Bioinformatics and machine learning-based identification of cell cycle-related genes and molecular subtypes in endometrial cancer

Jingying Pan[a,b,1], Shuhan Huang[c,1], Bidong Fu[c,1], Ruiyu Zhang[b], Minqin Zhou[c], Zichuan Yu[c], Hong Zeng[c], Xitong Geng[c], Yanting Zhu[d], Hao Zheng[c], Hao Wan[b], Xiaoyu Qu[e], Shengwei Tang[c], Yanying Zhong[a*]

[a.]Department of Obstetrics and Gynecology, Second Affiliated Hospital of Nanchang University, Nanchang, China;

[b.]First College of Clinical Medicine, Nanchang university, China;

[c.]Second College of Clinical Medicine, Nanchang university, China;

[d.]HuanKui Academy,Nanchang University, China;

[e.]School of Ophthalmology&optometry,Nanchang University, China.

[1]These authors contributed equally to this work.

***Correspondence:** Yanying Zhong, E-mail:qqqaz012@163.com, Department of Obstetrics and Gynecology, Second Affiliated Hospital of Nanchang University, Nanchang, China. 1 Minde Road, Nanchang, Jiangxi Province, 330000, China;

Abstract

Endometrial cancer is a common malignant tumor in women, with rising incidence rates and an unoptimistic prognosis. DSN1 is a kinetochore protein-coding gene that affects centromere assembly and progression in cell cycles, which is associated with adverse predictions for many cancers. However, the role of DSN1 in UCEC has not yet been reported. We identified the UCEC-related gene module and obtained the differential genes. Then we constructed a diagnostic model and identified the subtype of the molecule and its association with predictions. Subsequently, we identified DSN1 as the core gene and predicted its predictive value. Furthermore, using bioinformatics methods, we found DSN1 was associated with certain clinical characteristics and experimentally validated the expression in cancer tissues of DSN1. Pathway enrichment analysis identified DSN1 as a cell cycle-associated protein, which was validated by WB. The protein interaction network also revealed DSN1 was significantly associated with NDC80. Then we explored the correlation of DSN1 and immune cells and immune cell infiltration and found that DSN1 may affect Th2 enrichment by affecting CCL7 and CCL8. Drug susceptibility analysis showed DSN1 was sensitive to cisplatin and resistant to sunitinib. In conclusion, DSN1 was a novel biomarker that contributes to prognosis and treatment.

Introduction

Endometrial cancer (UCEC) is the fourth most familiar cancer in women（1）. The morbidity of UCEC has increased by 132% in the last three decades, and the mortality

rate has also increased (2, 3). Most patients with early-stage endometrial cancer are relatively manageable by the hysterectomy and adjuvant radiotherapy (4). However, the current prognosis for terminal UCEC patients remains poor, with a five-year survival rate of only 15% for stage IV patients (5). Although chemotherapy is the standard treatment for patients with terminal or relapsing endometrial cancer, the outcome is unsatisfactory (6). So there is still a need to find ways to assist in pre-diagnosis to ameliorate the prognosis of UCEC. Biomarkers play a significant part in the identification, early diagnosis, and disease prevention and monitoring during treatment. Therefore, it is significant to find biomarkers that can enhance the prognosis of endometrial cancer and assist in early diagnosis.

DSN1(DSN1 component of MIS12 kinetochore complex) is a widely distributed protein in the centromere, and its coding gene is located on chromosome 20q11.23. DSN1 plays a vital part in many biological processes, for example, mitosis and cell cycle. Some studies have indicated that chromosomal instability is connected with the high expression of DSN1. And chromosomal instability can lead to chromosomal structural and quantitative abnormalities, which is a prominent feature of human cancers. DSN1 has been reported to have profound significance in the progression of various tumors, including breast carcinoma (7), colorectal carcinoma (8), hepatocellular carcinoma (9), and sarcoma (10). For example, in colorectal and hepatocellular carcinomas, DSN1 affects cell cycle progression and is closely related to its clinicopathological features, and in vitro, DSN1 significantly promotes

osteosarcoma cell proliferation (11). However, the function of DSN1 in UCEC has not been reported.

With the bioinformatics and experimental method, we investigated the potential function and mechanism of DSN1 in UCEC. By the LASSO regression and consensus cluster analysis, we confirmed the core genes and determined the molecular subtypes of UCEC and their relationship with prognosis. Then, survival and random forest analysis were utilized to identify the key gene, DSN1. The expression of DSN1 in UCEC and its correlation with prognosis were analyzed by R software, qRT-PCR, and Western Blot experiments. Subsequently, Linkedomics and GSEA were utilized to investigate the pathway enrichment and latent biological functions of DSN1 and verified by WB. Using STRING, we constructed a DSN1 protein interaction network. In addition, the relationship between DSN1 and tumor-infiltrating immune cells, chemokines, and drug interactions in UCEC were also investigated. Our study suggested that DSN1 is a novel biomarker of UCEC that contributes to prognosis and treatment.

Materials and methods

Clinical samples

The specimens of all 50 patients from the Second Affiliated Hospital of Nanchang University were pathologically identified with UCEC between December 2023 and September 2024. While the tissues were still fresh, total proteins were removed. The study was approved by the ethics committee of Nanchang University's Second

Affiliated Hospital (No. Review [2023] No. (165)). This is a retrospective study. As it analyzed existing medical records and archived samples, obtaining informed consent was infeasible. The ethics committee of Nanchang University's Second Affiliated Hospital has approved the waiver of informed consent for this clinical research. And after the data is collected, information can be obtained that can identify individual participants.

Data Collection and Proceed

Sequencing and clinicopathological data of UCEC patients were derived from the TCGA database (https://cancergenome.nih.gov), including 59 normal samples and 554 tumor samples. The mRNA expression profiles (Number: GSE17025, GSE39099, and GSE106191) were derived from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Determination of differentially expressed genes (DEGs)

The gene sets were normalized, and difference analysis was carried out using the "limma" package to get DEGs between each comparison group and the control group. We used a P-value less than 0.05 and |log2 fold change (FC) |> 1 to determine DEGs.

Cell culture

The commercially accessible UCEC cell line HEC-1-B was utilized to matrix cells, while the detailed method can be referred to in this article (12).

Transfection with shRNA

Specific sequences were designed to manipulate the expression of DSN1. ShRNAs were cloned into the lentiviral pLKO.1 vector, which was purchased commercially

from GenePharma Biothch in Shanghai, China.

RNA extraction and qRT-PCR

The TRIzol reagent (Invitrogen, USA) per the manufacturer's instructions was applied to isolate total RNA from the cells. Invitrogen's M-MLV reverse transcription kit was used to do reverse transcription on total RNA. The Agilent Technologies AriaMx Real-Time PCR System (Agilent, USA) and SYBR Green PCR Master Mix (Roche, Switzerland) were utilized for the qRT-PCR analysis. The relative expression was computed using the 2 (delta delta threshold cycle) (2Ct) technique with tubulin as a normalized internal control.

Western blotting analysis and antibodies

Western blotting analysis was carried out. Primary antibodies included anti-DSN1 (1:1000, Cell Signaling Technology 4724), anti-NDC80 (1:1000, Santa Cruz, sc-515550), anti-Cyclin D1 monoclonal antibody (1:1000, Abcam, ab16663), anti-Tubulin monoclonal antibody (1:1000, Proteintech, 11224-1-AP).

WGCNA (Weighted correlation network analysis)

Using WGCNA, we uncovered interesting gene modules that are closely connected to other genes. First, transpose the concatenated matrix file, then screen out the genes with expression variance in the first quartile before constructing the correlation matrix. The weighted adjacency matrix is then transformed into a topological overlap matrix (TOM) to assess network connectivity, and the hierarchical clustering approach is utilized to create the TOM clustering tree. Genes were sorted into modules based on their expression patterns using a weighted correlation coefficient.

The relationship between modules and clinical characteristics was further investigated to identify modules linked with UCEC for investigation.

Gene function enrichment analysis

GSEA was implemented by applying R package ClusterProfiler 4.1.0, and genomic enrichment analysis of differentially expressed genes of TCGA and GEO was performed. Prominent functional and pathway differences between high and low DSN1 expression groups were also analyzed by GSEA. Based on gene expression profiles and phenotypic grouping, we kept the minimum gene set at 5 and the largest gene set at 5,000. p values less than 0.05 and FDR less than 0.1 were deemed statistically significant. Then, R packages "clusterProfiler" and "org.hh.egg.db" were applied to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis to further analyze the associated pathways of 164 genes and 20 genes in UCEC. An adjusted p-value of less than 0.05 was considered significantly enriched.

Machine learning identification of candidate genes associated with UCEC

Using the least absolute shrinkage and selection operator (LASSO), the final core genes were screened. A diagnostic model was applied to determine a risk score for each patient. The random forest algorithm is an algorithm based on bagging. Bagging is an integration method that randomly divides a data set into multiple subsets to build a basic decision tree. We used a random forest algorithm to score these genes.

Identification of molecular-related subtypes by consensus clustering

The samples are categorized by determining the consensus matrix through

consensus clustering, and the K-value for the number of clusters is set between 2 and 9. The optimal K value was determined when the cumulative distribution function index reached an approximate maximum. Correlated subtypes were established based on consensus clustering of 20 gene expression levels. These results were used to define Cluster 1, Cluster 2, and Cluster 3. Heat maps were plotted using the ggplot2 software package to visualize the expression of these 20 genes in the different clusters.

LinkedOmics Database Analysis

The LinkedOmics database (http://www.linkedomics.org/login.php) was used to evaluate 32 TCGA cancer-related datasets. Its "LinkFinder" module has been used to investigate differential expression genes associated with DSN1 in the TCGA UCEC. If the P-value and false discovery rate are both less than 0.05, the gene will be considered a significant related gene.

Protein-protein interaction network construction

The STRING database (https://string-db.org) allows us to explore the genomic relationships between genes encoding proteins and thus infer the biological activity of proteins. Our study utilized the STRING database to explore the top 500 co-expressed DSN1 genes and selected medium-confidence genes with a confidence level of 0.9, which were then displayed by Cytoscape software to generate an overall PPI network.

TIMER Database Analysis

The Tumor Immune Estimation Resource (TIMER) database could provide immune infiltration information for different cancer types systematically. We utilized

"Gene", "Diff Exp", and "Correlation" modules to acquire immune-related information and gene expression level information.

TISIDB analysis

The TISIDB database (http://cis.hku.hk/TISIDB/) could be used to investigate the association between individual genes and immunity in the setting of cancer. Using the "Chemokine" module, we investigated the link between DSN1 expression and cytokines in UCEC.

Comparative toxicogenomics database

The CTD (http://ctdbase.org/) integrates the interactions of specific genes with relevant compounds and contains literature coverage of a number of compounds. From this, we got several drugs that regulate DSN1 expression.

Statistical Analysis

In this study, all statistical studies were accomplished by R software (version 3.6.3/4.0.3/4.1.2).

Survival and clinicopathological features analysis

The R packages involved are "pROC", "timeROC" and "ggplot2". By using logistic regression, we investigated the association between DSN1 and clinicopathological features. In addition, DSN1 and the prognostic model use the "Survival" software package for survival regression fitting. And survival analysis for different subgroups. Then, to compare various survival factors, our team developed time-dependent receiver operating characteristic curves (ROC) and used the “rms” package to build a nomogram model and visualize it.

198 **Correlation analysis**

199 Using the R tools (limma, ggplot2, and heatmap packages), we conducted gene
200 expression correlation analysis. The correlation of DSN1 with DSN1-related genes.
201 The correlation of CCL7 and CCL8 with Th2 cells.

202 **Calculation of immune infiltration**

203 We used R-packet GSVA and 24 immune cell markers to calculate the immune
204 infiltration corresponding to DSN1.

205

206 **Result**

207 **Modular cluster analysis of endometrial cancer-related genes using weighted**
208 **gene co-expression network analysis (WGCNA).**

209 Aimed to explore critical modules and genes linked with endometrial cancer
210 (UCEC), we used the mRNA expression matrix of TCGA-UCEC, GSE17025,
211 GSE39099, and GSE106191. First, we identified 3922 and 9037 differential genes for
212 GEO (P-value <0.05) and TCGA (P-value<0.05, FDR<0.05, and |logFC|>1)(Figure
213 1A-B), which were then used for functional enrichment analysis and WGCNA. The
214 GSEA analysis revealed significant links between UCEC-related differential genes
215 and cell cycle regulation and immunological pathways (Figure 1C, D). We then
216 separately constructed the WGCNA model and identified the gene co-expression
217 modules (Figure 1E, F). We found that the yellow modules containing 371 genes were
218 most relevant to the UCEC for the TCGA-UCEC database, while it contained 429
219 genes for the GEO. Furthermore, we conducted a cross-analysis of the genes and

obtained a set of 164 genes (Figure 1G), which was analyzed by GO and KEGG, showing that these genes were closely related to DNA replication, cell cycle, and p53 signaling pathways (Figure 1H, I). Furthermore, random forest algorithms showed higher scores for the 15 genes with higher predictive values in the gene concentration (Figure 1J). Overall, through a comprehensive analysis of multiple datasets, we identified modules and genes that were closely related to UCEC, and they were dramatically gathered in cell cycle-related pathways.

Figure 1. Identification of key gene modules in WGCNA. (A-B) The volcano map of the DEGs screening from the TCGA-UCEC and GSE17025, GSE39099, and GSE106191 data sets. (C-D) GSEA analysis of DEGS. (E) WGCNA analysis identified the module related to UCEC in the TCGA database as the yellow module (r = 0.4). (F) WGCNA analysis confirmed the module most associated with UCEC in the GEO dataset as the yellow module (r = 0.42). (G) The Venn figure depicted the TCGA-UCEC, GEO-UCEC intersection of related genes. (H, I) The GO and KEGG pathway analysis of the gene set containing 164 genes. (J) The random forest analysis.

LASSO regression analysis further built a diagnostic model containing 20 genes.

To build the predictive model and improve the accuracy of the prediction, we employed an LASSO regression analysis of the 164 genes. Then, we refined 20 core genes (Figure 2A, B). Simultaneously, they had higher levels of expression in cancer tissue than normal tissue (Figure 2C). Based on these key genes, we constructed a

diagnostic model whose ROC curve analysis AUC values reached 1, showing high diagnostic accuracy (Figure 2D). GO and KEGG pathway analyses were used to understand the biological functions of genes, showing that the genes were highly concentrated in the pathways related to cell cycle regulation (Figure 2E).

Figure 2. A LASSO regression diagnosis model was constructed. (A-B) LASSO regression analysis was used to determine the most critical model genes, and a new gene set containing 20 genes was obtained. (C) Expression profiles of 20 genes. (D) Diagnostic ROC curves of key gene sets in UCEC. (E) GO and KEGG pathway enrichment analysis of gene sets.

Identification of molecular subtypes of endometrial cancer-related genes.

To study the biological differences between the different subtypes of UCEC, the ConsensusClusterPlus package was used to conduct a consistent clustering analysis of the expression profiles of 20 model genes. CDF curve results showed that when k = 3, the classification was dependable and steady (Figure 3A-C). Meanwhile, the specific expression of the 20 genes in the three clusters was presented in the hot chart (Figure 3D). Immediately, we compared the survival conditions of these three clusters (C1, C2, and C3), suggesting that the C3 cluster had the worst prognostic effect and that there were no significant differences between C2 and C1 (Figure 3E). Overall, there may be some differences in prognosis between the different subtypes of these UCEC-related genes.

Figure 3. Identification of the molecular subtypes of genes associated with endometrial cancer. (A-C) Consensus clustering determined three UCEC clusters with different enrichment scores. (D) A heat map depicting the expressions of 20 genes in the three clusters. (E) The UCEC OS curve of patients between different subtypes.

DSN1 had the highest research value based on 20 gene survival analyses.

To explore the predictive and diagnostic value of the genes, we used R software to map the survival curves. It indicated that the eight genes, such as DSN1, had a significant and strong correlation between their high levels of expression and the patient's poor overall survival (OS), suggesting that they may be potential biological markers for disease diagnosis and prognosis assessment (Figure 4A). In contrast, the results of the remaining 12 genes showed no statistically significant differences (as shown in Supplementary Figure 1A). Furthermore, according to the random forest chart score (Figure 1J), the DSN1 gene had the highest score, and its specific role in UCEC had not yet been explored, which greatly aroused our interest. So we further explored its potential biological function and its role in the cancer process. The diagnostic ROC curve of DSN1 showed a high diagnostic value of AUC of 0.966 (Figure 4B). We further constructed a nomogram containing the parameters of the age and different expressions for DSN1. And the calibration chart validates the calculated scores that can accurately predict the 1-, 3-, and 5-year survival chances (Figure 4C, D).

Figure 4. Survival analysis of model genes and identification of core genes. (A) OS curves of model genes with statistically significant survival. (B) Diagnostic ROC curve of DSN1 expression in endometrial cancer. (C) Nomogram to predict 1-, 3-, or 5-year OS rates in patients with endometrial cancer. (D) Calibration curve of the nomogram.

DSN1 was associated with multiple clinical pathology characteristics in UCEC.

To explore the link between DSN1 and the clinical characteristics in UCEC, we conducted gene expression and clinical pathological characteristics analysis. It showed the expression of DSN1 in tumor tissue was significantly higher ($P < 0.001$) (Figure 5A, B). Next, we detected the links between gene expression and clinical pathological characteristics. In terms of OS, significantly increased DSN1 was observed in UCEC-dead patients ($P < 0.05$) (Figure 5C). Meanwhile, vitally increased DSN1 was observed in the group over 60 (Figure 5D). As the tumor developed, the expression of DSN1 also increased significantly with the increase in histologic grade (Figure 5E). In the clinical stage, DSN1 expression in Stage I was prominently lower than that in Stage III (Figure 5F). Subsequently, we mapped the total survival curve under different clinical pathological characteristics, assessing their prognosis and diagnostic value. It demonstrated that the overall survival rate was lower with high expression of DSN1 for a worse prognosis, no matter if the group was for different ages (Figure 5G, H). The OS for patients with high expression of DSN1 was low in prognosis at G1 and G2, while G3 was without significant differences (Figure 5I, J).

The OS for patients with high expression of DSN1 was low in prognosis at Stage I and Stage II, while at Stage III and Stage IV, there were no significant differences (Figure 5K, L). The qRT-PCR was performed on the tumor tissues and corresponding non-tumor tissues of 50 UCEC patients, and it showed that the mRNA expression levels of 33 of these samples were significantly upregulated (Figure. 5M). According to the results of WB, the experiment DSN1 protein levels were significantly increased in cancer tissue (Figure 5N, O). In conclusion, the expression of DSN1 was significantly increased in cancer tissues, and its differential expression may have different effects under different clinicopathological characteristics.

Figure 5. Correlation between DSN1 expression and clinicopathological features. (A) DSN1 mRNA levels in tumor and normal tissues based on the TCGA database. (B) Based on the TCGA database, DSN1 was differentially expressed in UCEC-paired samples. The expression level of DSN1 under the different clinical pathological features, including (C) OS event, (D) age, (E) histologic grade, and (F) clinical stage (*p < 0.05, **p < 0.01, ***p < 0.001). The overall survival curves of patients with differential DSN1 expression under different clinicopathological characteristics include (G) age≤60, (H) age＞60, (I) histological grade: G1 and G2, (J) histological grade: G3, (K) clinical stage: stage Ⅰ and stage Ⅱ, and (L) stage Ⅲ and stage Ⅳ. (M) Quantitative Real-time PCR analysis of DSN1 mRNA level in 50 cases of UCEC tissues and corresponding normal tissues. Left, a log 2 (T/N) value >0 indicated that DSN1 expression was overexpressed in the UCEC samples; right, a log 2 (T/N) value <0 indicated that DSN1 expression was downregulated in the UCEC

329 samples. (N,O) Determination and quantification of DSN1 protein levels in UCEC
330 tissues and paired non-tumor tissues by western blot.
331
332 **Exploring the potential biological functions of DSN1 in UCEC**
333 To study the biological significance of DSNI in UCEC, we found DSN1
334 co-expression genes (Figure 6A). The hot charts, respectively, showed the top 50
335 genes that were positively and negatively related to DSN1 (Figure 6B, C). Then, we
336 explored the potential pathways for DSN1 in UCEC. GO analysis suggested that
337 when DSN1 expression was elevated, pathways such as DNA replication were
338 activated (Figure 6D). KEGG analysis showed that when DSN1 expression was
339 elevated, pathways such as the cell cycle, DNA replication, and pyrimidine
340 metabolism were activated (Figure 6E). Meanwhile, GSEA analysis indicated that the
341 elevation of DSN1 expression was closely related to pathways such as the cell cycle,
342 DNA replication, and purine metabolism (Figure 7A-I), showing that DSN1 can affect
343 cell cycle-related pathways.
344 **Figure 6. Using Linkedomics to investigate potential biological processes in**
345 **which DSN1 was involved.** (A) Volcano plot of DSN1 co-expressed genes. (B, C)
346 UCEC DSN1 top 50 positive correlation and negative correlation gene heat maps. (D,
347 E) GO and KEGG pathway enrichment analysis of DSN1 in endometrial carcinoma.
348 **Figure 7. GSEA pathway enrichment analysis.** Using GSEA, we found that DSN1
349 participated in potential biological processes, including (A) cell cycle, (B) DNA
350 replication, (C) basal transcription factors, (D) purine metabolism, (E) P53 signaling

pathway, (F) regulation of autophagy, (G) nucleotide excision repair, (H) mismatch repair, and (I) RNA polymerase.

Construction of PPI networks associated with DSN1

Protein interactions play an irreplaceable role in life processes such as cell cycle regulation. To study the specific mechanisms of DSN1 in cell cycles, we built a network of DNS1 protein interactions (Figure 8A). The first 50 genes of the interaction were shown in Figure 8B. We then screened the top 10 genes of the PPI score for a correlation and survival analysis, showing that all 10 were significantly positive, but only 6 of them had statistically significant survival (Figure 8C, D (Supplementary Figure 2A, B)). Among them, NDC80 has the highest correlation with DSN1, and NDC80 is highly expressed in a variety of cancers, such as liver cancer（13）and epithelial ovarian cancer（14）. In addition, NDC80 can also promote the progression of glioma by affecting the cell cycle（15）. Therefore, we speculate that DSN1 can influence the UCEC progression by influencing NDC80. WB revealed that NDC80 and Cyclin D1 protein levels dropped when DSN1 was knocked down (Figure 8E). As a result, we postulated that DSN1 may contribute to cancer progression by regulating NDC80 to stimulate the cell cycle.

Figure 8. The establishment of the DSN1-related protein-protein interaction network and the analysis of the hub genes. (A, B) DSN1-associated protein-protein interaction (PPI) network. (C) Correlation between DSN1 and the expression of NDC80, SPC25, BUB1, CENPA, BUB1B, and CENPI in UCEC. (D) Prognostic

373 analysis of related genes. (E) Western blot was used to detect DSN1, NDC80, and
374 Cyclin D1 protein expression in UCEC cells stably transfected with the control
375 shRNA or the DSN1 shRNA.

376

377 **Relevance of DSN1 expression to immune cell infiltration**

378 To reveal the relationship between DSN1 expression and the UCEC immune
379 response, the immune microenvironment was analyzed between the DSN1 differential
380 expression groups. Multiple immune cells in the DSN1 low-expression group were
381 significantly enriched. For the DSN1 high-expression group, the enrichment of T
382 helper cells, Tgd cells, and Th2 cells was higher (Figure 9A). Simultaneously, it
383 demonstrated that DSN1 was negatively related to the vast majority of immune cells,
384 while it was remarkably positive for Th2 cells, T helper cells, TGD, and Tcm (Figure
385 9B). Next, we delved into the relationship between DSN1 expression and immune
386 infiltration. It revealed that DSN1 expression was negatively associated with the
387 immune cell infiltration of B cells, CD4+ T cells, and dendritic cells. However, there
388 was a positive correlation with neutrophils and no statistical significance for CD8+T
389 cells and macrophages (Figure 9C). Then, it was found that DSN1 expression was
390 negatively correlated with most chemokines and significantly positively correlated
391 with CCL7 and CCL8 expression in UCEC (Figure 9D, E, F). Furthermore, the Th2
392 cells and chemokines CCL7 and CCL8 were significantly positively related (Figure
393 9G, H). Overall, DSN1 may affect the enrichment of Th2 cells in UCEC.

Figure 9. Analysis of the correlation between the DSN1 expression and the immune level and the chemokines. (A) Box plots showed the differences in immune microenvironment between DSN1 high and low expression groups. (B) The lollipops of the correlation of DSN1 expression and immune cells. (C) Correlation of DSN1 expression in UCEC with infiltration levels of B cell, CD8+ T cell, CD4+ T cell, macrophage, neutrophil, and dendritic cell. (D) Heat map of the correlation between DSN1 expression and chemokines in pan-cancer. (E, F) Scatter plot of the correlation between CCL7, CCL8, and DSN1 expression. (G, H) Scatter plots of the correlation between CCL7 and CCL8 expression and Th2 cell enrichment.

DSN1 drug interaction and drug sensitivity analysis

To explore interactions of DSN1 and cancer treatment drugs, we established drug interaction networks. It showed that seven kinds of drugs could promote the expression of DSN1. However, 11 drugs could inhibit the expression of DSN1 (Figure 10A). We compared cgp2014 and CTD, and cisplatin and sunitinib were found in both them (Figure 10B). IC50 results showed that, when DSN1 had high expression, the cisplatin IC50 was lower, and the sunitinib IC50 was higher than that when DSN1 had low expression (Figure 10C-D). Moreover, it indicated that when DSN1 was highly expressed, the patients were sensitive to cisplatin and were resistant to sunitinib.

Figure 10. Gene-drug interactions and drug-susceptibility analyses. (A) CTD showed the interaction network between DSN1 and the drug. (B) The Venn figure

showed the drug sensitivity results from the two databases. (C-D) The box plots showed the cisplatin IC50 and sunitinib IC50 of DSN1 differential expression.

Discussion

Endometrial cancer is one of the most prevalent gynecologic malignancies in developed nations. The incidence rate of UCEC in China was 10.54/100,000, and the fatality rate was 2.53/100,000, based on data from the National Cancer Center for the year 2022 (16). In addition, due to the increasing prevalence of obesity and metabolic syndrome, the incidence and mortality of UCEC are still increasing significantly (17). While most patients with early-stage UCEC can be cured with hysterectomy and adjuvant radiotherapy, the prognosis remains poor for patients with late-stage UCEC, with a five-year survival rate of roughly 48% for stage III and 15% for stage IV (18). However, there are currently no biomarkers for the routine clinical diagnosis and prognosis of UCEC (19), so finding new, reliable biomarkers for UCEC and learning more about its molecular biology are urgently needed. In this study, we created a diagnostic model for UCEC by combining multiple databases for comprehensive bioinformatics analysis and experimental verification. We also revealed the expression level and prognostic significance of the key gene DSN1 in the model in UCEC and further explored the biological mechanism of DSN1 promoting the occurrence and progression of UCEC.

In this study, we combined three endometrial cancer mRNA microarray datasets from the GEO database (GSE17025, GSE39099, and GSE106191) and the

endometrial cancer datasets from the TCGA database for differential gene analysis. We then performed WGCNA analysis of these differential genes and obtained 164 core genes that were significantly associated with UCEC in both databases. Then, through pathway enrichment analysis, we found that these 164 core genes were substantially enriched in cell cycle-related pathways, such as mitotic cell cycle transition, regulation of cell cycle transition, DNA replication, and other signaling pathways. At the same time, through random forest analysis, the top 15 genes with the highest score among 164 core genes were found. We then performed lasso regression analysis on these 164 genes and obtained 20 core genes for UCEC diagnosis. In addition, we also used consensus cluster analysis to divide UCEC into three major subtypes based on these 20 genes and constructed associated prognostic models.

Then, through the survival analysis of these 20 genes and taking into account the random forest algorithm mentioned above, DSN1 can influence patient prognosis in addition to having a major impact on the random forest algorithm. At the same time, DSN1, as a kinetochore protein-coding gene, is widely distributed in centromere (20) and is crucial for preserving the stability of centromere structure and proper chromosome separation (10). Previous studies have shown that DSN1 can alter the cell cycle, which can accelerate the growth of certain malignancies. For example, in colorectal cancer, DSN1 can be stabilized in an M6A-related manner that promotes cancer progression (21); in breast cancer, DSN1 expression is significantly upregulated and is strongly linked to a poor prognosis and decreased survival (7). However, the expression of DSN1 in endometrial cancer and its biological function

460 remain unclear. Our investigation revealed a correlation between the prognosis and
461 expression level of DSN1 and several clinicopathological characteristics of UCEC,
462 such as age, histological grade, and clinical stage. Further, we also confirmed that the
463 expression of DSN1 in endometrial carcinoma tissues was greater than that in normal
464 control tissues by qRT-PCR and western blot, which was consistent with
465 bioinformatics analysis. In summary, DSN1's mRNA and protein levels in UCEC
466 have changed, but more research is still needed to determine the precise mechanism of
467 action.

468 In order to further reveal the specific biological mechanism of DSN1 affecting
469 UCEC, we conducted the enrichment analysis of the GOKEGG pathway and the
470 GSEA pathway. The outcomes demonstrated that DSN1 was considerably activated in
471 cell cycle, DNA replication, chromosome separation, and other pathways. It is well
472 known that cell cycle is one of the main cellular mechanisms influencing the
473 development of malignant tumors (22) and having a significant impact on metastasis,
474 immunity, and tumor metabolism (23). The high expression of DSN1 in UCEC may
475 promote the growth of tumor cells by influencing the cell cycle. We further verified
476 this hypothesis through experiments. The experimental results demonstrated that
477 Cyclin D1 protein level decreased following DSN1 knockdown. These results suggest
478 that DSN1 knockdown can inhibit the cell cycle, thereby inhibiting tumor growth. In
479 summary, our study revealed that DSN1 is likely to promote cell proliferation and
480 cancer progression in UCEC by promoting the UCEC cell cycle.

481 Subsequently, through PPI analysis, we found that DSN1 had protein interaction

with NDC80, SPC25, BUB1, CENPA, BUB1B, and CENPI. Among them, NDC80 is the gene with the highest correlation with DSN1, and as the outer component of the centromere and the regulator of spindle check points, NDC80 is highly expressed in a variety of cancers. NDC80 knockdown in pancreatic cancer can cause cell cycle halt, which slows tumor development (24). NDC80 knockdown in gastric cancer can significantly reduce cell growth in vitro and in vivo, preventing cell division during the G2/M phase (25). Furthermore, research has shown that NDC80 is significantly overexpressed in endometrial cancer and is associated with a poor prognosis (26). Thus, we hypothesize that DSN1's interaction with NDC80 may have an impact on the onset and progression of UCEC. Through experiments, we found that in the shDSN1 group, NDC80 and Cyclin D1 protein levels were significantly reduced. Through these experiments, we have reason to believe that DSN1 affects the cell cycle through NDC80 and thus mediates the incidence and growth of tumors.

Given the growing body of research demonstrating the pivotal role the tumor microenvironment plays in tumor initiation, development, metastasis, and response to therapy (27), we looked at the relationship between DSN1 expression and immune cell infiltration in UCEC. The results demonstrated a negative correlation between DSN1 expression and the majority of immune cells' infiltration abundance and a positive correlation with Th2 cells' infiltration abundance. Th2 cells have the function of promoting tumor progression and immunosuppression, which is often linked to a poor prognosis of various tumors (28), and Th1/Th2 balance restoration is crucial in the therapy of cancers (29). Then, we examined the chemokines linked to DSN1

504 expression to look into the connection between DSN1 and immune cell infiltration in
505 more detail. The findings demonstrated that DSN1 was negatively correlated with
506 most chemokines in UCEC but positively correlated with CCL7 and CCL8. And
507 CCL7 and CCL8 were also positively correlated with Th2 cells in UCEC. According
508 to previous studies, CCL7 can enrich Th2 cells during scar formation, thus affecting
509 the local expression levels of Th1 and Th2 cells (30). In vitiligo, CCL8 can deflect T
510 cells toward pathogenic Th2 cells (31). Therefore, we speculate that DSN1 is likely to
511 affect the infiltration level of Th2 cells by regulating CCL7 and CCL8, thus
512 promoting tumor progression and immunosuppression. In summary, taking advantage
513 of the properties of DSN1 that may increase the proportion of Th2 cells, it is possible
514 to try to combine DSN1 inhibitors with immune checkpoint inhibitors to further
515 enhance the effect of immunotherapy in UCEC patients.

516 In conclusion, our study integrated various bioinformatics approaches across
517 multiple datasets to ultimately establish a diagnostic and prognostic model for UCEC,
518 in which DSN1 is expected to be a biomarker for early diagnosis and treatment of
519 endometrial cancer. Bioinformatics analysis and experiments confirmed that DSN1
520 was overexpressed in UCEC and correlated with some clinicopathological features.
521 At the same time, we also tested how DSN1 affects the cell cycle through
522 experiments. Finally, we proposed and predicted the interaction of DSN1 with
523 NDC80, as well as analyzed drugs that were sensitive to cells with high DSN1
524 expression.

525 However, we need to acknowledge several limitations related to this study. First,

we obtained data on UCEC mainly from online public databases, and larger studies are still needed in the future to validate and strengthen our findings. Secondly, we only verified the interaction of DSN1 with NDC80 and did not further verify their specific mode of action through experiments.

Acknowledgements

We thank the mentioned public databases for providing us the data and analytical tools.

Reference

1. Zheng W. Molecular Classification of Endometrial Cancer and the 2023 FIGO Staging: Exploring the Challenges and Opportunities for Pathologists. CANCERS. [Journal Article]. 2023 2023/8/15;15(16).

2. Ogunmuyiwa J, Williams V. Emerging Advances in Endometrial Cancer: Integration of Molecular Classification into Staging for Enhanced Prognostic Accuracy and Implications for Racial Disparities. CANCERS. [Journal Article; Review]. 2024 2024/3/16;16(6).

3. Dobrzycka B, Terlikowska KM, Kowalczuk O, Niklinski J, Kinalski M, Terlikowski SJ. Prognosis of Stage I Endometrial Cancer According to the FIGO 2023 Classification Taking into Account Molecular Changes. CANCERS. [Journal Article]. 2024 2024/1/17;16(2).

4. Coll-de LRE, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S,

548 Colas E. Prognostic Biomarkers in Endometrial Cancer: A Systematic Review and
549 Meta-Analysis. J CLIN MED. [Journal Article; Review]. 2020 2020/6/17;9(6).

550 5. Chen J, Wu S, Wang J, Han C, Zhao L, He K, et al. MCM10: An effective
551 treatment target and a prognostic biomarker in patients with uterine corpus
552 endometrial carcinoma. J CELL MOL MED. [Journal Article; Research Support,
553 Non-U.S. Gov't]. 2023 2023/6/1;27(12):1708-24.

554 6. Ayodele A, Obeng-Gyasi E. Exploring the Potential Link between PFAS
555 Exposure and Endometrial Cancer: A Review of Environmental and
556 Sociodemographic Factors. CANCERS. [Journal Article; Review]. 2024
557 2024/2/28;16(5).

558 7. Peng Q, Wen T, Liu D, Wang S, Jiang X, Zhao S, et al. DSN1 is a prognostic
559 biomarker and correlated with clinical characterize in breast cancer. INT
560 IMMUNOPHARMACOL. [Journal Article]. 2021 2021/12/1;101(Pt B):107605.

561 8. Chuang TP, Wang JY, Jao SW, Wu CC, Chen JH, Hsiao KH, et al.
562 Over-expression of AURKA, SKA3 and DSN1 contributes to colorectal adenoma to
563 carcinoma progression. Oncotarget. [Journal Article]. 2016
564 2016/7/19;7(29):45803-18.

565 9. Li FN, Zhang QY, Li O, Liu SL, Yang ZY, Pan LJ, et al. ESRRA promotes
566 gastric cancer development by regulating the CDC25C/CDK1/CyclinB1 pathway
567 via DSN1. INT J BIOL SCI. [Journal Article; Research Support, Non-U.S. Gov't].
568 2021 2021/1/20;17(8):1909-24.

569 10. Sun C, Huang S, Ju W, Hou Y, Wang Z, Liu Y, et al. Elevated DSN1

570 expression is associated with poor survival in patients with hepatocellular
571 carcinoma. HUM PATHOL. [Journal Article; Research Support, Non-U.S. Gov't].
572 2018 2018/11/1;81:113-20.

573 11. Li Q, Liang J, Chen B. Identification of CDCA8, DSN1 and BIRC5 in
574 Regulating Cell Cycle and Apoptosis in Osteosarcoma Using Bioinformatics and
575 Cell Biology. TECHNOL CANCER RES T. [Journal Article]. 2020
576 2020/1/1;19:1079233253.

577 12. Huang D, Xiao F, Hao H, Hua F, Luo Z, Huang Z, et al. JARID1B promotes
578 colorectal cancer proliferation and Wnt/beta-catenin signaling via decreasing CDX2
579 level. CELL COMMUN SIGNAL. [Journal Article; Research Support, Non-U.S.
580 Gov't]. 2020 2020/10/27;18(1):169.

581 13. Huang M, Yang S, Tai W, Zhang L, Zhou Y, Cho W, et al. Bioinformatics
582 Identification of Regulatory Genes and Mechanism Related to Hypoxia-Induced
583 PD-L1 Inhibitor Resistance in Hepatocellular Carcinoma. INT J MOL SCI. [Journal
584 Article]. 2023 2023/5/13;24(10).

585 14. Yu X, Pan M, Jiang L, Liu K. NDC80 Kinetochore Complex Serves as a
586 Potential Prognostic Predictor and Correlates with Immune Infiltrates in Epithelial
587 Ovarian Cancer Patients. INT J GEN MED. [Journal Article]. 2024
588 2024/1/20;17:1789-805.

589 15. Ye W, Liang X, Chen G, Chen Q, Zhang H, Zhang N, et al. NDC80/HEC1
590 promotes macrophage polarization and predicts glioma prognosis via single-cell
591 RNA-seq and in vitro experiment. CNS NEUROSCI THER. [Journal Article]. 2024

592 2024/7/1;30(7):e14850.

593 16. Li W, Xu Y, Zeng X, Tan J, Wang Y, Wu H, et al. Etiological relationship
594 between lipid metabolism and endometrial carcinoma. LIPIDS HEALTH DIS.
595 [Journal Article; Review]. 2023 2023/8/4;22(1):116.

596 17. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al.
597 Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality
598 Worldwide for 36 Cancers in 185 Countries. CA-CANCER J CLIN. [Journal Article].
599 2021 2021/5/1;71(3):209-49.

600 18. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh
601 N. Endometrial cancer. LANCET. [Journal Article; Research Support, Non-U.S.
602 Gov't; Review]. 2022 2022/4/9;399(10333):1412-28.

603 19. Behrouzi R, Barr CE, Crosbie EJ. HE4 as a Biomarker for Endometrial
604 Cancer. CANCERS. [Journal Article; Review]. 2021 2021/9/23;13(19).

605 20. Navarro-Mendoza MI, Perez-Arques C, Panchal S, Nicolas FE, Mondo SJ,
606 Ganguly P, et al. Early Diverging Fungus Mucor circinelloides Lacks Centromeric
607 Histone CENP-A and Displays a Mosaic of Point and Regional Centromeres.
608 CURR BIOL. [Journal Article; Research Support, N.I.H., Extramural; Research
609 Support, Non-U.S. Gov't; Research Support, U.S. Gov't, Non-P.H.S.]. 2019
610 2019/11/18;29(22):3791-802.

611 21. Wang X, Lu X, Wang P, Chen Q, Xiong L, Tang M, et al. SRSF9 promotes
612 colorectal cancer progression via stabilizing DSN1 mRNA in an m6A-related
613 manner. J TRANSL MED. [Journal Article]. 2022 2022/5/4;20(1):198.

614 22. Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer.
615 NATURE. [Journal Article; Review]. 2001 2001/5/17;411(6835):342-8.
616 23. Cheung AH, Hui CH, Wong KY, Liu X, Chen B, Kang W, et al. Out of the
617 cycle: Impact of cell cycle aberrations on cancer metabolism and metastasis. INT J
618 CANCER. [Journal Article; Research Support, Non-U.S. Gov't; Review]. 2023
619 2023/4/15;152(8):1510-25.
620 24. Meng QC, Wang HC, Song ZL, Shan ZZ, Yuan Z, Zheng Q, et al.
621 Overexpression of NDC80 is correlated with prognosis of pancreatic cancer and
622 regulates cell proliferation. AM J CANCER RES. [Journal Article]. 2015
623 2015/1/20;5(5):1730-40.
624 25. Qu Y, Li J, Cai Q, Liu B. Hec1/Ndc80 is overexpressed in human gastric
625 cancer and regulates cell growth. J GASTROENTEROL. [Journal Article; Research
626 Support, Non-U.S. Gov't]. 2014 2014/3/1;49(3):408-18.
627 26. Chen Y, Yao Y, Zhang L, Li X, Wang Y, Zhao L, et al. cDNA microarray
628 analysis and immunohistochemistry reveal a distinct molecular phenotype in serous
629 endometrial cancer compared to endometrioid endometrial cancer. EXP MOL
630 PATHOL. [Comparative Study; Journal Article; Research Support, Non-U.S. Gov't].
631 2011 2011/8/1;91(1):373-84.
632 27. Xiao Y, Yu D. Tumor microenvironment as a therapeutic target in cancer.
633 PHARMACOL THERAPEUT. [Journal Article; Research Support, N.I.H.,
634 Extramural; Research Support, Non-U.S. Gov't; Review]. 2021 2021/5/1;221:107753.
635 28. Palucka AK, Coussens LM. The Basis of Oncoimmunology. CELL. [Journal

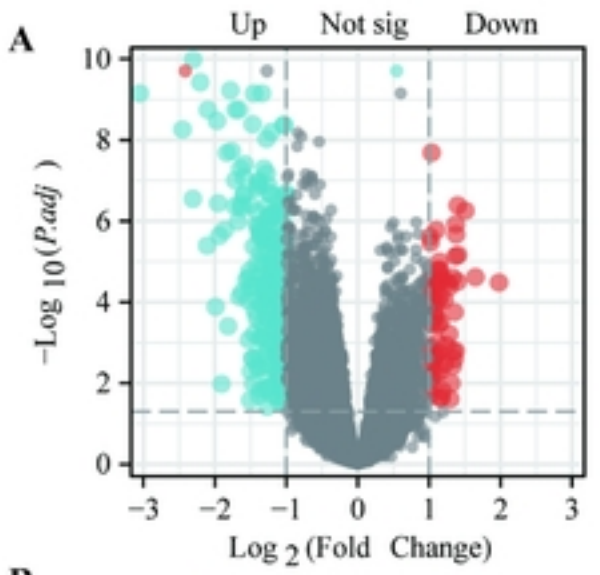
636 Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't;
637 Research Support, U.S. Gov't, Non-P.H.S.; Review]. 2016 2016/3/10;164(6):1233-47.
638 29. Sheu BC, Lin RH, Lien HC, Ho HN, Hsu SM, Huang SC. Predominant
639 Th2/Tc2 polarity of tumor-infiltrating lymphocytes in human cervical cancer. J
640 IMMUNOL. [Journal Article; Research Support, Non-U.S. Gov't]. 2001
641 2001/9/1;167(5):2972-8.
642 30. Chen B, Li H, Xia W. Imiquimod regulating Th1 and Th2 cell-related
643 chemokines to inhibit scar hyperplasia. INT WOUND J. [Journal Article]. 2019
644 2019/12/1;16(6):1281-8.
645 31. Jin R, Zhou M, Lin F, Xu W, Xu A. Pathogenic Th2 Cytokine Profile
646 Skewing by IFN-gamma-Responding Vitiligo Fibroblasts via CCL2/CCL8.
647 CELLS-BASEL. [Journal Article; Research Support, Non-U.S. Gov't]. 2023
648 2023/1/4;12(2).

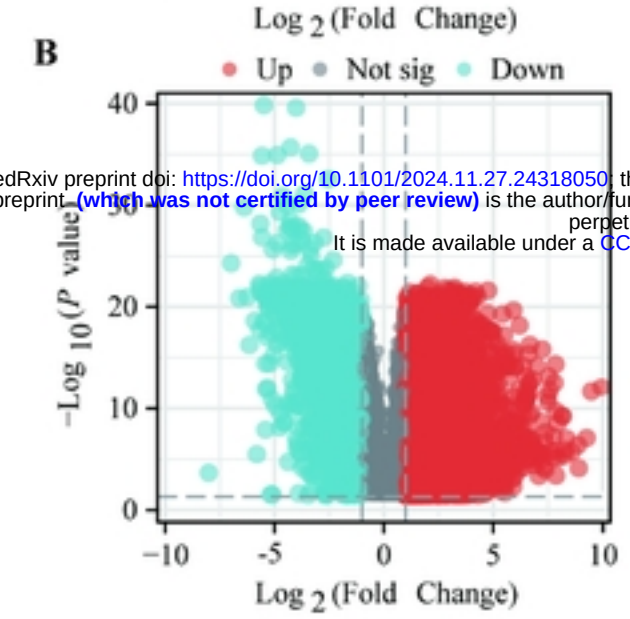
649

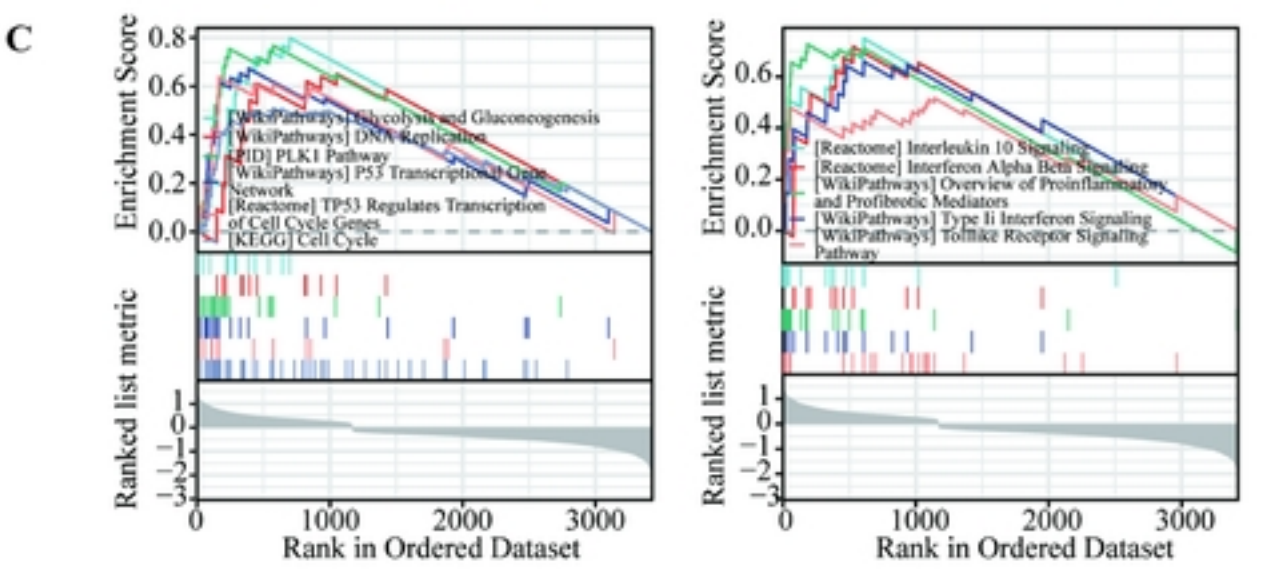
650 **Supplementary Figure captions**

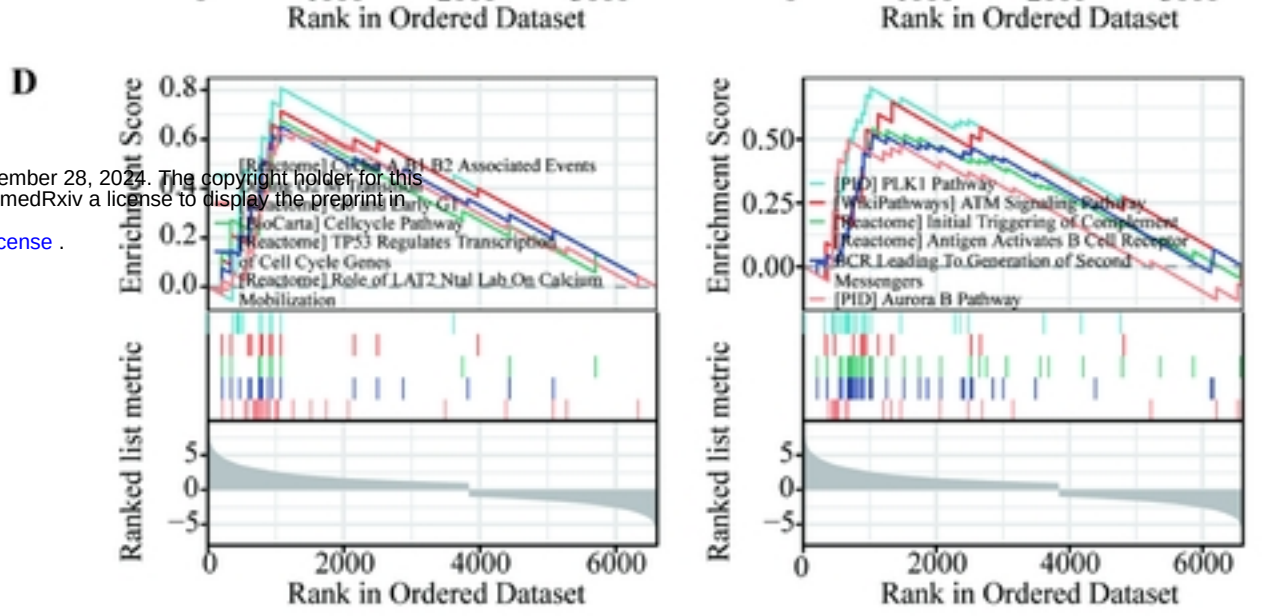
651 **Supplementary Figure 1. The survival curves of 12 genes for which they were not**
652 **statistically significant.** (A) 12 genes with no statistically significant survival curves.
653 **Supplementary Figure 2. Correlation analysis and survival analysis of four genes**
654 **for which survival analysis was not meaningful.** The correlation scatter plot for
655 genes without survival statistical significance for (A) AURKB, MAD2L1, NUF2, and
656 ZWINT. The survival curves of the genes include (B) AURKB, MAD2L1, NUF2, and
657 ZWINT.

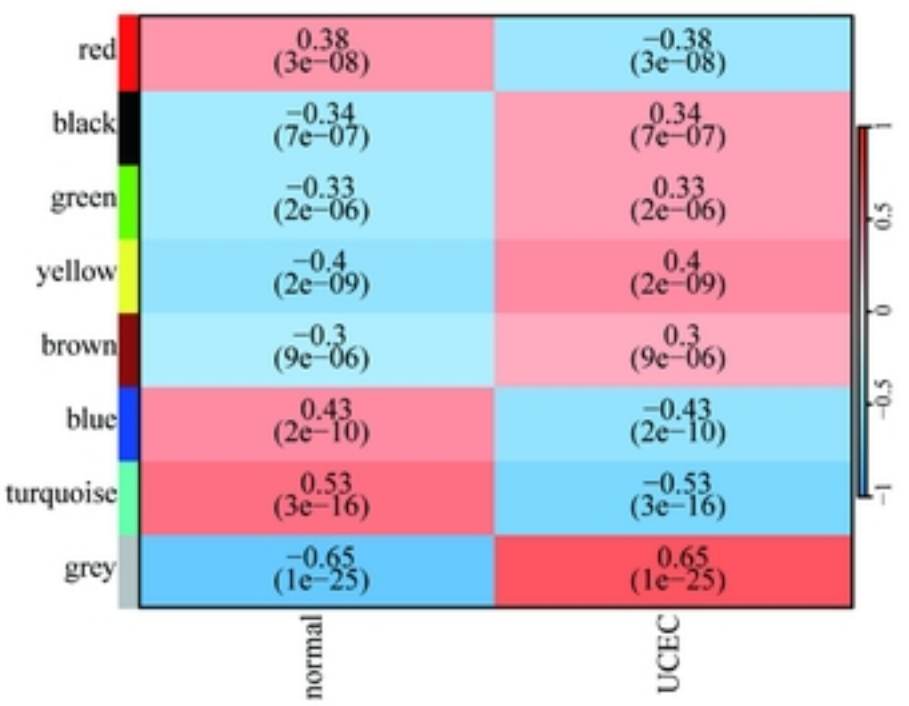
658

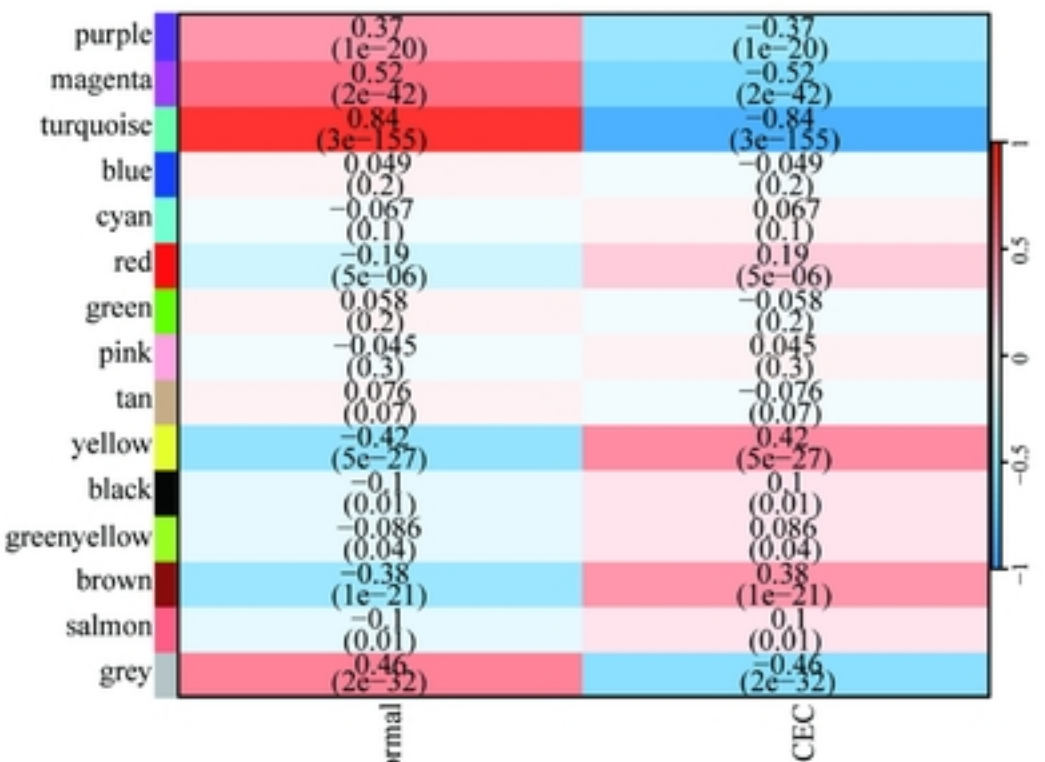


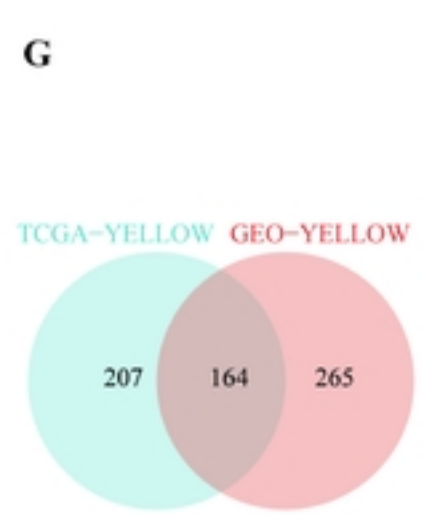


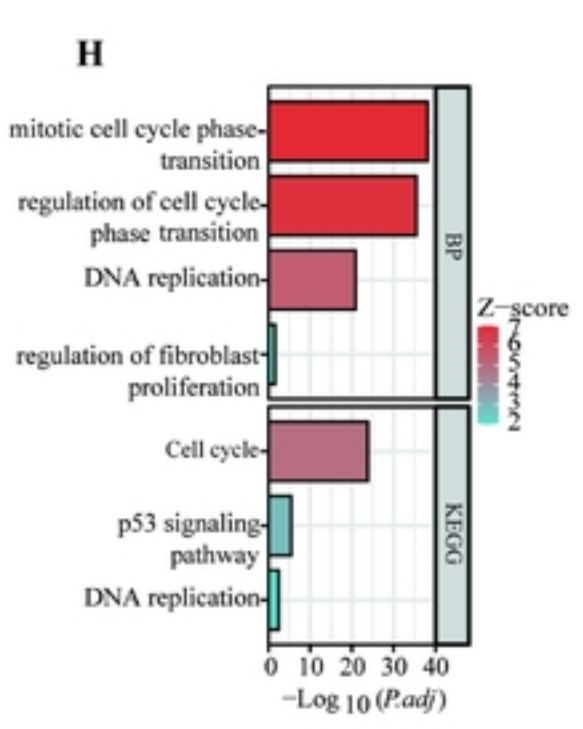


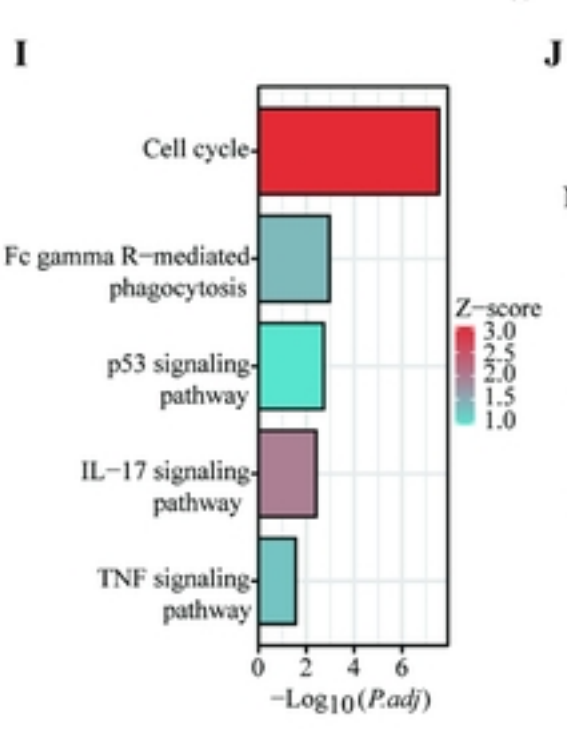












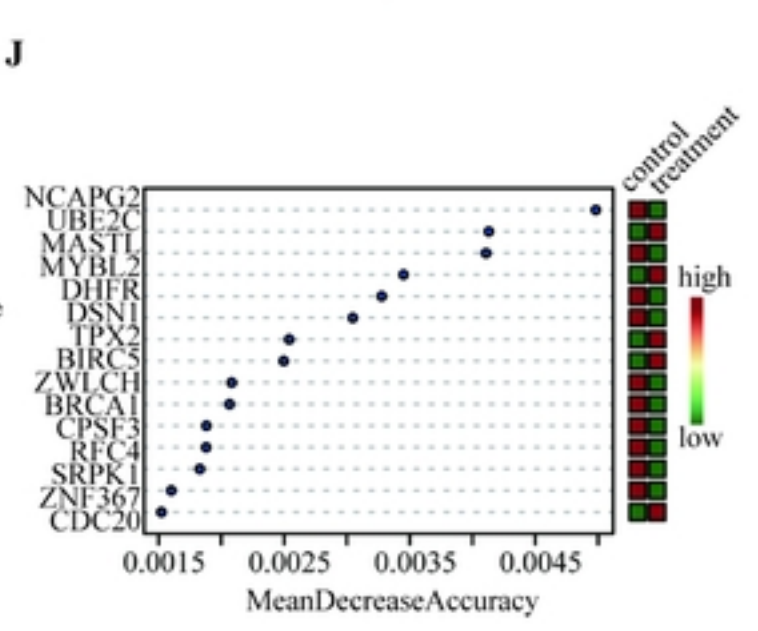
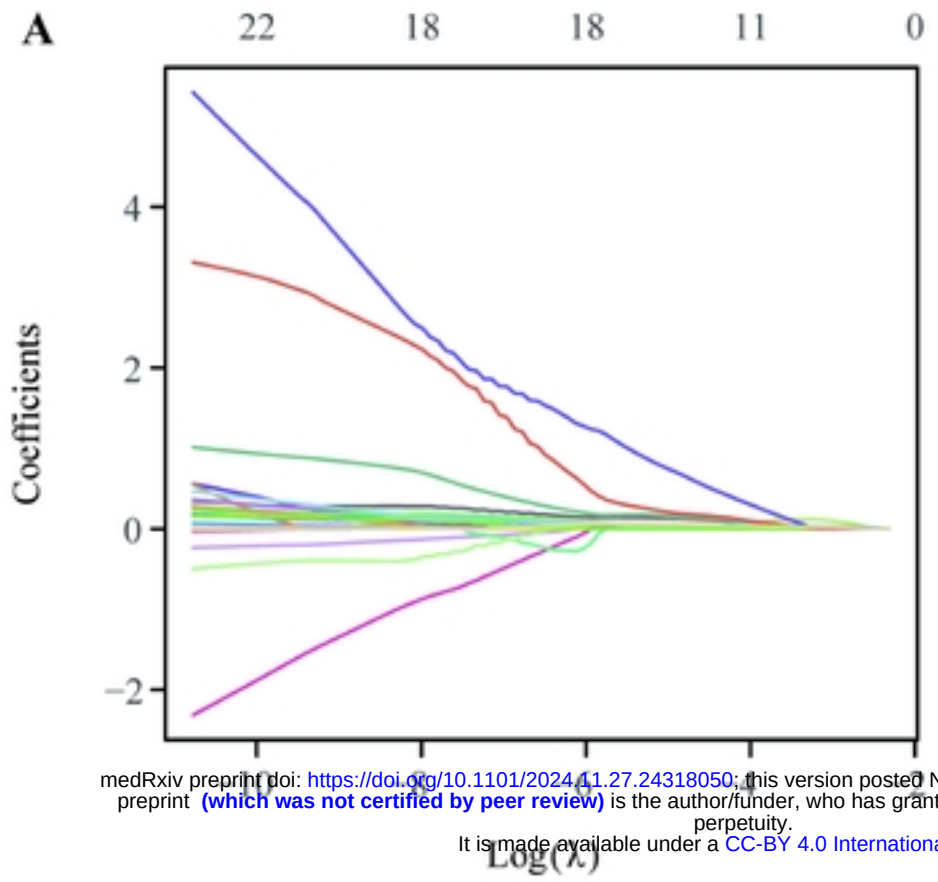
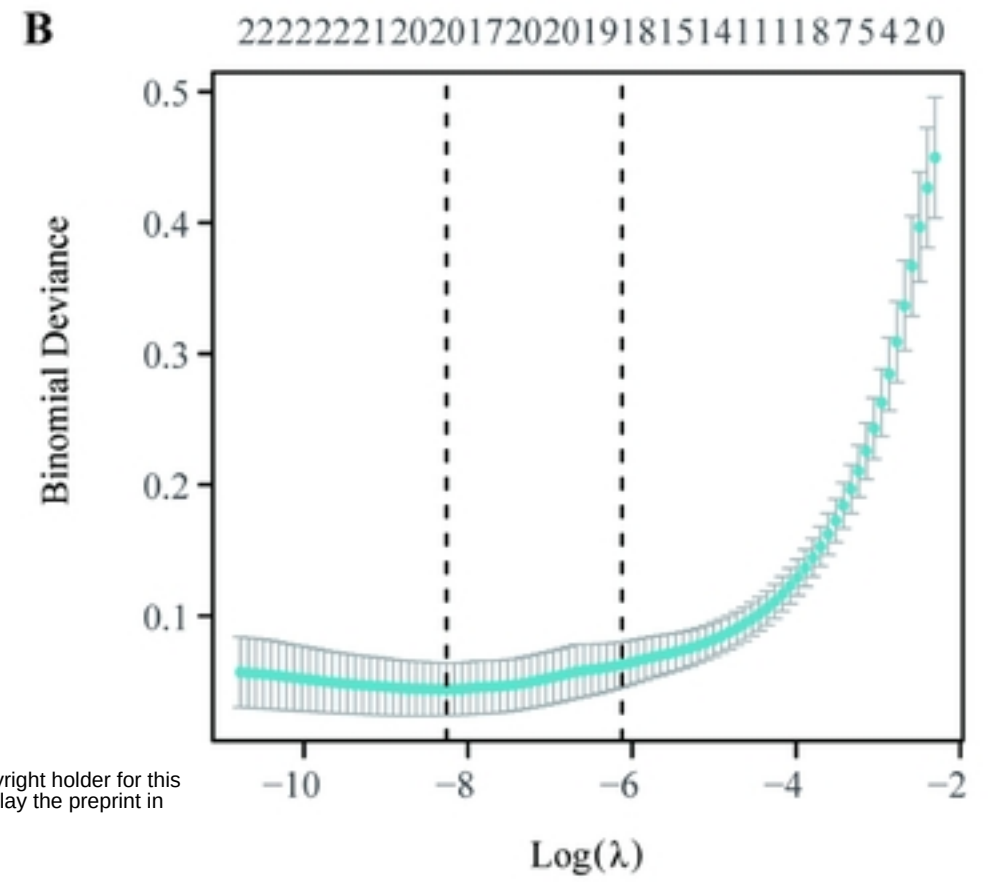
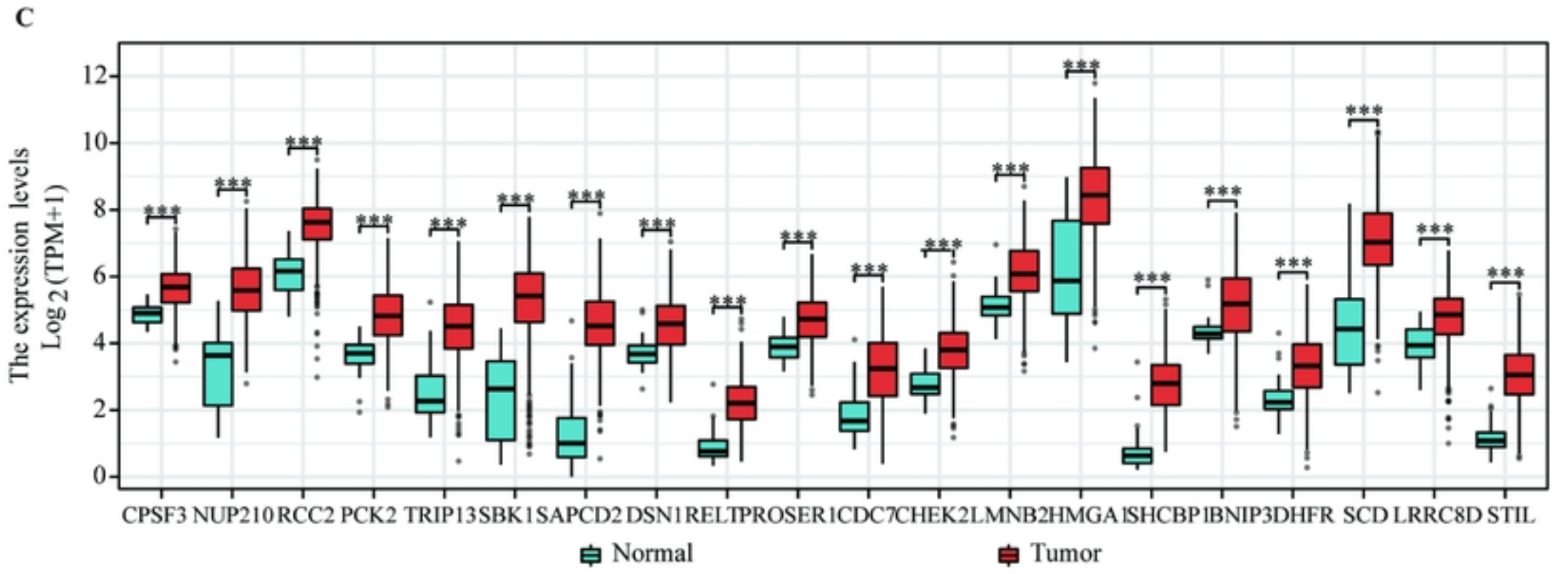
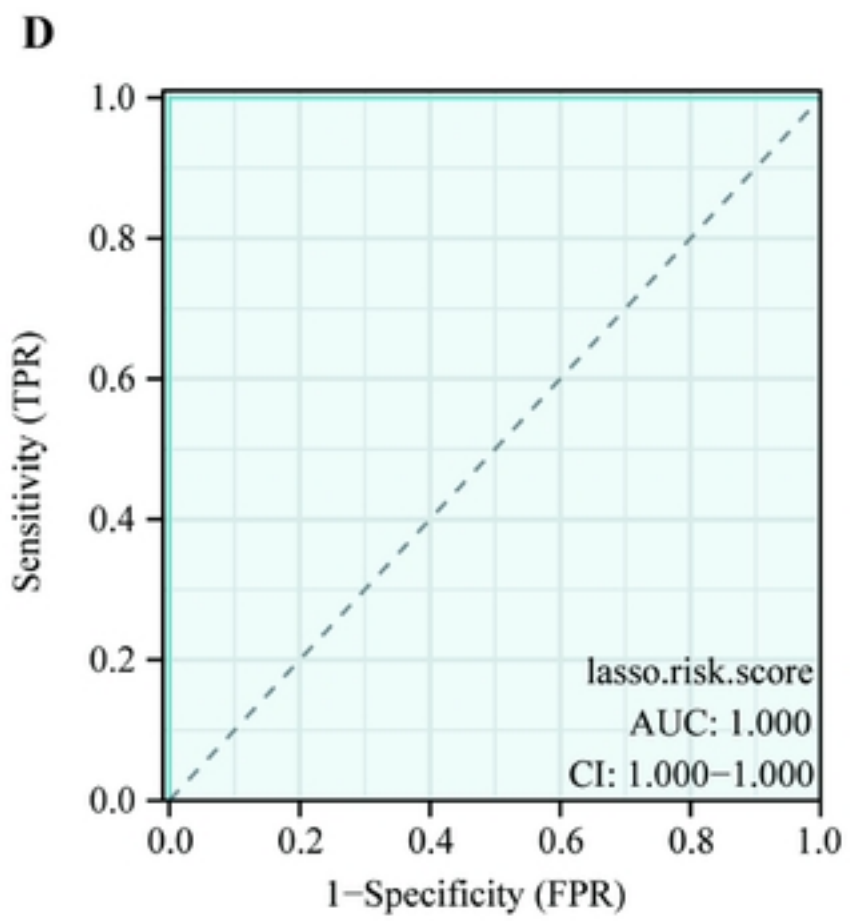


Figure 1









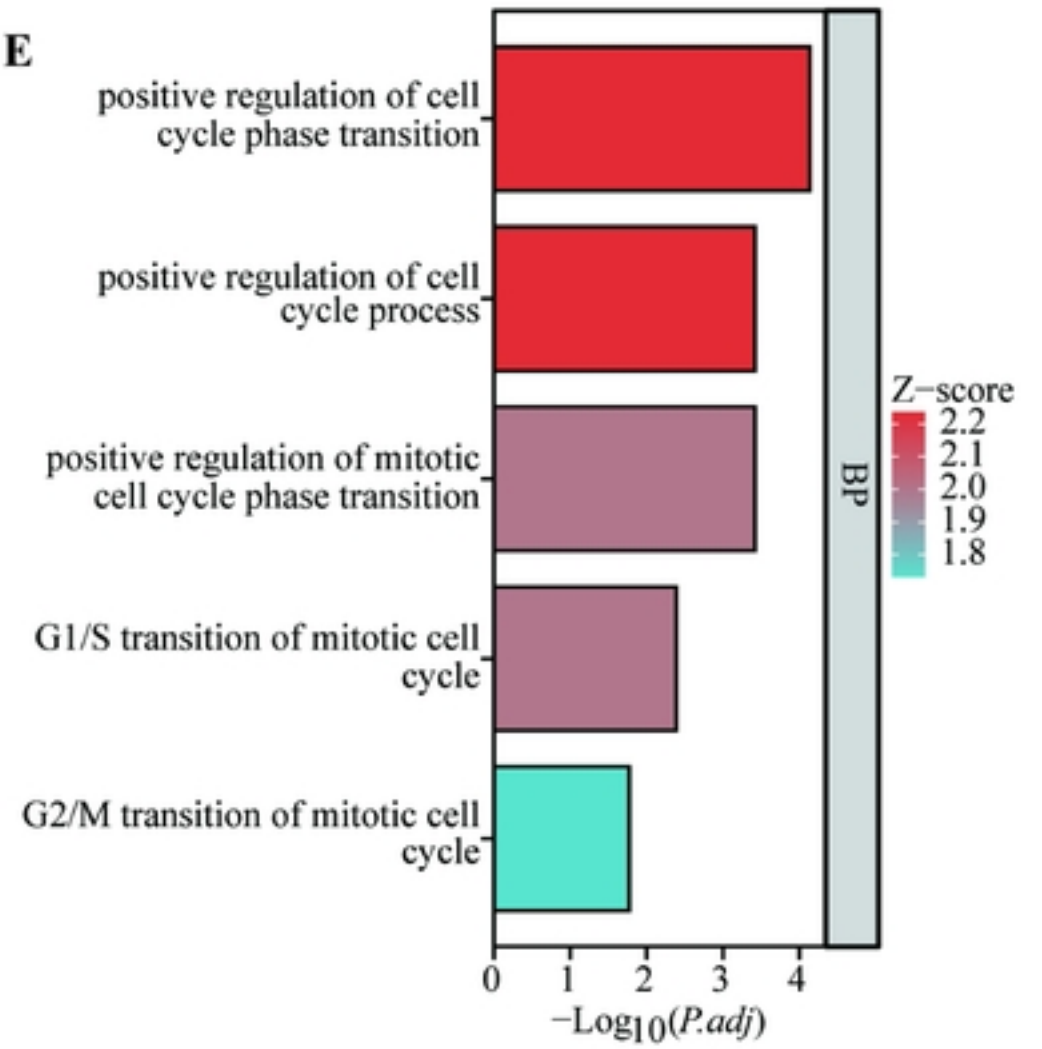
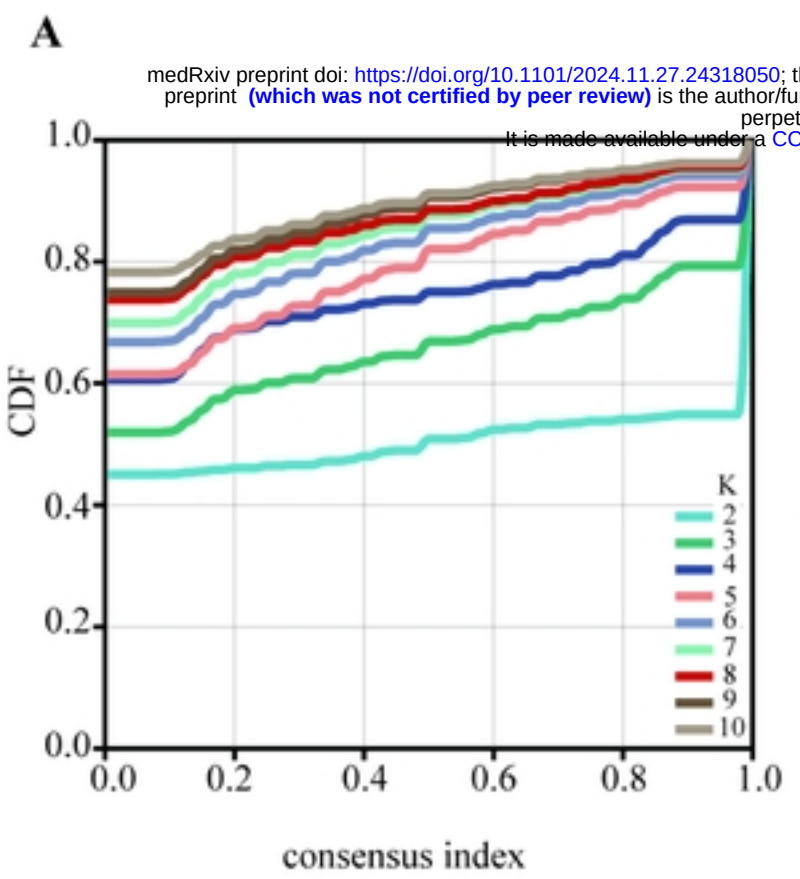
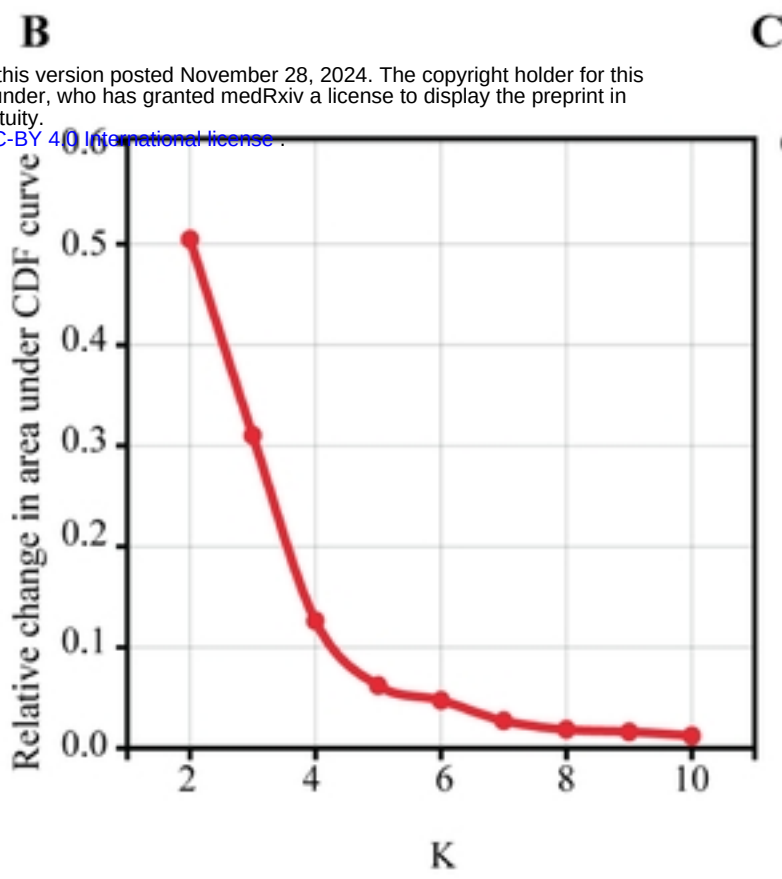
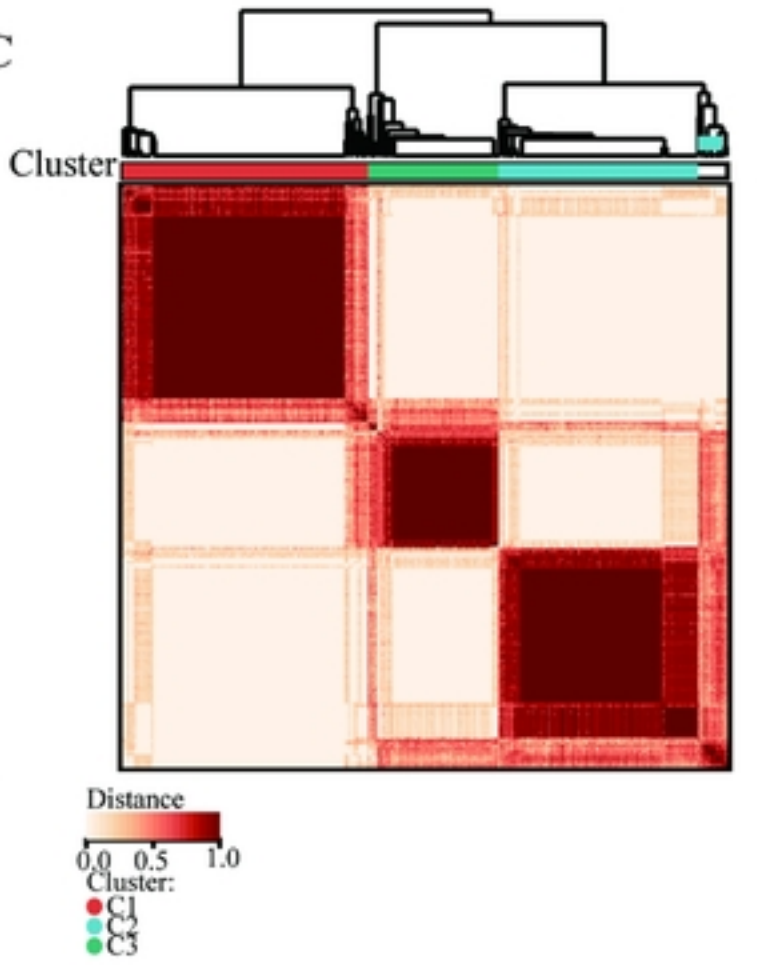
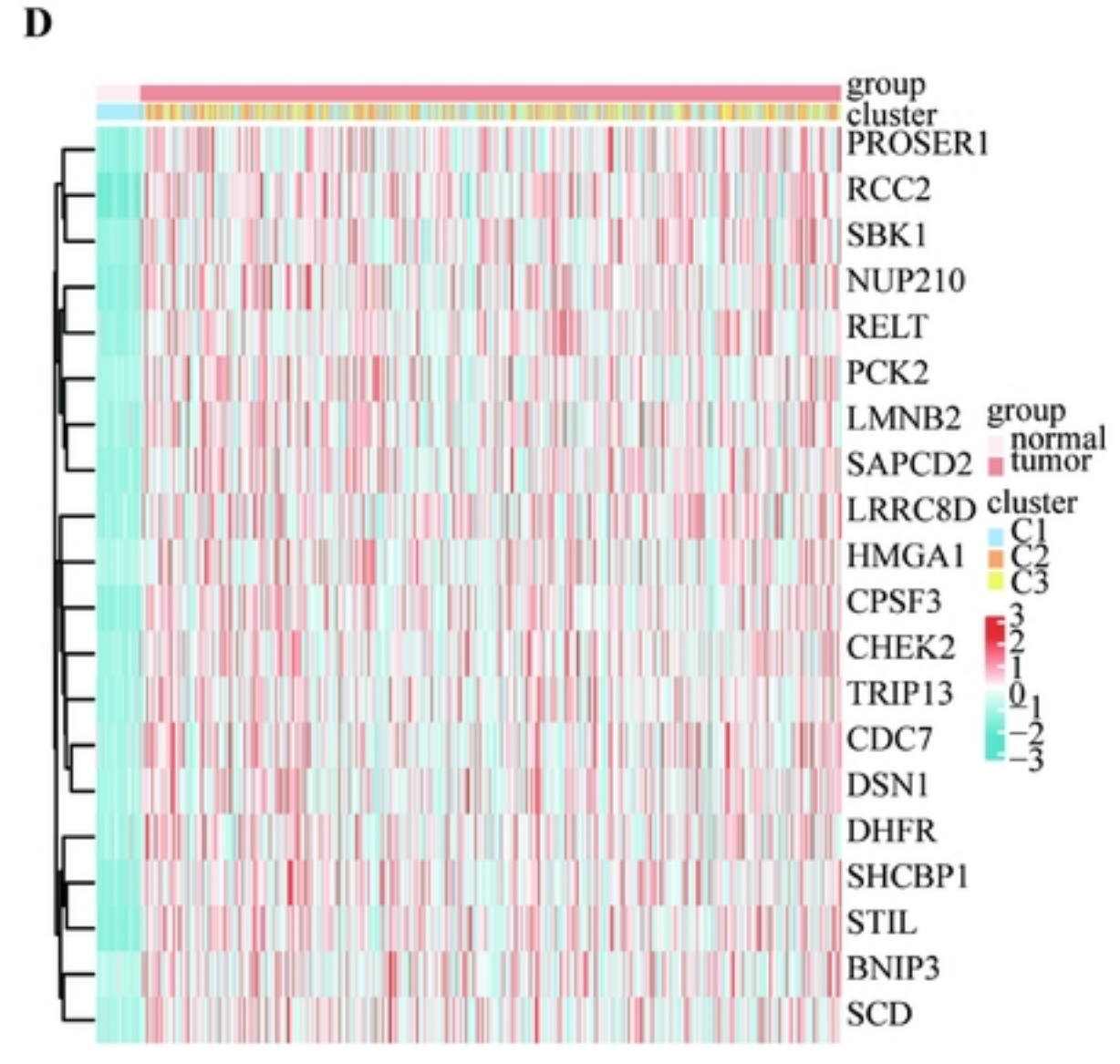


Figure 2









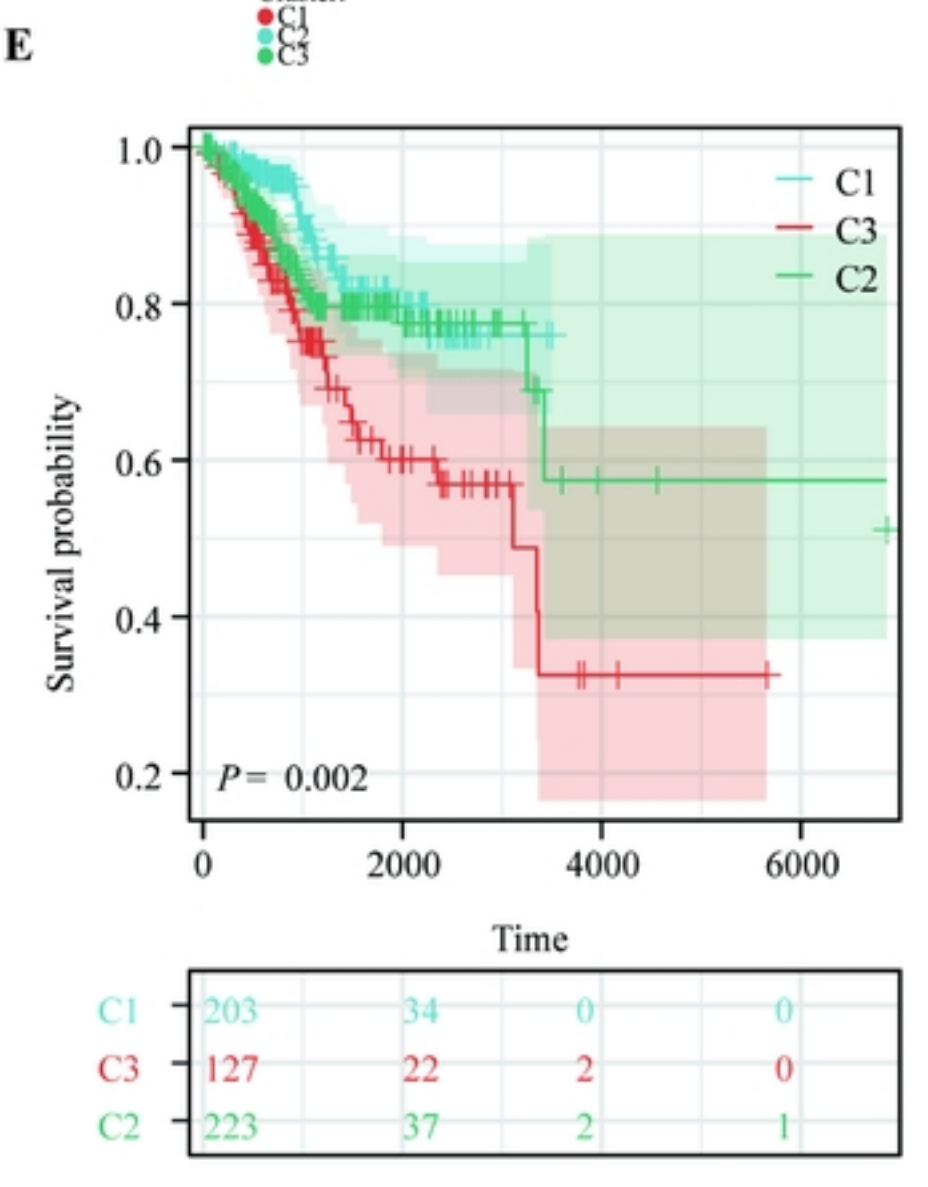
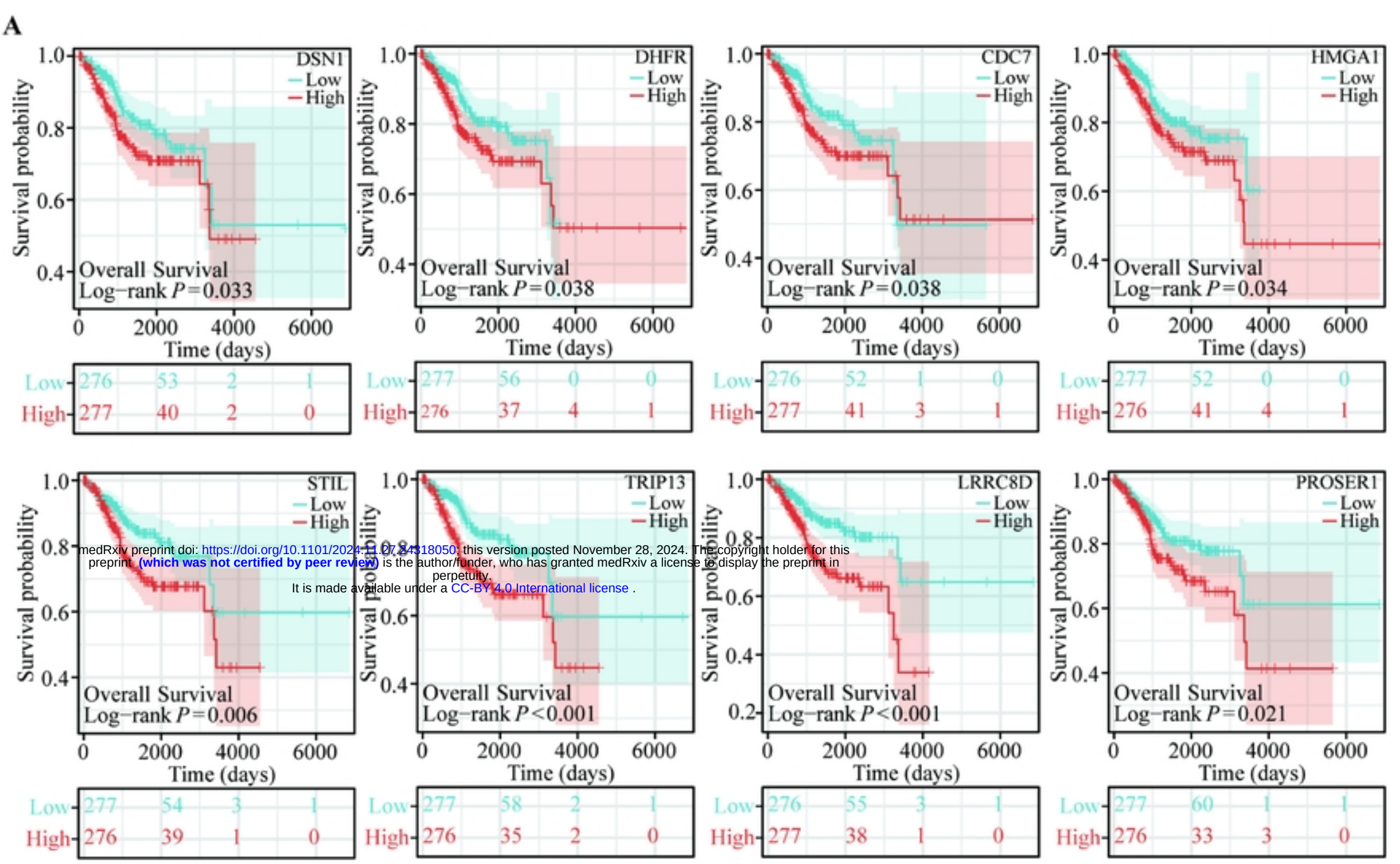
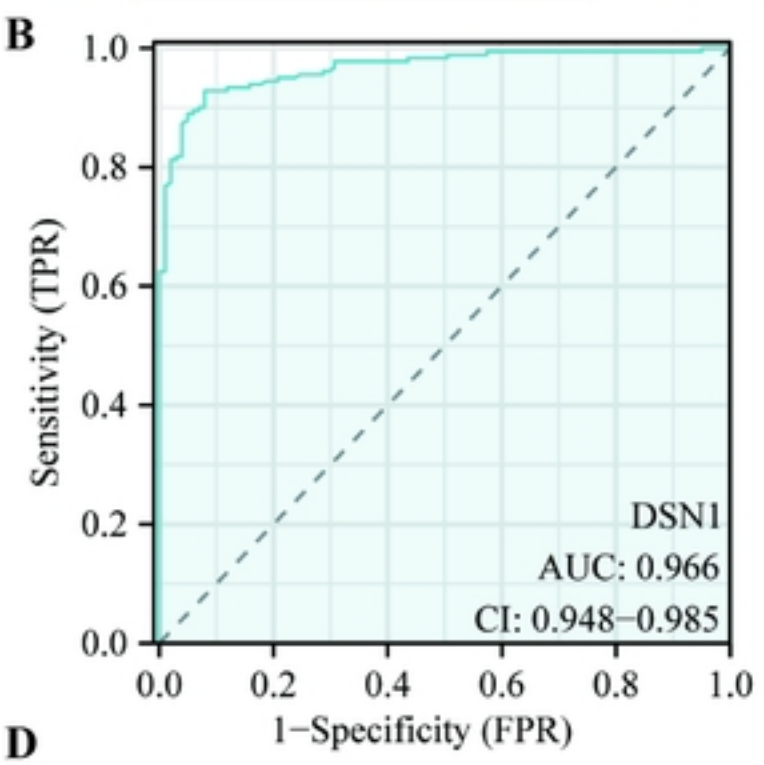
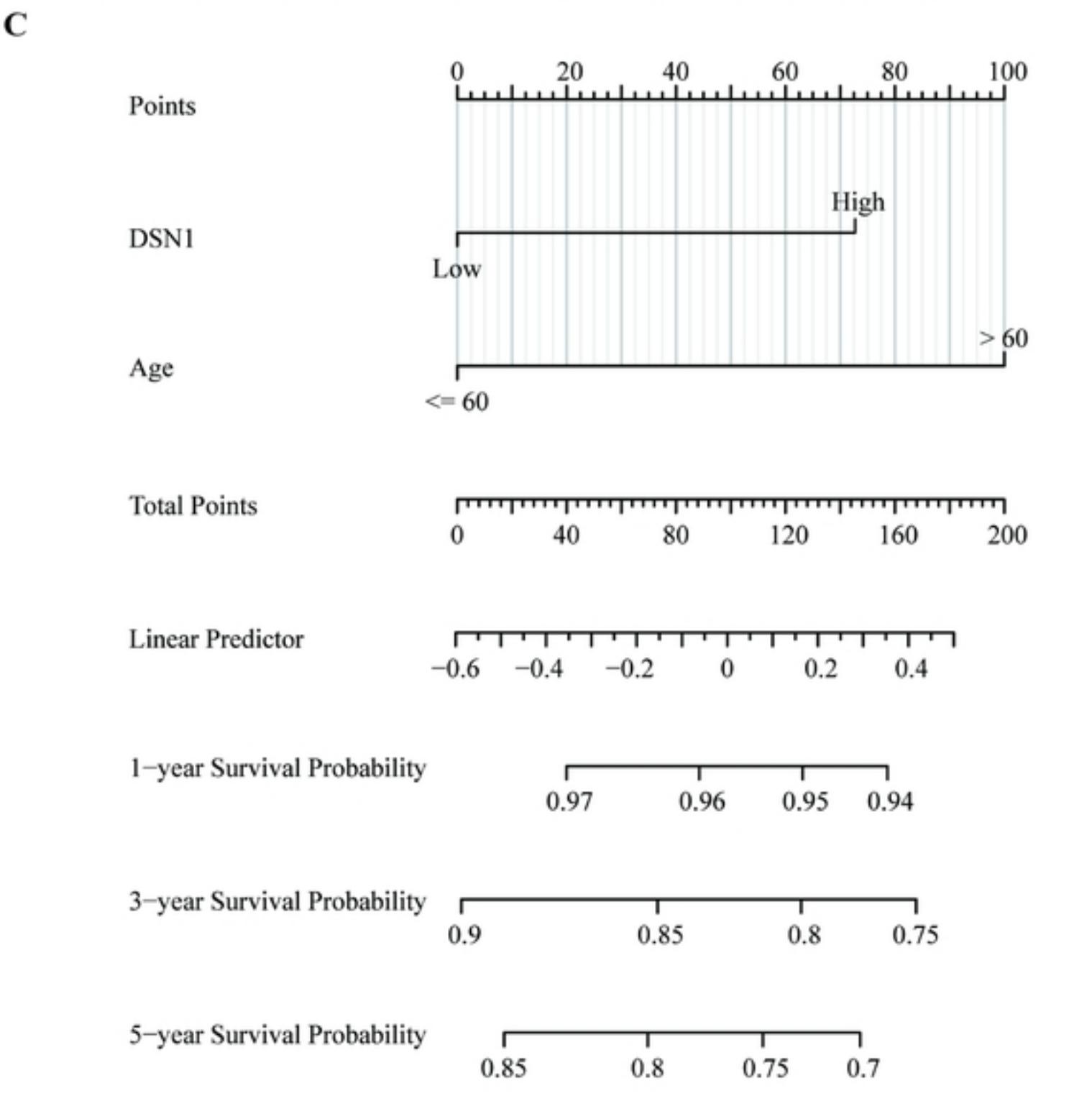


Figure 3







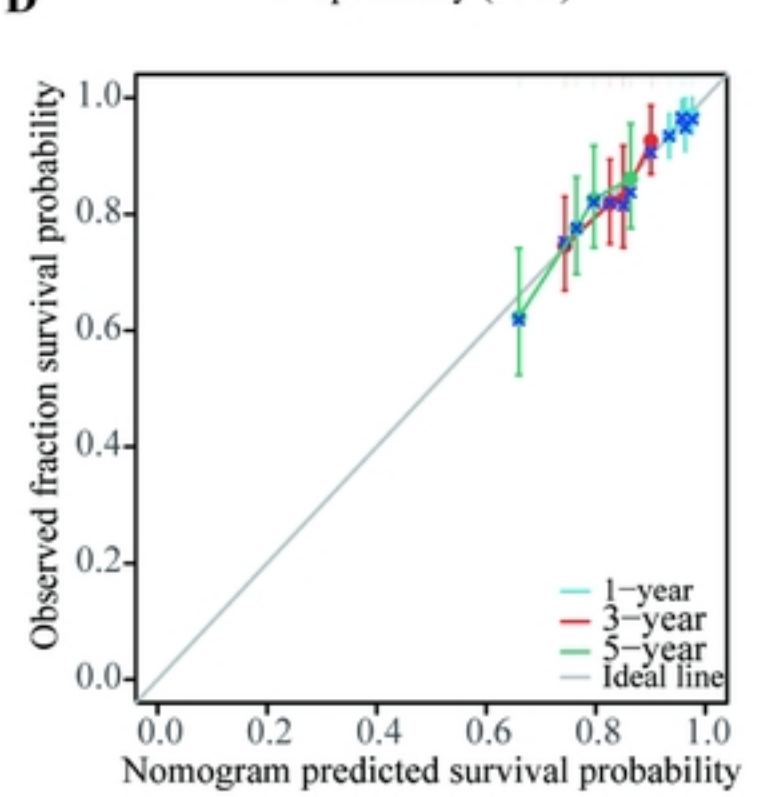
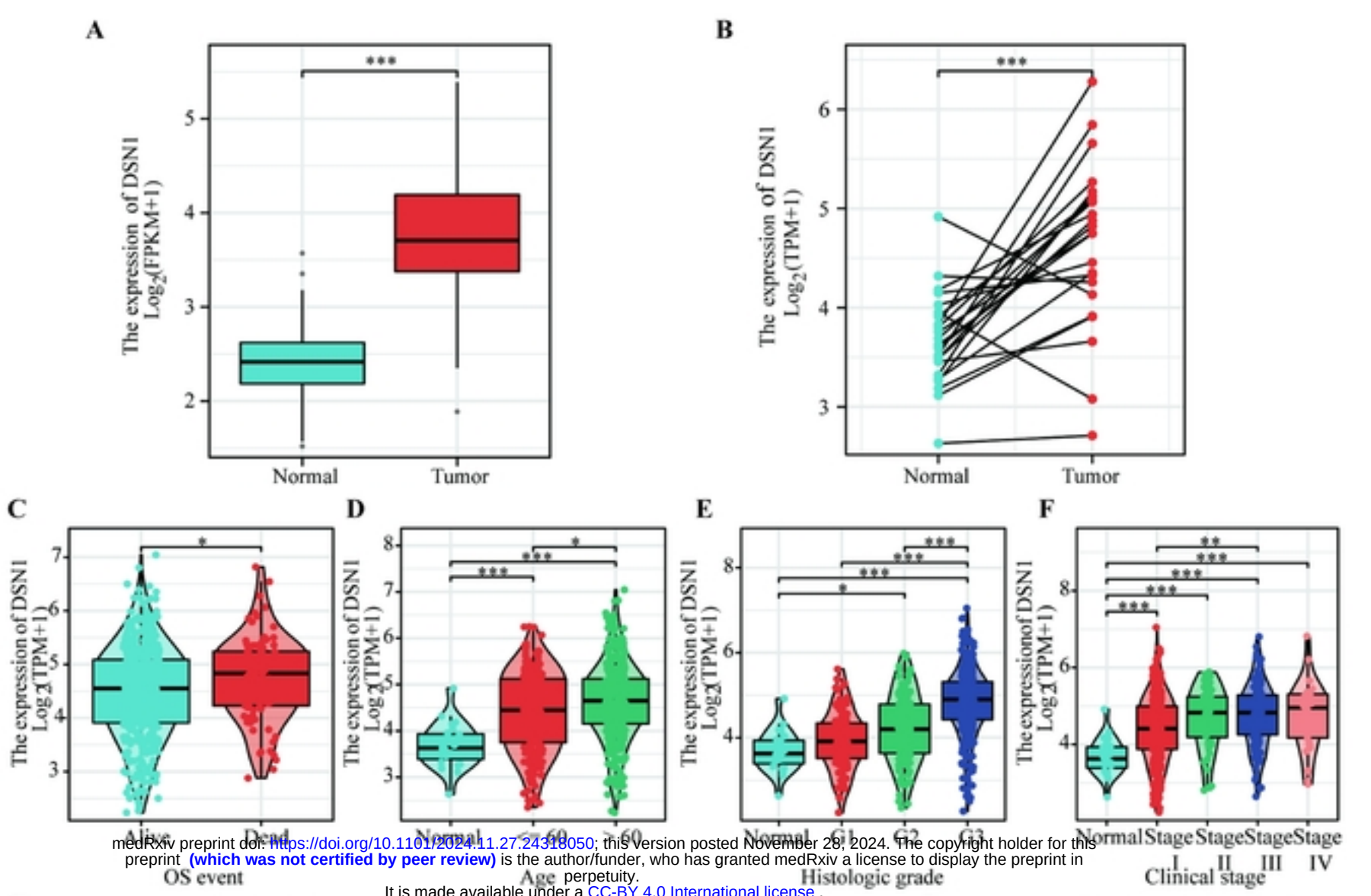
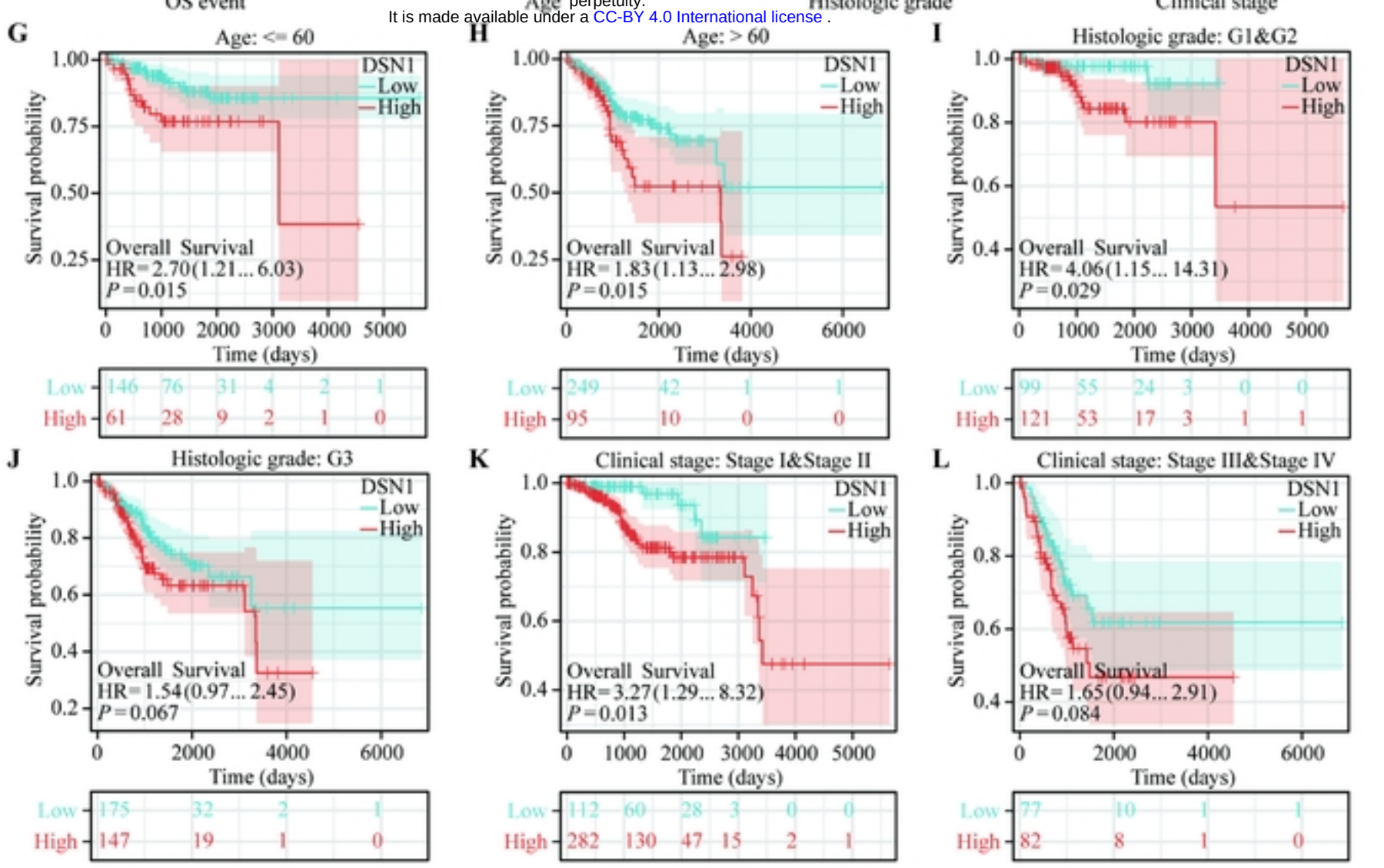


Figure 4





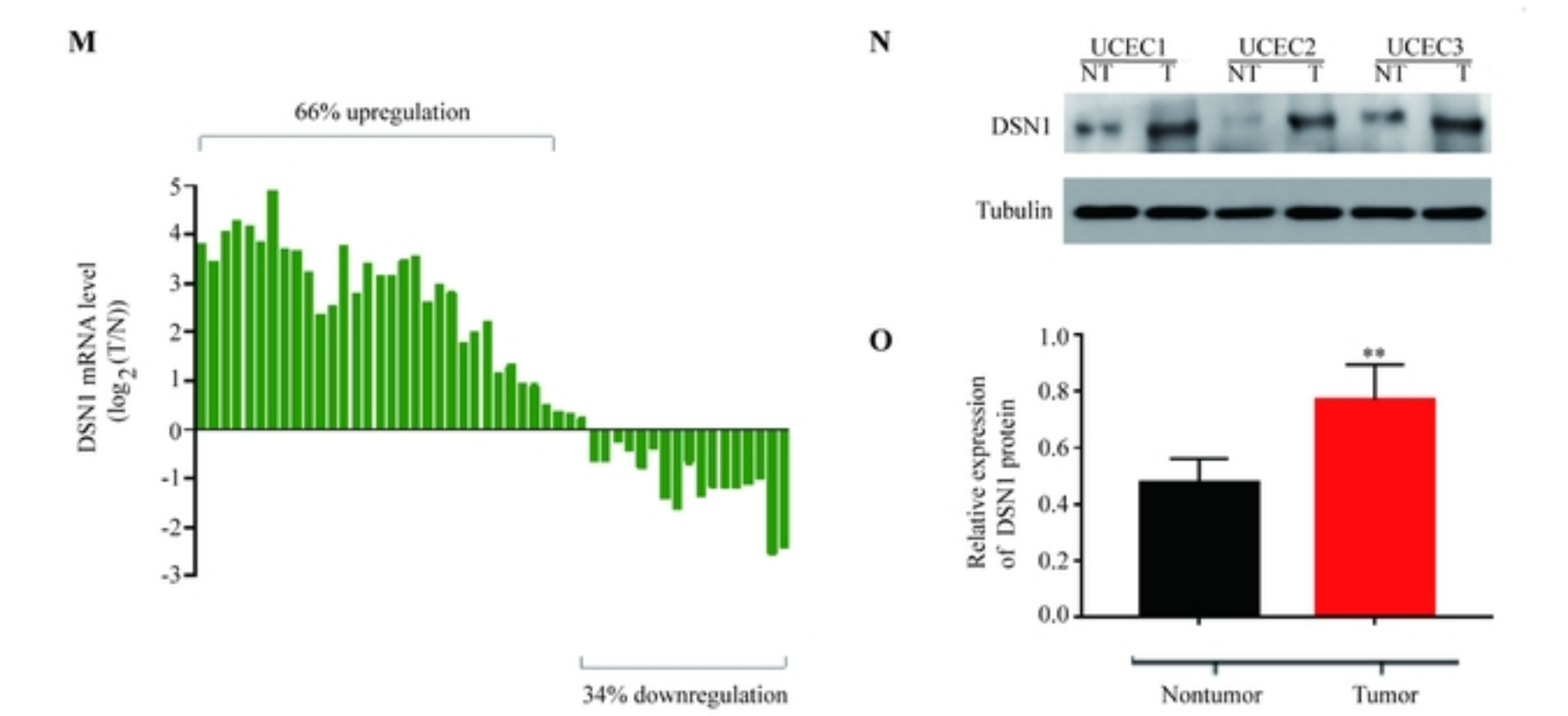


Figure 5

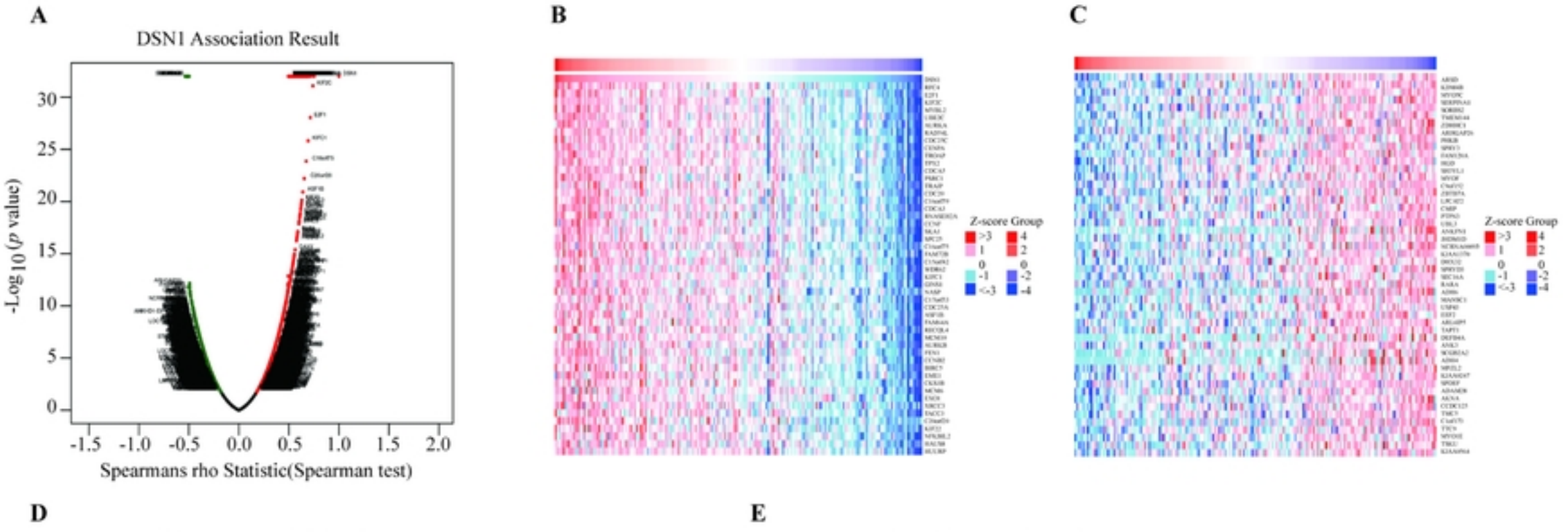
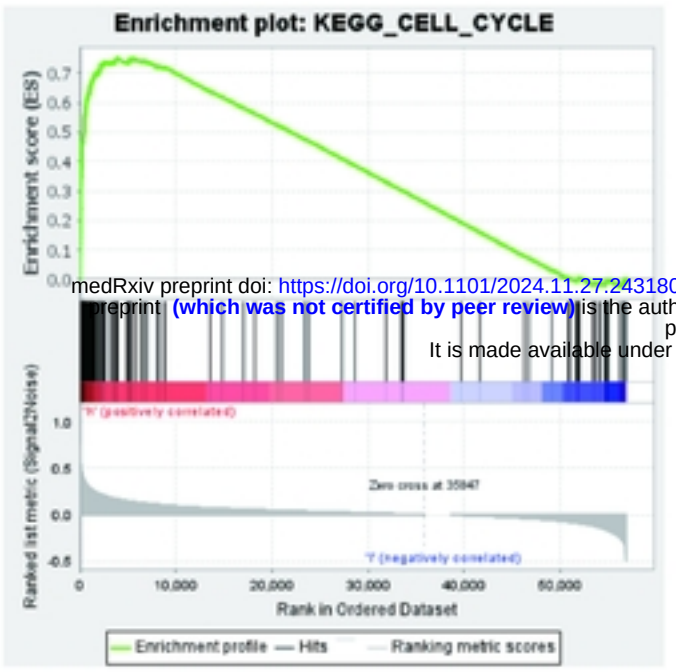
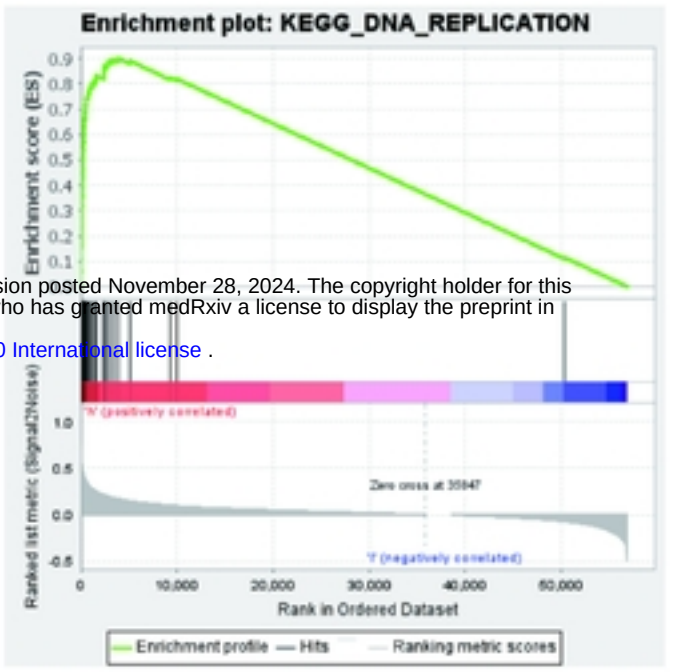
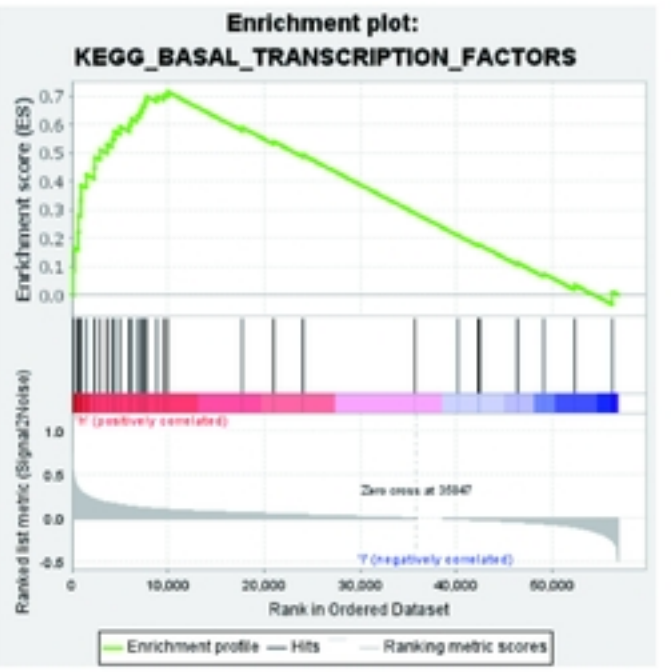
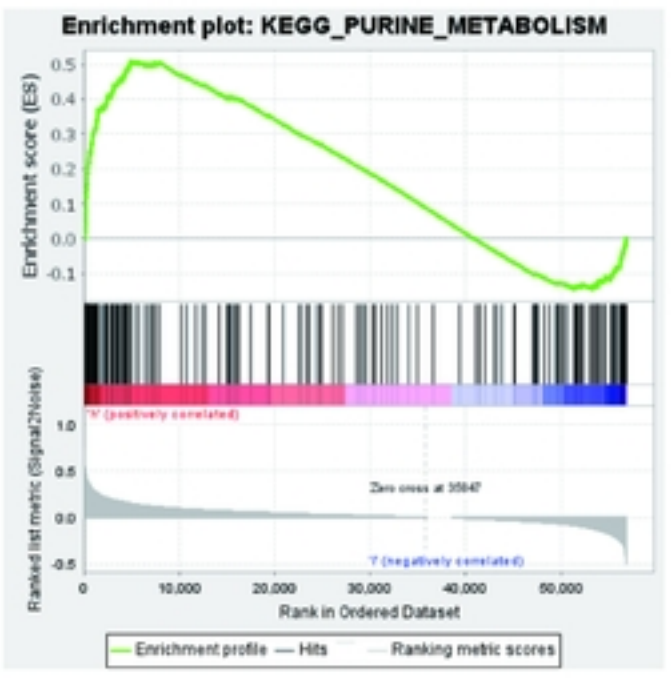

A
DSN1 Association Result
-Log10 (p value)
Spearmans rho Statistic(Spearman test)
B
Z-score Group
C
Z-score Group
D
FDR ≤ 0.05
FDR > 0.05
chromosome segregation
DNA replication
spindle organization
mitotic cell cycle phase transition
cell cycle G2/M phase transition
double-strand break repair
telomere organization
cell cycle checkpoint
chromatin assembly or disassembly
interstrand cross-link repair
DNA recombination
ncRNA processing
negative regulation of cell cycle process
regulation of chromosome organization
RNA splicing
ribonucleoprotein complex biogenesis
RNA 3'-end processing
nucleotide-excision repair
rRNA metabolic process
mRNA processing
base-excision repair
tRNA metabolic process
platelet degranulation
regulation of inflammatory response
type 2 immune response
inositol lipid-mediated signaling
NIK/NF-kappaB signaling
skin development
leukocyte migration
regulation of cell projection assembly
cilium or flagellum-dependent cell motility
cellular defense response
heart valve development
interleukin-6 production
positive regulation of response to external stimulus
positive regulation of cell motility
interleukin-8 production
response to chemokine
chemokine production
tissue remodeling
organ or tissue specific immune response
acute inflammatory response
Normalized Enrichment Score
E
FDR ≤ 0.05
FDR > 0.05
Cell cycle
Spliceosome
DNA replication
Proteasome
Nucleotide excision repair
Fanconi anemia pathway
RNA transport
Homologous recombination
Ribosome biogenesis in eukaryotes
Mismatch repair
Progesterone-mediated oocyte maturation
Ribosome
Oocyte meiosis
Base excision repair
Pyrimidine metabolism
mRNA surveillance pathway
Aminoacyl-tRNA biosynthesis
Parkinson disease
MicroRNAs in cancer
RNA polymerase
One carbon pool by folate
Cellular senescence
Glycosphingolipid biosynthesis
Osteoclast differentiation
Type I diabetes mellitus
Drug metabolism
Proteoglycans in cancer
Parathyroid hormone synthesis, secretion and action
Toxoplasmosis
Rheumatoid arthritis
NF-kappa B signaling pathway
Chagas disease (American trypanosomiasis)
Cytokine-cytokine receptor interaction
IL-17 signaling pathway
Th17 cell differentiation
Circadian rhythm
Graft-versus-host disease
Tyrosine metabolism
Mucin type O-glycan biosynthesis
Leishmaniasis
Inflammatory bowel disease (IBD)
Staphylococcus aureus infection
Amoebiasis
Hematopoietic cell lineage
TNF signaling pathway
Complement and coagulation cascades
Normalized Enrichment Score

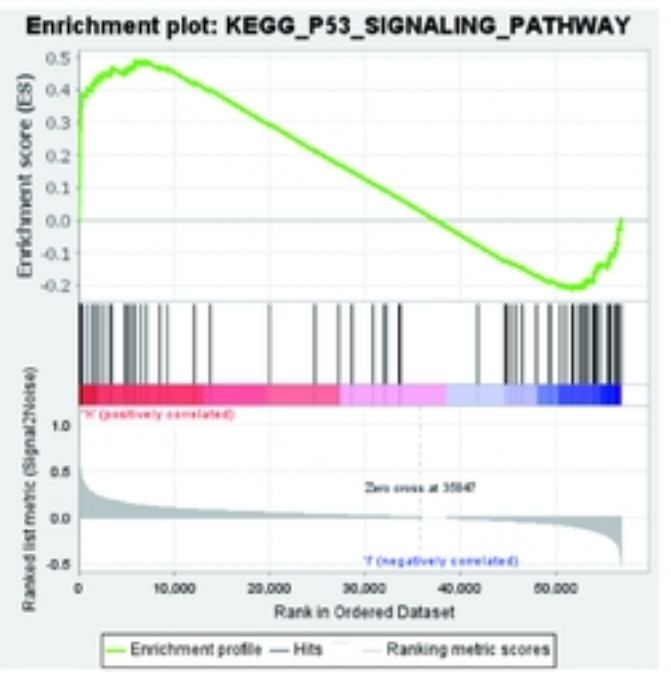
Figure 6

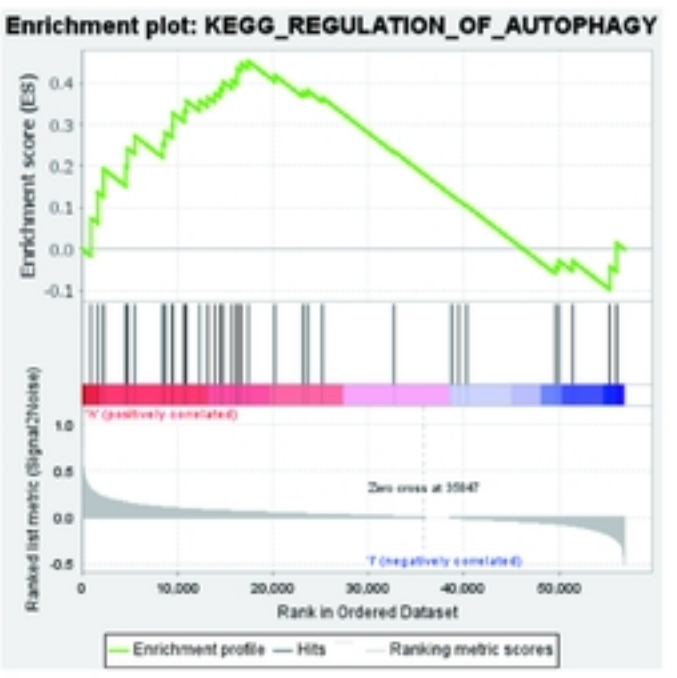


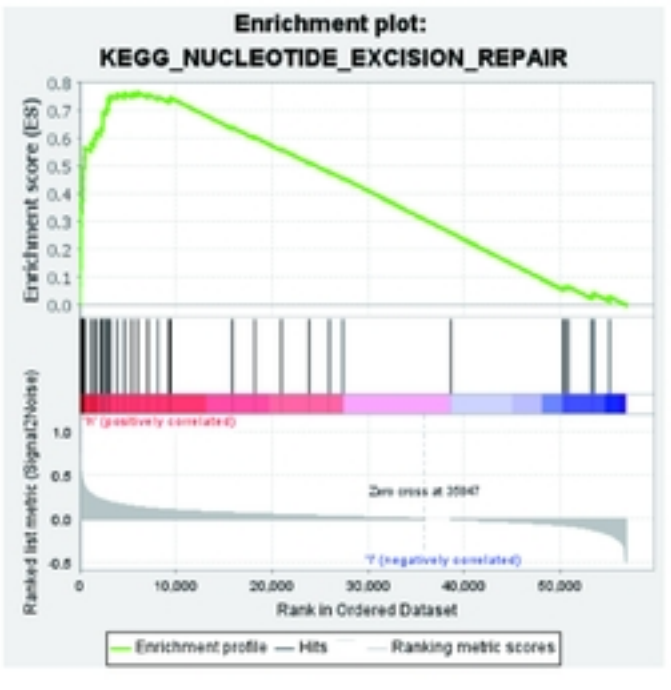


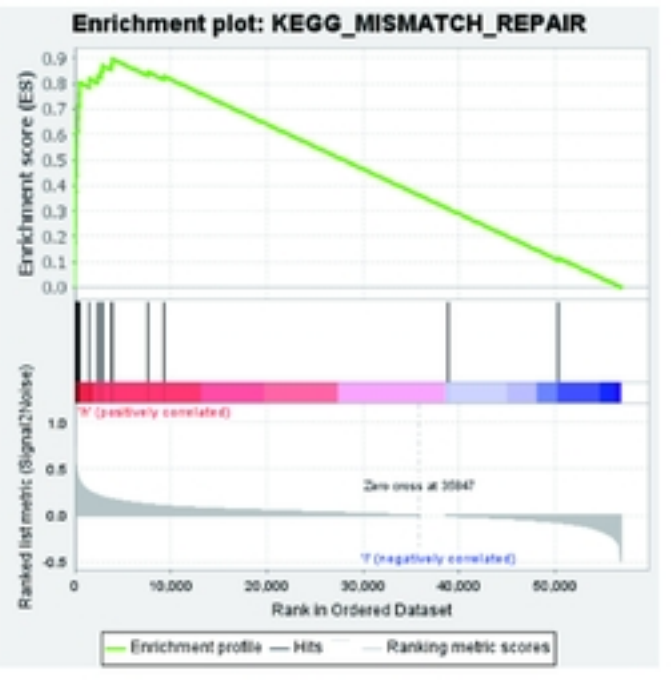












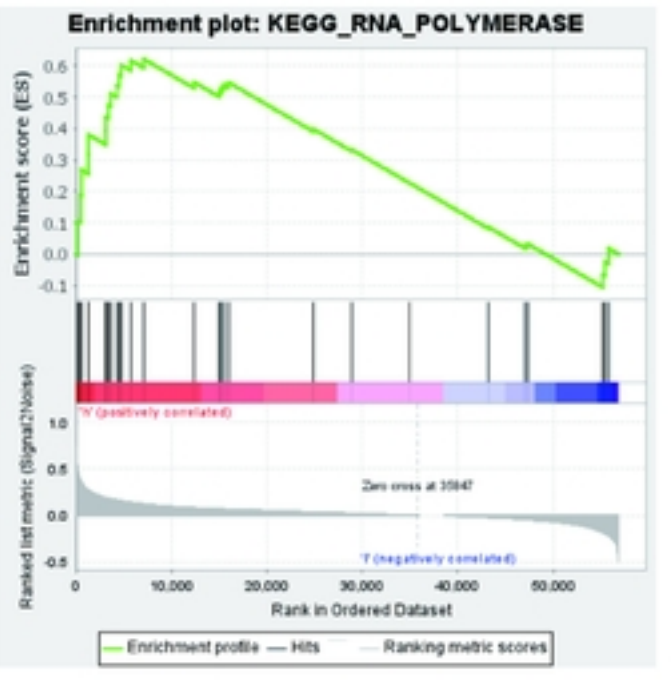


Figure 7

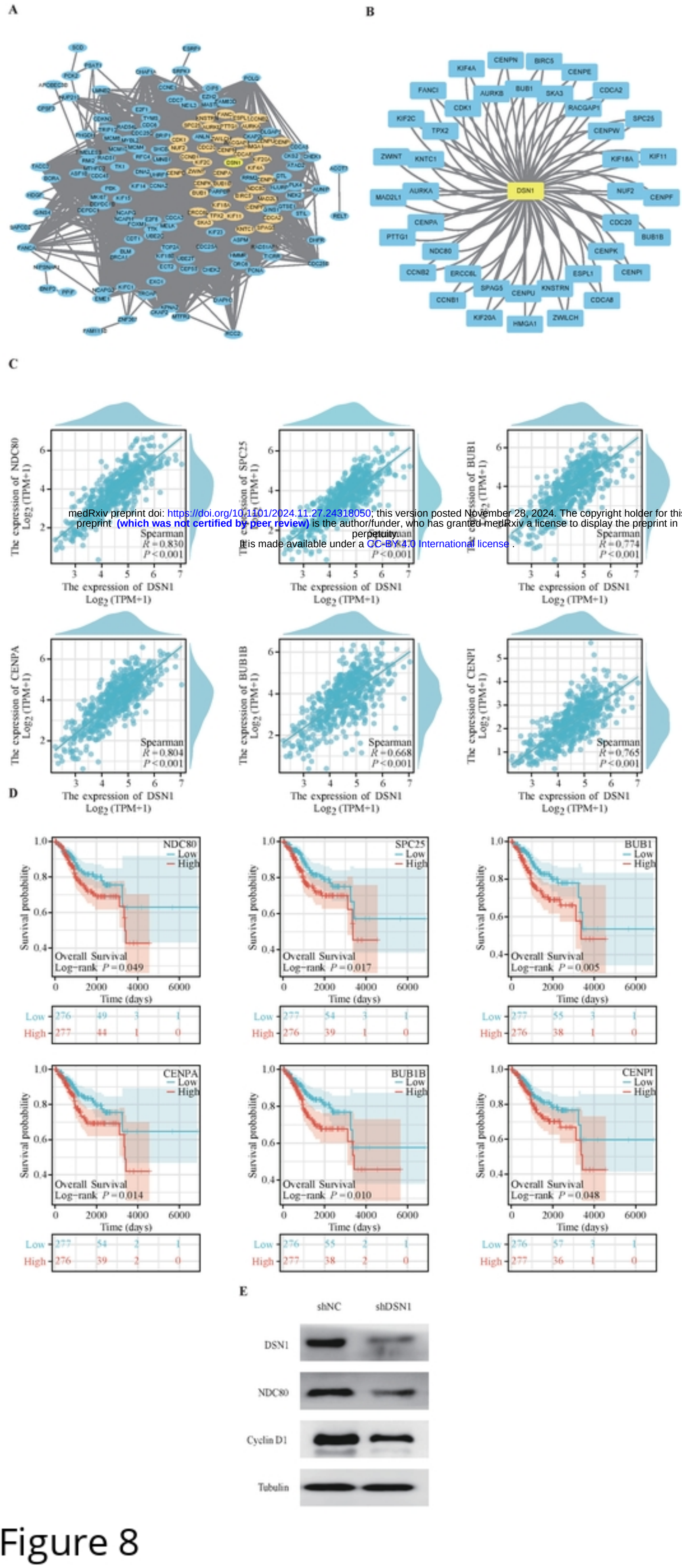


Figure 8

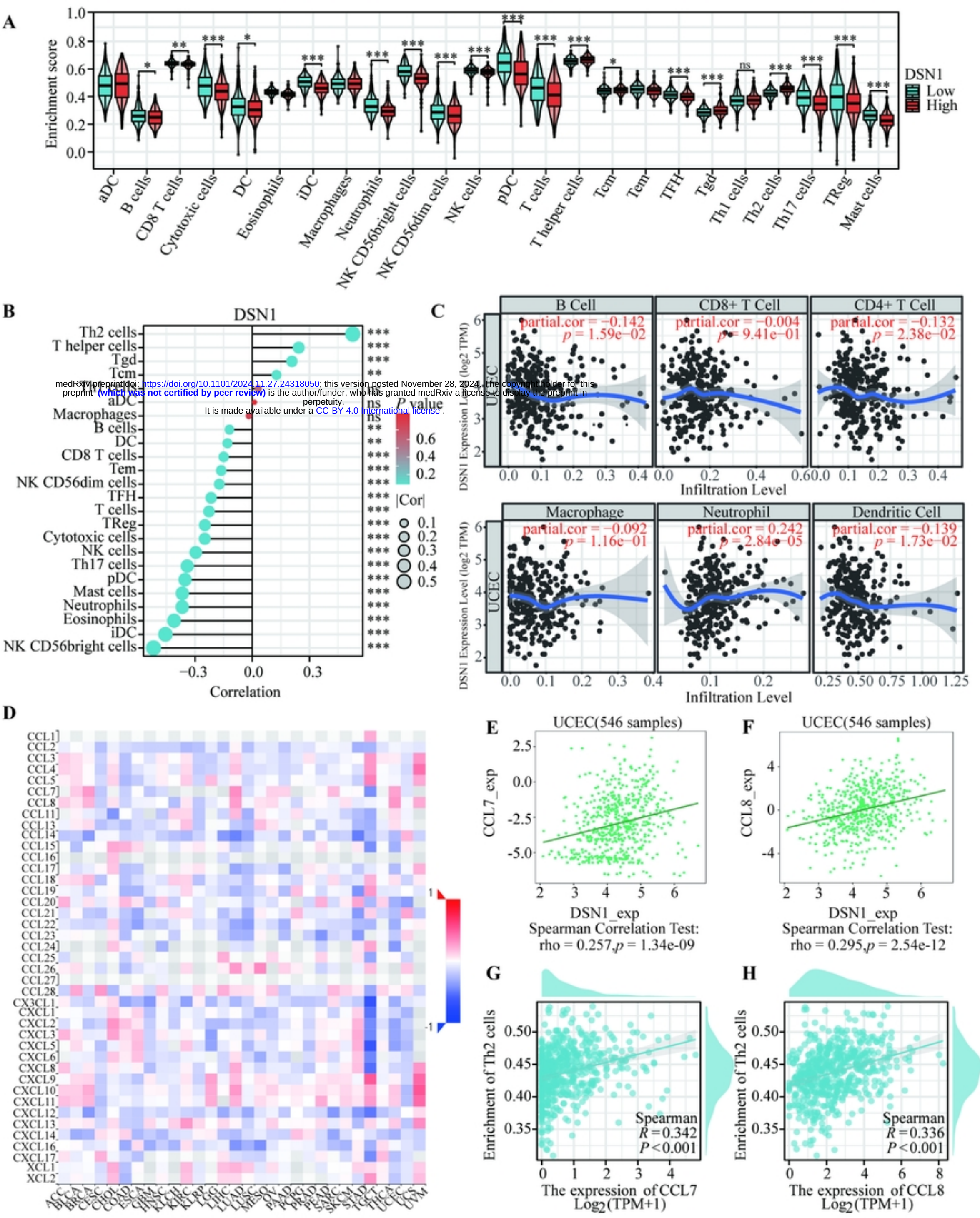


Figure 9

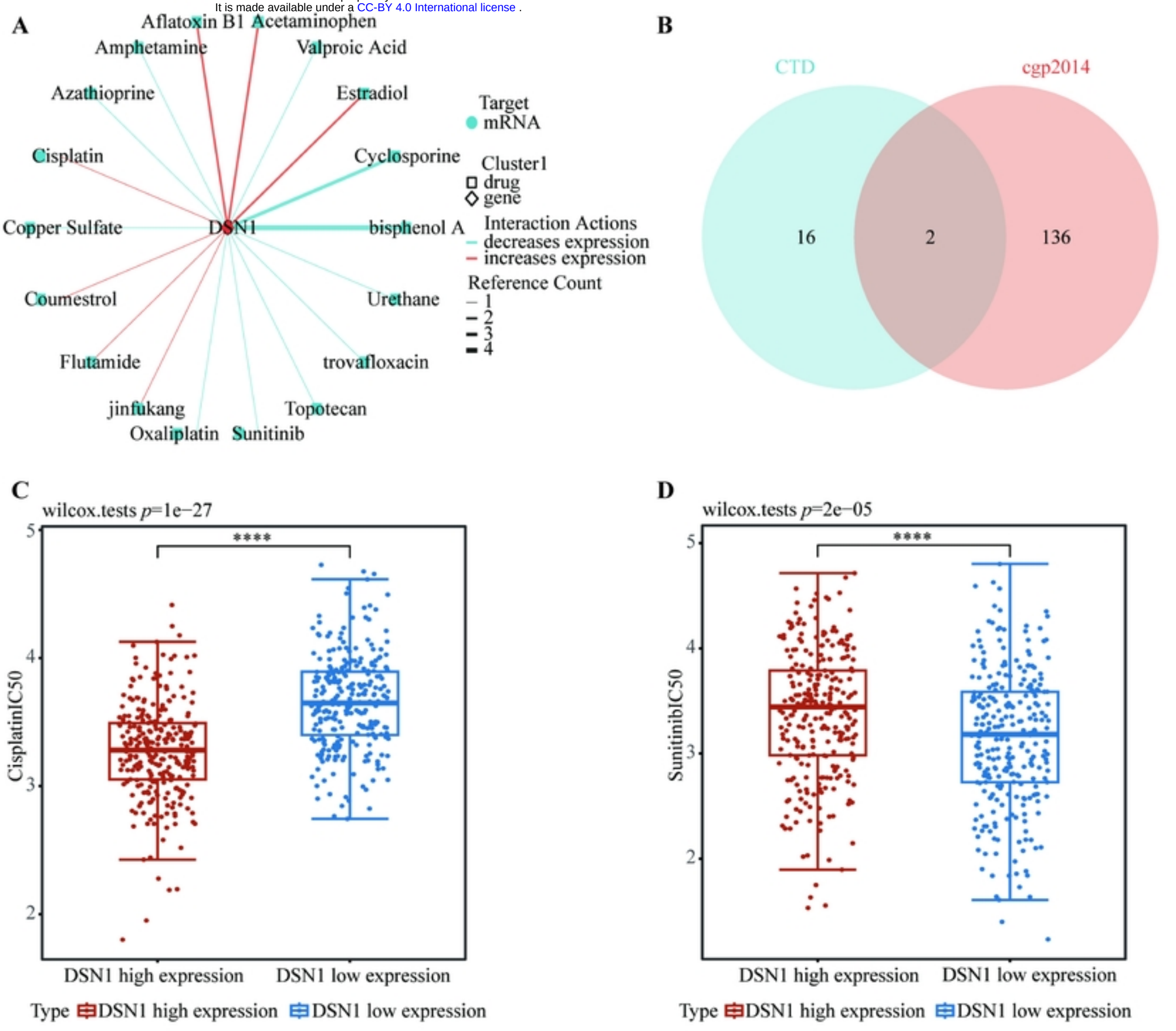


Figure 10