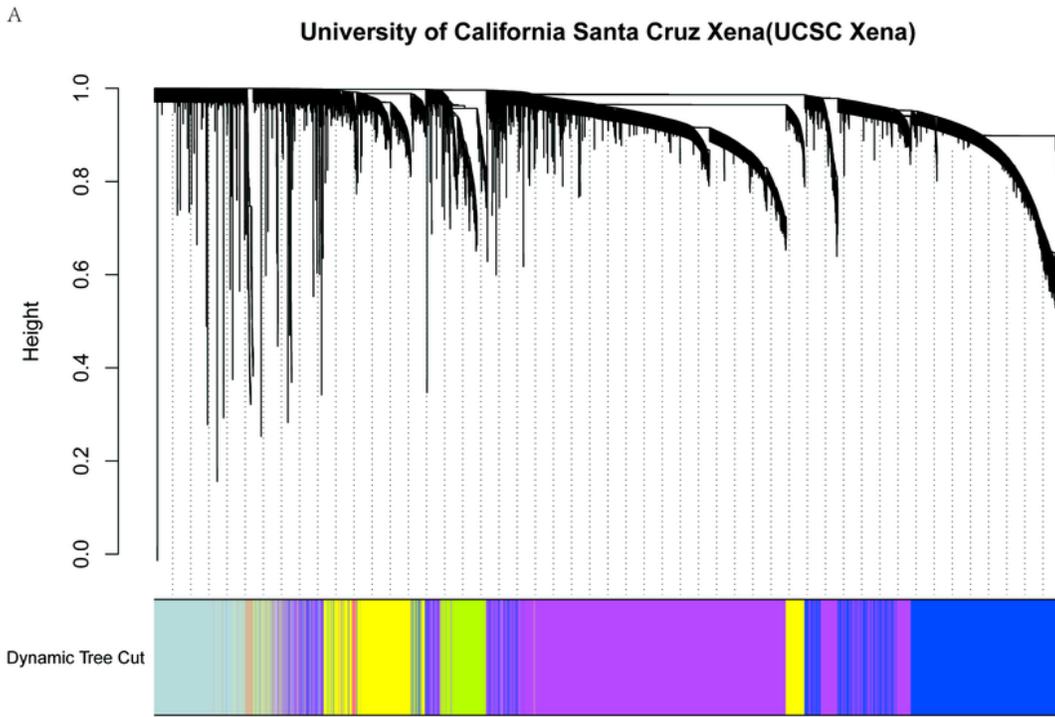
**Figure 1**

Heat map and volcano map of DEGs from the UCSC Xena database. (A) Heat map of the top 50 up-regulated and top 50 down-regulated DEGs obtained from the identification. The red part indicates up-regulated genes, and the green part indicates down-regulated genes. (B) Volcano plot with cut off value set to  $|\logFC| > 1, P\text{-value} < 0.05$ . Red dots indicate up-regulated genes, green dots indicate down-regulated genes, and black represents non-significant genes.



**Figure 2**

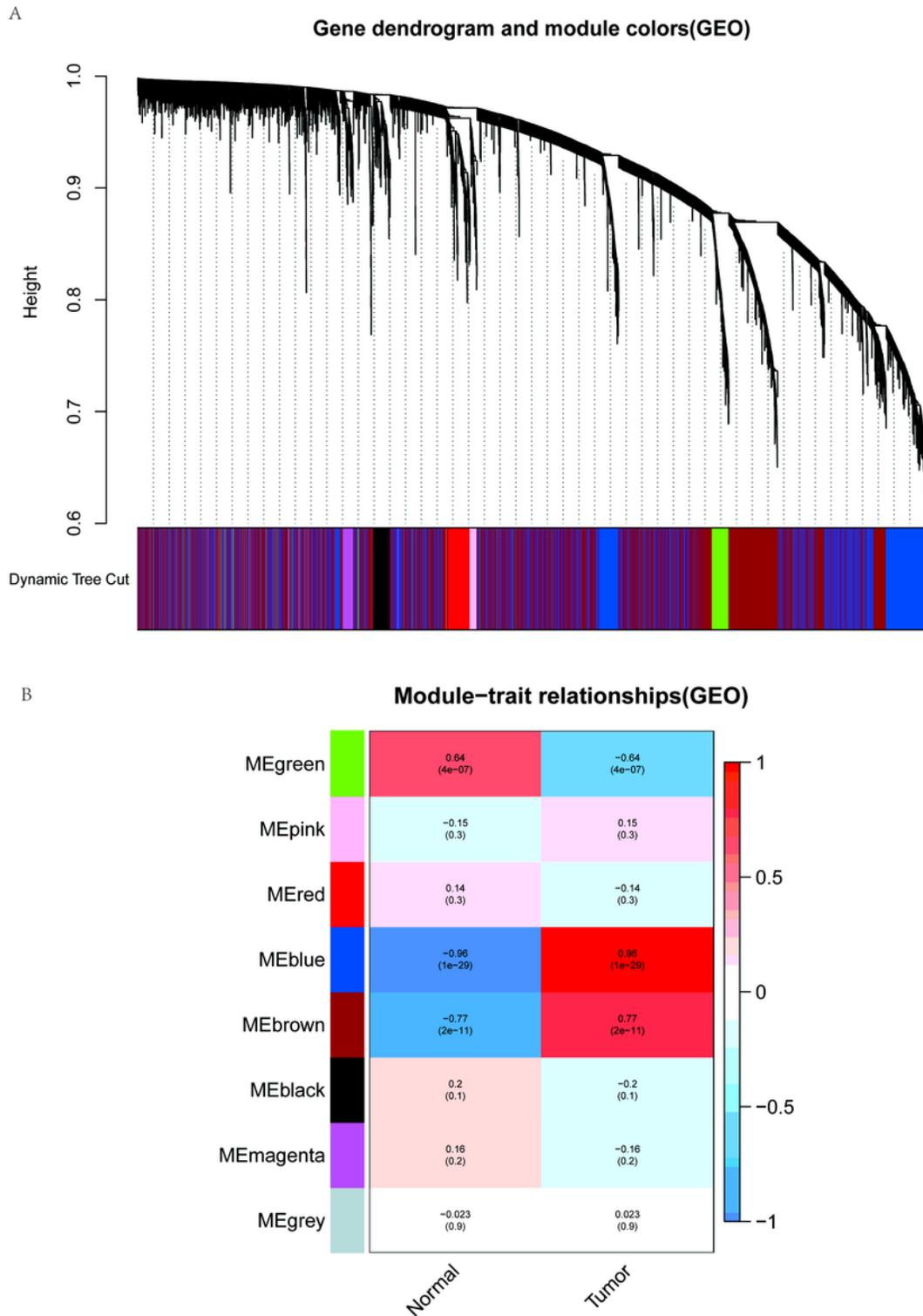
Heat map and volcano map of GSE3189. (A) Heat map of the top 50 up-regulated and top 50 down-regulated DEGs obtained from the identification. The red part indicates up-regulated genes, and the green part indicates down-regulated genes. (B) Volcano plot with cut off value set to  $|\log_{FC}| > 1, P\text{-value} < 0.05$ . Red dots indicate up-regulated genes, green dots indicate down-regulated genes, and black represents non-significant genes.



**Figure 3**

Identification of modules associated with clinical information in the UCSC Xena database. (A) Clustered tree diagram of co-expression network modules sorted by gene-level clustering of matrices obtained by Equation 1-TOM. Each color represents a different co-expressed gene. (B) Relationship diagram of the features of the modules. Each row corresponds to a color module, and each column corresponds to a

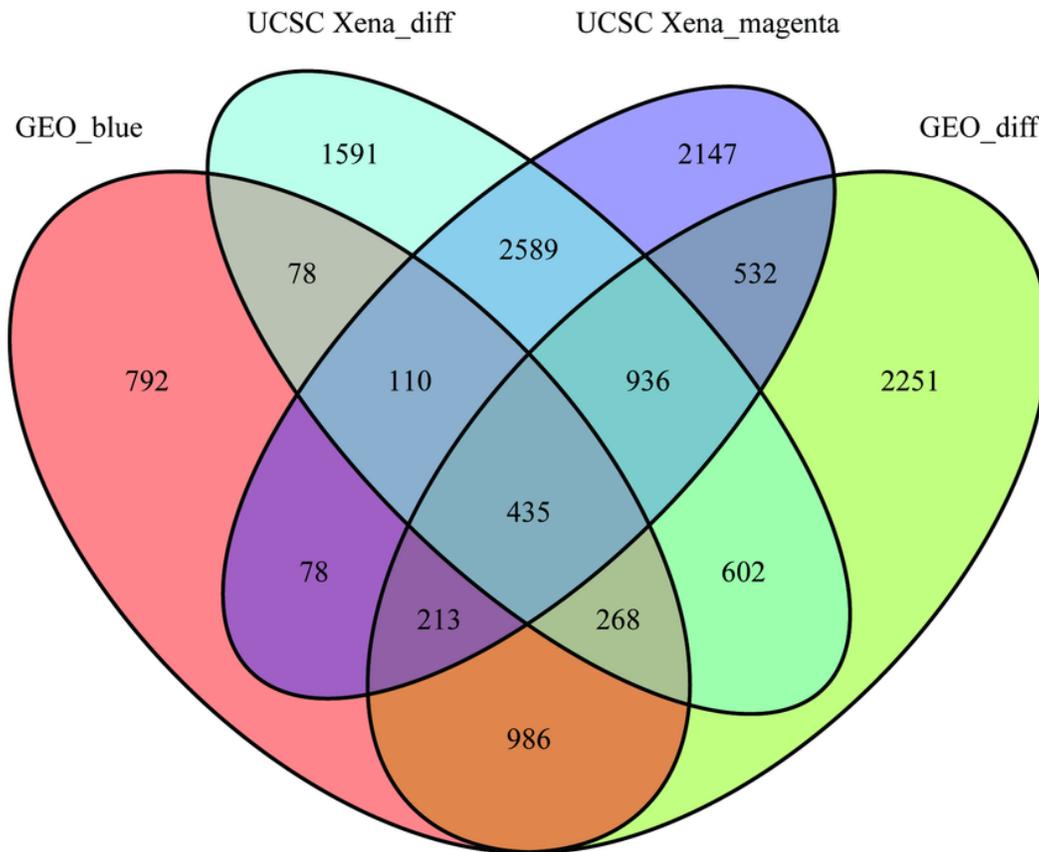
clinical feature (normal and cancer). Each cell contains the correlation and P-value of the corresponding module.



**Figure 4**

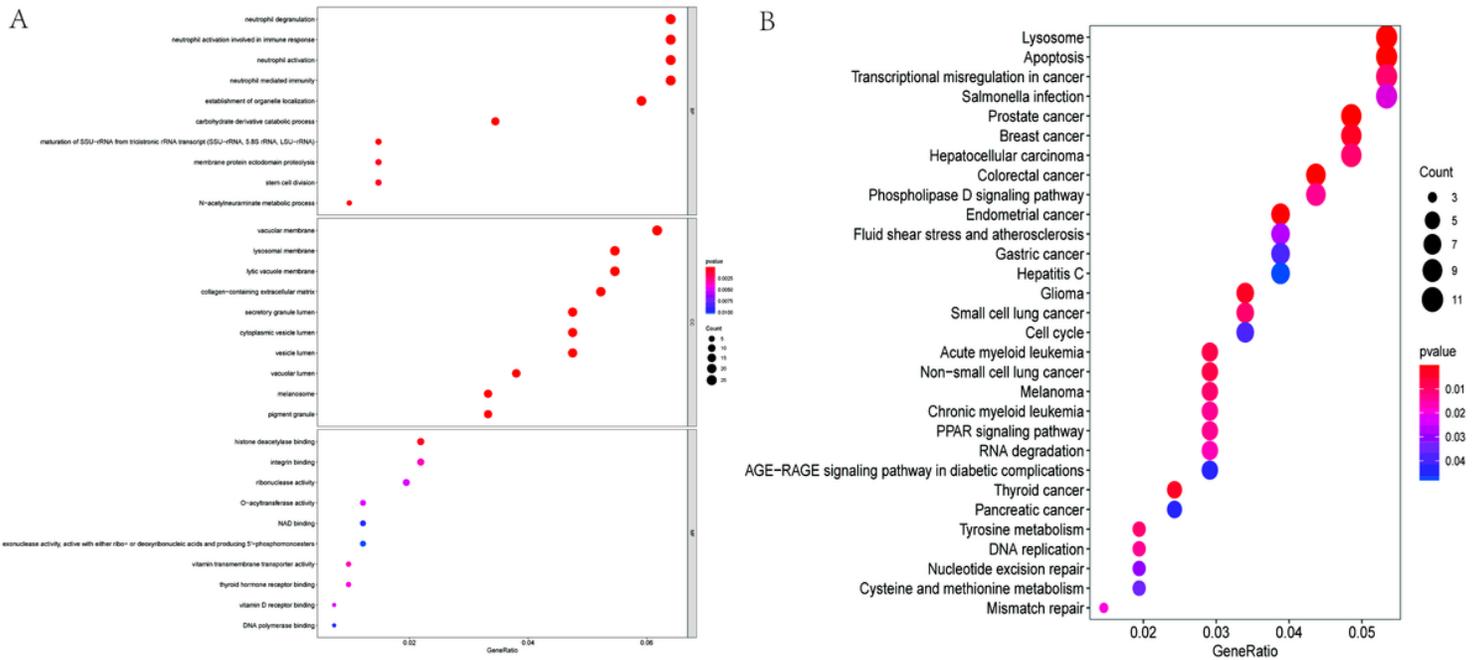
Identification of modules related to clinical information in the GSE3189 database. (A) Clustered tree diagram of co-expression network modules sorted by gene-level clustering of matrices obtained by Equation 1-TOM. Each color represents a different co-expressed gene. (B) Relationship diagram of the

features of the modules. Each row corresponds to a color module, and each column corresponds to a clinical feature (normal and cancer). Each cell contains the correlation and P-value of the corresponding module.



**Figure 5**

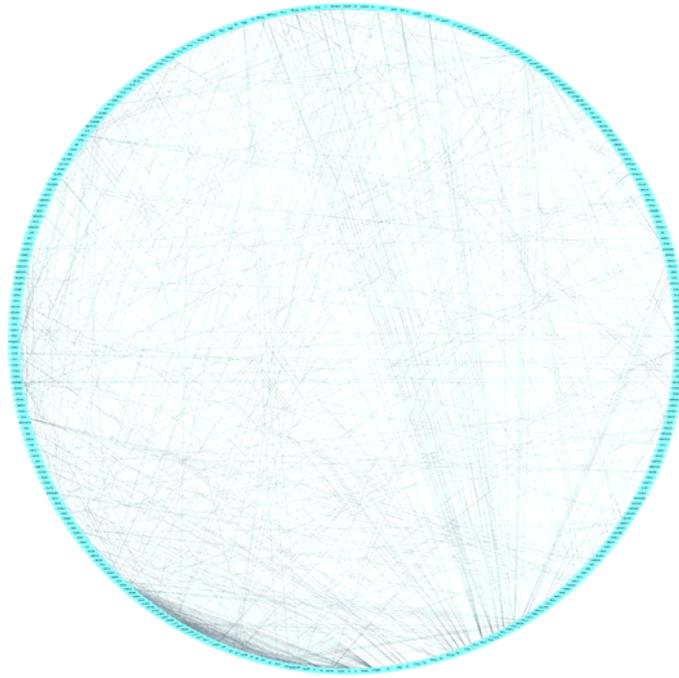
The Venn plots between DEGs and co-expression modules. Geo\_blue indicates the most significant module identified by WGCNA analysis of GSE3189; UCSC Xena\_diff indicates DEGs identified by the UCSC Xena database; UCSC Xena\_magenta indicates DEGs identified by WGCNA analysis of the most significant modules out; GEO\_diff indicates DEGs identified by GSE3189.



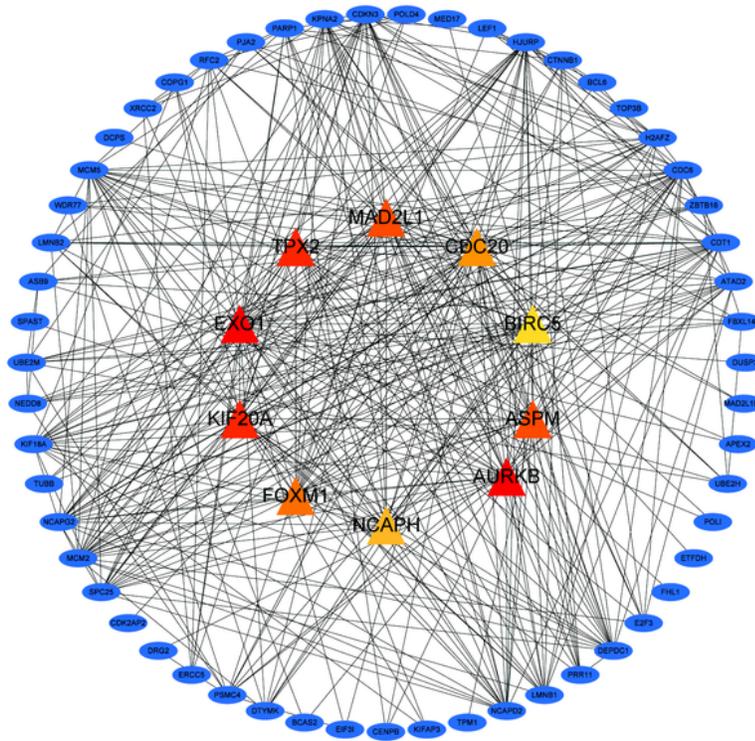
**Figure 6**

GO enrichment analysis and KEGG pathway enrichment analysis of DEGs. (A) shows the GO enrichment analysis of DEGs; (B) shows the DEGs' KEGG pathway enrichment analysis.

A

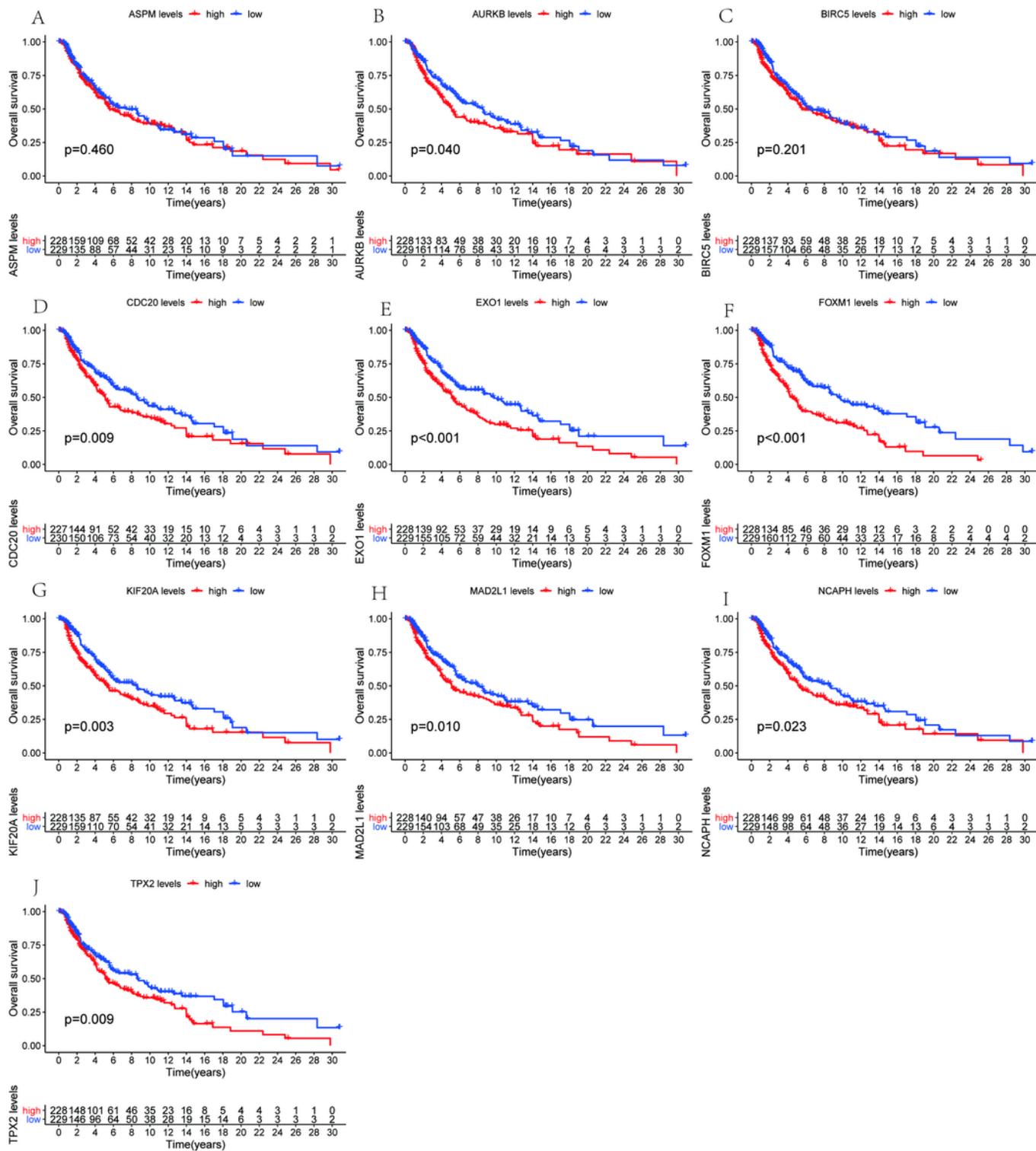


B



**Figure 7**

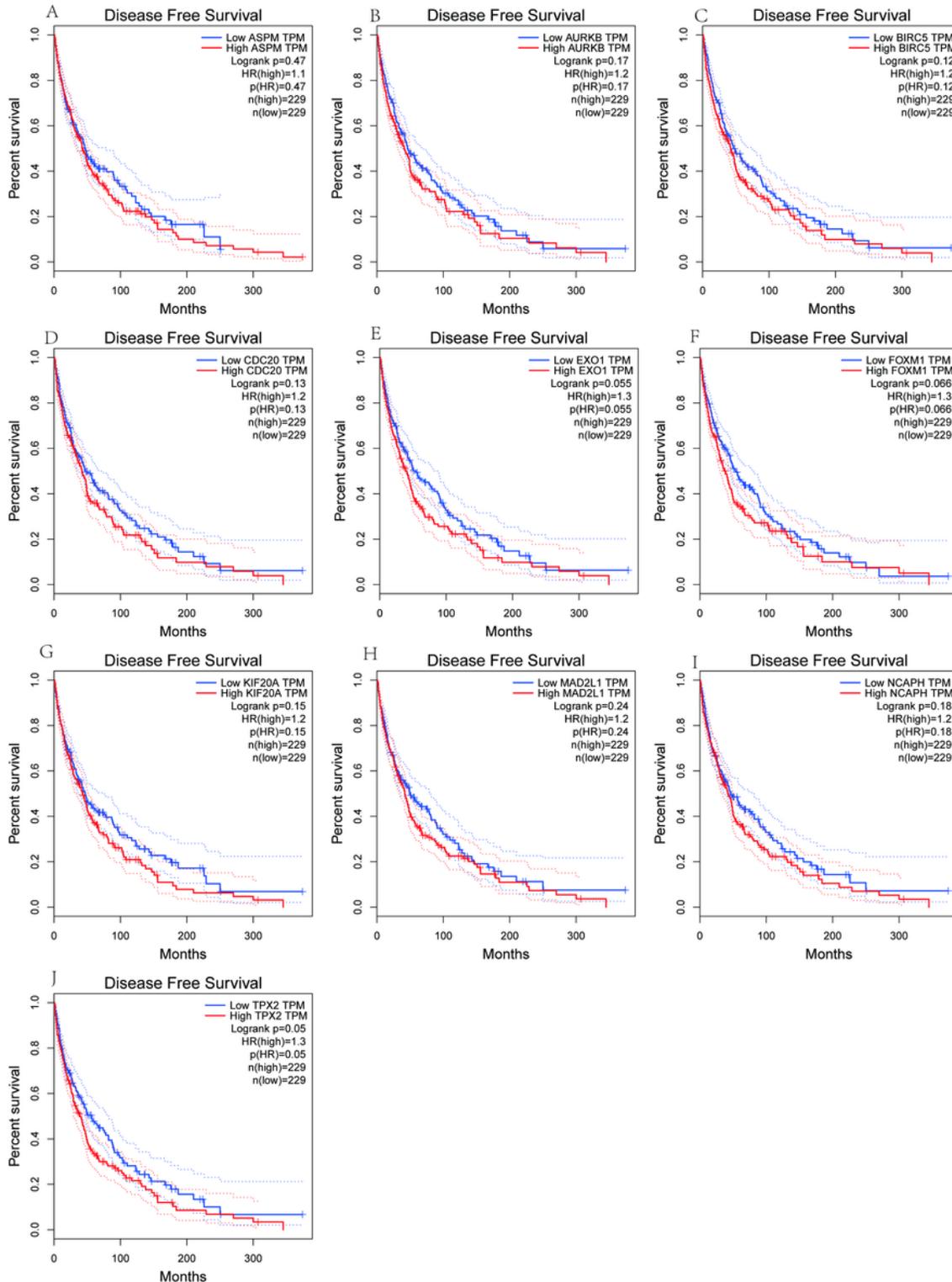
Protein-protein interaction (PPI) network diagram of DEGs and visualization of the Hub gene. (A) shows the PPI network diagram derived from the STRING input library; (B) shows the Hub gene identified by Cytohubba.



**Figure 8**

Overall survival (OS) analysis of the top 10 Hub genes in melanoma patients from the UCSC Xena database. (A) Survival analysis of ASPM in melanoma. (B) Survival analysis of AURKB in melanoma. (C) Survival analysis of BIRC5 in melanoma. (D) Survival analysis of CDC20 in melanoma. (E) Survival analysis of EXO1 in melanoma. (F) Survival analysis of FOXM1 in melanoma. (G) Survival analysis of

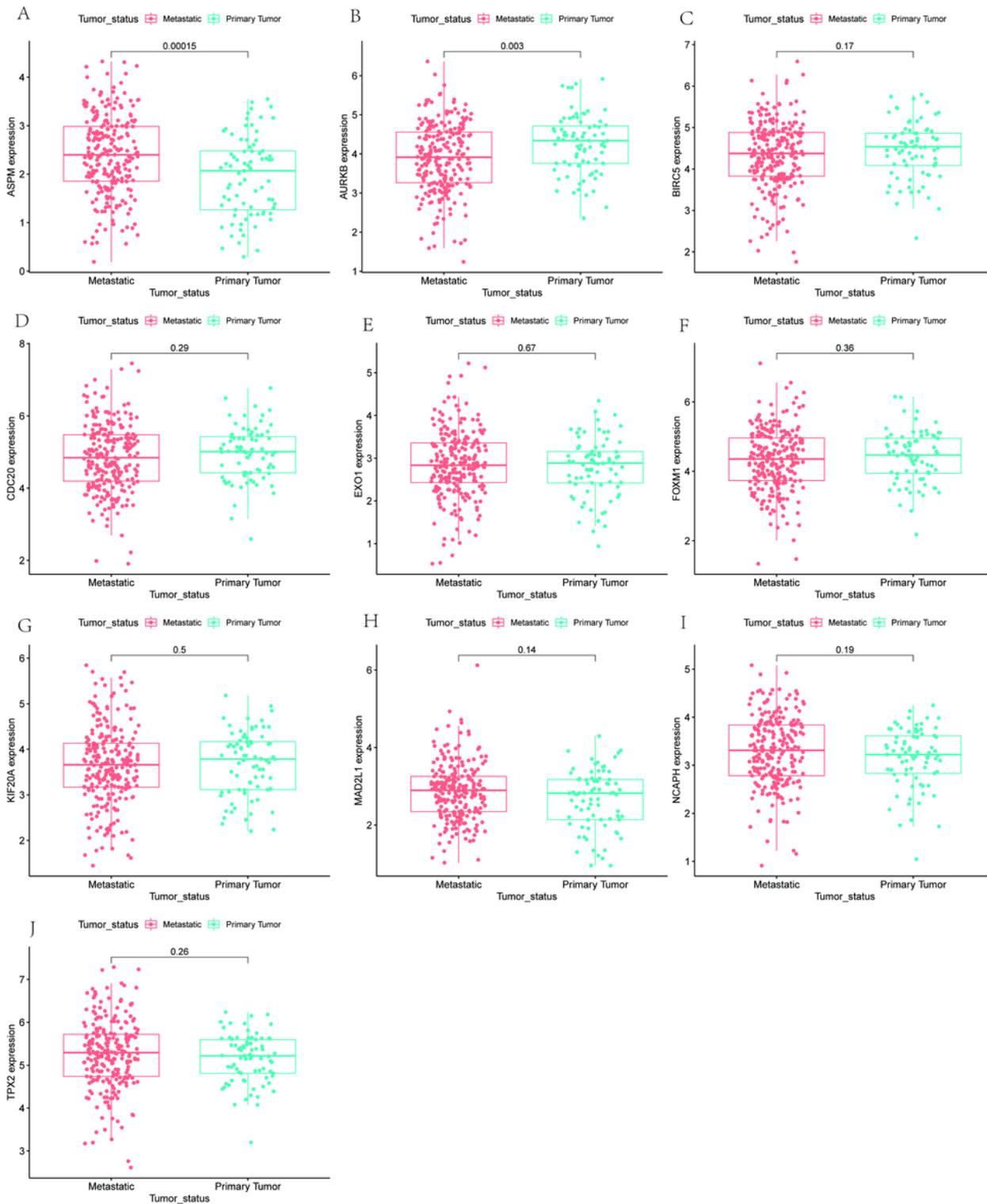
KIF20A in melanoma. (H) Survival analysis of MAD2L1 in melanoma. (I) Survival analysis of NCAPH in melanoma. (J) Survival analysis of TPX2 in melanoma.



**Figure 9**

Disease-free survival (DFS) analysis of the top 10 Hub genes in melanoma patients from the GEPIA2 database. (A) Survival analysis of ASPM in melanoma. (B) Survival analysis of AURKB in melanoma. (C) Survival analysis of BIRC5 in melanoma. (D) Survival analysis of CDC20 in melanoma. (E) Survival

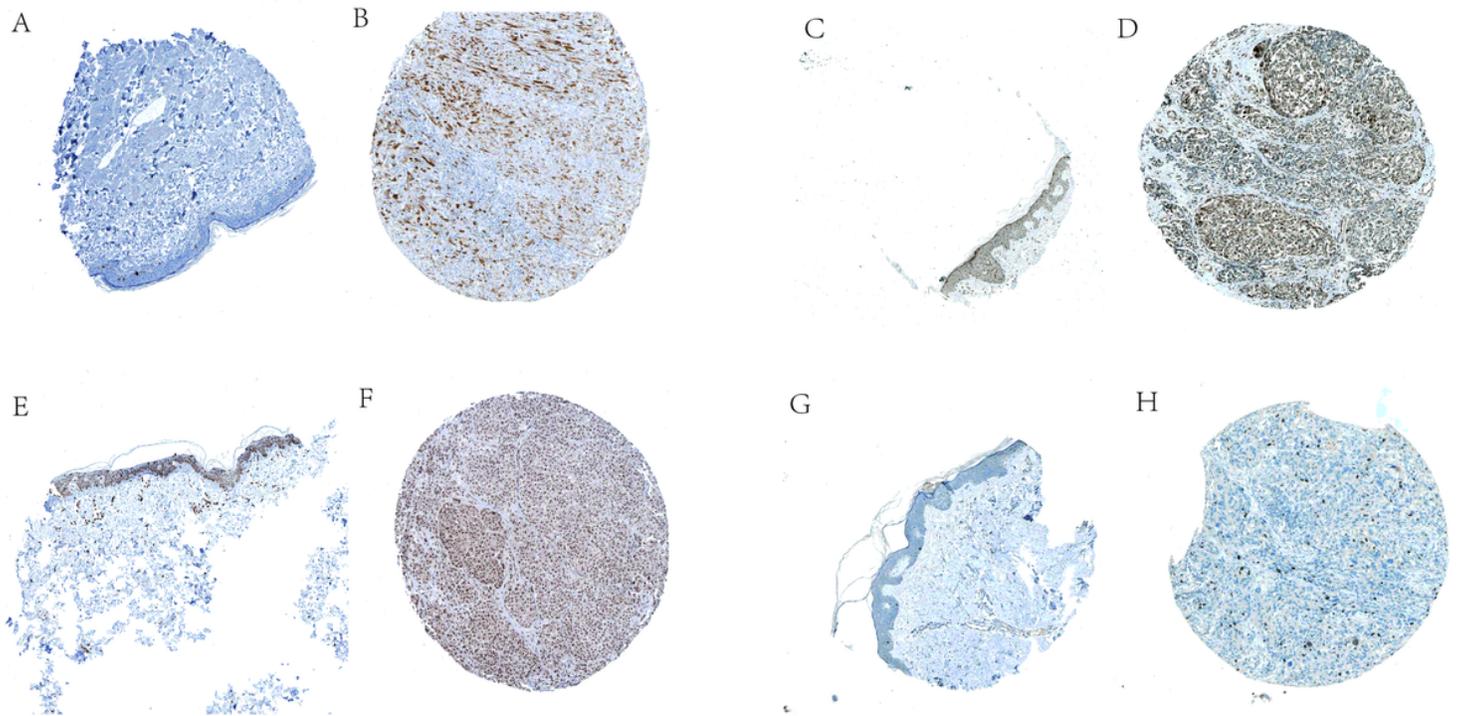
analysis of EXO1 in melanoma. (F) Survival analysis of FOXM1 in melanoma. (G) Survival analysis of KIF20A in melanoma. (H) Survival analysis of MAD2L1 in melanoma. (I) Survival analysis of NCAPH in melanoma. (J) Survival analysis of TPX2 in melanoma.



**Figure 10**

Validation of the expression levels of the top 10 Hub genes in melanoma and normal tissues from the UCSC Xena database in patients developing metastases and patients with primary tumors. (A)

Expression levels of ASPM in the metastatic and primary groups in melanoma. (B) Expression levels of AURKB in the metastatic and primary groups in melanoma. (C) The expression level of BIRC5 in the metastatic and primary melanoma groups. (D) The expression level of CDC20 in the metastatic and primary groups of melanoma. (E) Expression levels of EXO1 in the metastatic and primary groups in melanoma. (F) Expression levels of FOXM1 in the metastatic and primary melanoma groups. (G) Expression levels of KIF20A in the metastatic and primary melanoma groups. (H) The expression level of MAD2L1 in the metastatic and primary groups in melanoma. (I) The expression level of NCAPH in the metastatic and primary melanoma groups. (J) The expression level of TPX2 in the metastatic and primary groups of melanoma.



**Figure 11**

Immunohistochemical staining (IHC) validation of the HPA database for CDC20, FOXM1, KIF20A, and TPX2. (A) IHC staining of CDC20 in normal skin, no. CAB004525. (B) IHC staining of CDC20 in melanoma, no. CAB004525. (C) IHC staining of FOXM1 in normal skin, no. CAB017832. (D) IHC staining of FOXM1 in melanoma, no. CAB017832. (E) IHC staining for KIF20A in normal skin numbered HPA036910. (F) IHC staining for KIF20A in melanoma numbered HPA036910. (G) IHC staining for TPX2 in normal tissue numbered HPA005487. (H) IHC staining for TPX2 in melanoma. No. HPA005487.