ORIGINAL ARTICLE

CENPF as an independent prognostic and metastasis biomarker corresponding to CD4+ memory T cells in cutaneous melanoma

Mengzhi Li[1] | Jingling Zhao[1] | Ronghua Yang[2] | Ruizhao Cai[1] | Xusheng Liu[1] | Julin Xie[1] | Bin Shu[1] | Shaohai Qi[1]

[1]Department of Burns, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[2]Department of Burn Surgery, The First People's Hospital of Foshan, Foshan, China

Correspondence
Shu Bin and Shaohai Qi, Department of Burns, The First Affiliated Hospital of Sun Yat-sen University, NO. 58, Zhongshan 2 Road, Guangzhou 510080, China.
Emails: shubin@mail.sysu.edu.cn and qishaohaigzburns@163.com

Abstract

Owing to recent advances in immunotherapies, the overall survival of patients with skin cutaneous melanoma (SKCM) has increased; however, the 5-year survival rate of metastatic patients remains poor. Skin cutaneous melanoma—upregulated genes were screened via analysis of differentially expressed genes (GSE3189 and GSE46517), and metastasis-related oncogenes were identified via weighted gene coexpression network analysis of the GSE46517 dataset. As confirmed by the Tumor Immune Estimation Resource, we found highly expressed centromere protein F (CENPF) in SKCM and its metastases. Immunostaining of human melanoma tissues demonstrated high CENPF expression. According to the Kaplan-Meier survival curve log-rank test, receiver-operating characteristic curve, and univariate and multivariate analyses, the Cancer Genome Atlas (TCGA) database suggested CENPF be a typical independent predictor of SKCM. The CIBERSORT algorithm classified the types of the immune cells from GSE46517 and showed higher proportion of CD4+ memory-activated T cells in metastatic melanoma. Single-sample gene set enrichment analysis of TCGA data confirmed the correlation between CENPF and activated CD4+ T cells. Centromere protein F was positively correlated with tumor mutational burden and CD4+ memory T cell markers (interleukin [IL]-23A, CD28, and CD62L), negatively associated with memory T cell maintenance factors (IL-7 and IL-15) by correlation analysis. Moreover, immunofluorescence showed high coexpression of CENPF and IL23A, CD4 in melanoma. Upregulated CENPF might lead to premature depletion of CD4+ memory T cells and immunosuppression. Nomogram indicated CENPF clinical predictive value for 1-, 3-, 5-, and 7-year melanoma overall survival. Therefore, CENPF plays a vital role

Shu and Qi contributed equally to this work.

This work was done in Guangzhou, Guangdong, China.

in the progression and metastasis of melanoma and can be an effective therapeutic target.

KEYWORDS
CD4+ activated T cells, CD4+ memory-activated T cells, CENPF, cutaneous melanoma, metastasis, nomogram, TMB, tumor immunology

1 | INTRODUCTION

Skin cutaneous melanoma (SKCM) is the most aggressive skin cancer derived from melanocytes.[1] Owing to advances in immunotherapies, the overall and progression-free survival rates, as well as the quality of life, of patients with advanced melanoma have improved.[2] However, the 5-year survival rate of patients with stage IV SKCM remains low at only 19%.[3] The prognosis for patients with melanoma remains poor due to the occurrence of metastasis, drug resistance, and recurrence. Therefore, recent efforts have focused on screening for more effective underlying immune-related oncogenic mutations in advanced SKCM to identify new biomarkers and potential drug targets and improve the accuracy of melanoma prognosis.

Cluster of differentiation 4 (CD4) memory T cells are antigen-specific CD4+ T cells that persist after the expansion, contraction, and memory phases of primary T cell response; these cells provide a lifelong, immediate, and accelerated response upon future encounters with specific antigens.[4-6] CD4 memory-activated T cells increase in number after a secondary antigenic stimulation and differentiate into certain CD4+ T cell subsets against specific pathogens.[7] Recent studies have shown that CD4 memory-activated T cell infiltration was significantly higher in colorectal cancer tissues than in normal tissues and was more elevated in stages T1-2 than in stages T3-4.[8] Additionally, intratumoral CD4+ T cell density has been proven to be a reliable prognostic indicator in colorectal cancer patients.[9] It can be an adverse prognostic factor in many cancers, such as kidney, prostate, lung, and breast cancers.[10-13]

Centromere protein F (CENPF) is a ~350-kDa protein that encodes the centromere-kinetochore complex–associated protein, which assembles onto kinetochores during the late G2 phase of interphase.[14,15] It is associated with the progression of various human tumors, such as hepatocellular carcinoma (HCC),[16,17] breast cancer,[18] non-Hodgkin's lymphoma,[19] and lung adenocarcinoma,[20] and with poor prognosis. CENPF was shown to promote the proliferation and migration of HCC cells[16] and accelerate bone metastasis of breast cancer.[21] Additionally, CENPF participates in CD4+ T cell–related cell cycle regulation in HCC, which is regulated by the E2F1 feedback-interactive BRCA1 pathway.[22] CENPF was significantly associated with the count of CD4+ T cells in HCC.[1723] However, the immunomodulatory function of CENPF in SKCM and its expression in clinical melanoma specimens have not been investigated.

In this study, we confirmed that CENPF is highly expressed in melanoma and is a valuable predictor of SKCM prognosis and metastasis based on multiple online databases, such as the Gene Expression

Omnibus (GEO) database, The Cancer Genome Atlas (TCGA), and the Tumor Immune Estimation Resource (TIMER). According to immunofluorescence staining of clinical melanoma specimens, we found that CENPF was highly coexpressed with the CD4 and CD4+ memory T cell marker IL23A. Moreover, we constructed a valid clinical nomogram associated with CENPF expression that demonstrated a notable estimating ability. Our results show CENPF as a feasible immunotherapeutic strategy for patients with SKCM. The detailed workflow is shown in Figure 1.

2 | MATERIALS AND METHODS

2.1 | Differentially expressed gene (DEG) analysis

The microarray datasets GSE3189 and GSE46517 were downloaded from the GEO database and normalized via log2 transformation for analysis. The inclusion criteria of the GSE3189 dataset were as follows: melanoma or nevus tissues with more than 50% melanocyte content without mixed histology.[31] The melanocyte content of the GSE46517 samples was also determined. Based on these criteria, we selected 45 primary melanoma and 18 nevus samples from the GSE3189 dataset and 12 primary melanoma and six nevus samples from the GSE46517 dataset. Data filtering, original annotation, quality control, normalization, and DEG identification were performed in Rstudio (Version 1.2.5042) using the GEOquery, hgu133a.db, limma, stringr, and reshape2 packages. We used the VennDiagram function on the ImageJ online database (http://www.ehbio.com/ImageGP/) to identify DEGs.

2.2 | Identification of the hub genes

The online database STRING (https://string-db.org/) was used to generate a protein-protein interaction (PPI) network of DEGs.[43] Cytoscape (version 3.7.2) software was used to construct the molecular interaction networks, with a combined score of >0.4.[44] The MCODE clustering algorithm was used to identify highly interconnected regions.[45] The following criteria were used to define the regions: degree cutoff = 2, node score cutoff = 0.2, max depth = 100, and K-score = 2. To screen for specific functional genes, we assessed the overlap between the top 10 genes identified by four different methods (degree, edge-percolated component, maximal clique centrality, and maximum neighborhood component) based on the Cytoscape plug-in cytoHubba.[46]
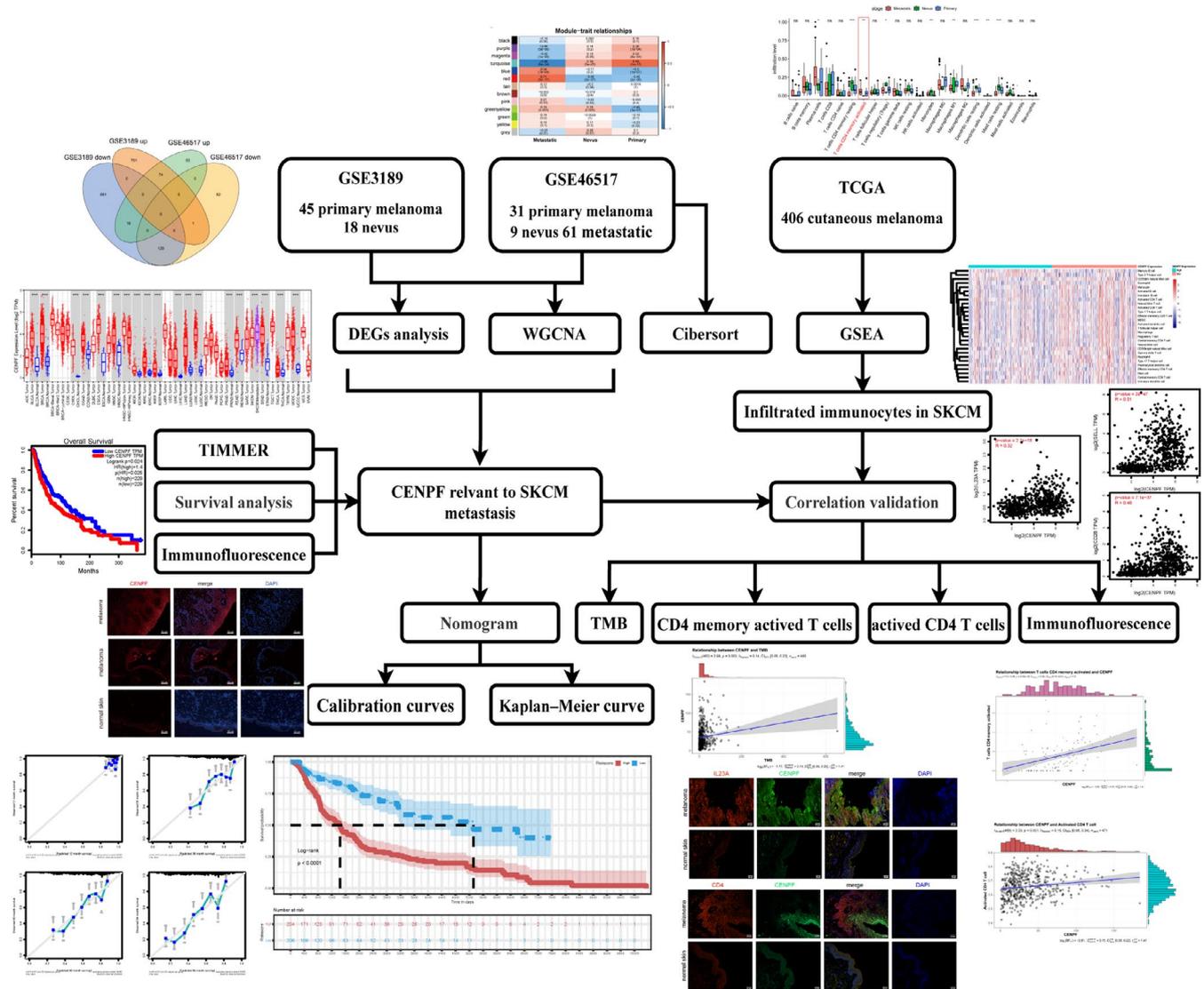
FIGURE 1 The detailed workflow of this study

Gene expression levels and survival rates were assessed using the boxplot function and survival analysis function according to Gene Expression Profiling Interactive Analysis (GEPIA).[47] The hub genes for subsequent analyses were defined based on the log-rank test results of the Kaplan-Meier (KM) estimator ($p < 0.05$).

2.3 | Ethical statement and immunofluorescence

This study was approved by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-sen University, with the patients' written informed consent. Six normal skin tissue samples and six melanoma tissue samples from patients who underwent surgical resection were collected. Immunofluorescence was used to assess the distribution and localization of CENPF in melanoma and normal skin tissues. Human melanoma formalin-fixed paraffin-embedded (FFPE) tissues were used for immunofluorescence staining. The representative FFPE sections were prepared to make epitopes available for antibody binding via dehydration, deparaffinization, and antigen

retrieval, and then, endogenous peroxidases were quenched before antibody staining. The sections were fixed in 4% paraformaldehyde for 30 minutes, washed three times with PBS, and incubated with 0.01% (v/v) Triton X-100 in 5% bovine serum albumin at 37℃ for 30 minutes. Slides were incubated with anti-rabbit CENPF (1:100, Abcam, ab224813) for 2 hours at room temperature, washed three times with PBS, and incubated with Alexa Fluor 555–conjugated goat anti-rabbit secondary antibodies (1:500, Beyotime Institute of Biotechnology) at room temperature for 1 hour in the dark. All sections were counterstained with 4,6-diamidino-2-phenylindole (DAPI) (Beyotime Institute of Biotechnology) for 5 minutes. Images were captured using an epifluorescence microscope (BX63, Olympus).

2.4 | Fluorescent multiplex immunohistochemical analysis

We also performed fluorescent multiplex immunohistochemical staining to detect the expression of CENPF and CD4 in melanoma

tissues. The preparation of multiplex immunohistochemical sections before serial staining was the same as that used for immunofluorescence. The primary antibody for CENPF (1:100, Abcam, anti-rabbit ab224813) was incubated overnight at 4°C and then incubated with the SignalStain® Boost Detection Reagent (HRP rabbit, #8114) at room temperature for 60 minutes. The antibody solution was removed, and the sections were washed with Tris-buffered saline with 0.1% Tween (TBST) three times for 5 minutes each. All the subsequent steps were performed in the dark. One hundred microliters of Opal 520 Reagent (1:100, Akoya Biosciences) were added to each slide and incubated for 10 minutes at room temperature in a humidified chamber; then, the slides were washed with TBST three times for 5 minutes per wash.

Subsequently, the antigen unmasking step was repeated for stripping. Using a microwave oven, the slides were boiled in 10 Mm sodium citrate buffer (Ph 6.0) at a sub-boiling temperature for 10 minutes and then cooled to room temperature for 30 minutes. CD4 antibodies (1:100, Abcam, anti-rabbit ab183685) and IL23A antibodies (1:100, Bioss antibodies, anti-rabbit bs-18146R) were incubated overnight at 4 °C and then incubated with SignalStain® Boost Detection Reagent (HRP rabbit, #8114). The sections were washed, and 100 μl Opal 570 Reagent (1:100, Akoya Biosciences) was added, followed by three TBST washes. All slides were counterstained with DAPI for 5 minutes and captured using an epifluorescence microscope (BX63, Olympus).

2.5 | Weighted gene coexpression network analysis (WGCNA)

To explore the relationship between metastasis, primary melanoma, and nevus, we utilized the WGCNA algorithm using the R package[48,49] to construct a coexpression network based on the GSE46517 dataset. After the cluster analysis, we created a scale-free network. Module eigengenes were used to estimate the correlation between modules and clinical features. The gene significance (GS) represents the mediator *p*-value for each gene to define the correlation between gene expression and clinical features. We selected the most significant module related to metastasis according to *p* < 0.01 and high GS value. CENPF was included in this module.

2.6 | CIBERSORT

To explore the relationship between CENPF and immune cell subtypes, we utilized the CIBERSORT package to estimate the proportions of 22 immune cell subtypes[50] and set the perm at 1000 and the threshold at *p* < 0.05. The Mann-Whitney U test was used to determine differences between immune cell subtypes in metastasis melanoma, primary melanoma, and nevus samples. Spearman correlation was used to measure the linear correlation between CENPF expression and immune cell subtypes.

2.7 | Single-sample gene set enrichment analysis (ssGSEA)

The ssGSEA algorithm of the R package "Gene Set Variation Analysis (GSVA)" was used to explore the immune cell types' infiltration level and immune-related genes.[51] Pearson's correlation coefficient was used to calculate the correlation between CENPF and immune cell types.

2.8 | TMB estimation and prognostic analysis

We downloaded the masked somatic mutation data from TCGA online database. Then, the "maftools" R package was used to analyze the somatic mutation data and transcriptome profiles.[52] Pearson correlation analysis was used to determine the relationship between CENPF expression and TMB.

2.9 | TIMER database analysis

The TIMER database website has comprehensive tools for the analysis of tumor-infiltrating immune cells across diverse carcinomas.[24] We used the Diff Exp module to analyze the expression of target genes in SKCM metastasis and SKCM tumors. Box plots show gene expression levels between tumor and adjacent normal tissues for CENPF across all TCGA tumors. The number of asterisks annotates the statistical significance computed by the Wilcoxon test.

2.10 | Construction and verification of a nomogram

Based on univariate and multivariate analyses, we used CENPF level, recurrence, radiation therapy, and TNM stage to construct the nomogram to predict the 1-, 3-, 5-, 7-year survival of patients with melanoma. This nomogram was built using "rms" and "survival" R packages. The survival rate was assessed using the KM survival curve log-rank test.

2.11 | Statistical analyses

The UCSC Xena platform (http://xena.ucsc.edu) was used to download data regarding CENPF and phenotypic variables from TCGA for melanoma (SKCM). The CENPF gene dataset was divided into high- and low-expression groups based on the median expression levels. General data parameters were assessed using descriptive statistics. The chi-square test was used to determine the clinical features of CENPF expression, which were significantly different between the representative phenotypes. Univariate Cox regression was used to analyze the associations of CENPF level, radiotherapy, recurrence, and TNM stage on the survival of patients with SKCM. Multivariate analysis of the variables that showed

significant associations was used to identify the variables with an independent impact on survival. The area under the receiver-operating characteristic (ROC) curve (AUC) of each independently associated factor was calculated to determine the best predictive power for each factor. MedCalc statistical software (MedCalc Software Ltd., v15.8; https://www.medcalc.org; 2015) was used to generate ROC curves.

A Cox regression model was established to determine independent prognostic factors affecting OS. The CENPF nomogram was built based on multivariable analysis. Factors such as CENPF level, recurrence, radiation therapy, and TNM stage were used to calculate the nomogram score. We then used calibration curves to test the performance of the CENPF nomogram. Finally, the KM survival curve log-rank test was used to estimate the survival probabilities of the low- and high-risk groups, as defined by the median of the nomogram scores.

3 | RESULTS

3.1 | Identification of DEGs in SKCM

Differentially expressed genes analysis was used to identify up-regulated and downregulated genes. As shown in the Venn diagram (Figure 2A), 1842 (836 upregulated and 1006 downregulated) and 355 (143 upregulated and 212 downregulated) DEGs were identified using the GSE3189 and GSE46517 datasets, respectively. A total of 74 upregulated and 129 downregulated genes were significantly differentially expressed. The overlapping upregulated and downregulated DEGs of the two gene sets are shown in Tables S1 and S2. Volcano plots illustrating the DEGs between SKCM and nevus control tissues in GSE3189 and GSE46517 datasets are presented as volcano plots in Figure 2B,C. The top 30 DEGs between tumor and normal tissues are presented as heatmaps in Figure 2D,E.

3.2 | Protein-protein interaction network construction and critical module analysis

A PPI network was constructed to determine the interactions between 203 identified genes using the STRING online database and Cytoscape software. The network comprised 149 nodes and 297 edges, and the top 100 nodes are illustrated in Figure 2F. Significantly upregulated and downregulated genes were identified based on the shape of the nodes. Based on the degree of importance of the Cytoscape MCODE plug-in, we found that CENPF belongs to the top significant module, which contains 36 edges and nine nodes, namely CENPF, SPAG5, cyclin B1 (CCNB1), TYMS, HJURP, ASPM, kinesin family member 20A (KIF20A), PBK, and CEP55. The intersection of the top 10 genes identified by four different methods using the Cytoscape plug-in cytoHubba was used to select four hub genes: CCNB1, cyclin-dependent kinase 2 (CDK2), CENPF, and KIF20A (Table 1).

The TIMER database website has comprehensive tools for the analysis of tumor-infiltrating immune cells across diverse carcinomas.[24] Additionally, according to the SKCM metastasis and tumor data in the TIMER database, CENPF expression was significantly higher in SKCM metastasis tumors than in SKCM primary tumors (Figure 3A).

To verify the expression of these genes using the GEO database, an ROC curve analysis was performed to evaluate the corresponding sensitivity and specificity of SKCM diagnosis using the data from the GSE3189 dataset. As shown in Figure 3B, the AUC value of CENPF was 0.954, demonstrating a significant association with the diagnosis. Moreover, the boxplot function of GEPIA showed that the transcriptional level of CENPF increased in 461 SKCM tissues compared with 558 normal tissues based on TCGA and Genotype-Tissue Expression (GTEx) datasets (Figure 3C). We confirmed that CENPF expression was higher in SKCM tissues than in normal ones via immunofluorescence analysis in six clinical melanoma and three normal skin samples (Figure 3D). These results suggest that CENPF has high sensitivity and specificity for SKCM diagnosis.

3.3 | High expression of CENPF is significantly correlated with poor survival

To investigate the impact of high CENPF expression on the prognosis of patients with SKCM, the OS analysis function in GEPIA was used, which showed that CENPF was significantly associated with survival time of patients with SKCM (Figure 3E). The baseline clinico-pathological characteristics of 410 paired patients with SKCM from TCGA database, including sex, age, recurrence, TNM stage, radiation therapy, and event, are detailed in Table 2.

Univariate analyses of various prognostic parameters of patients with SKCM demonstrated that high expression of CENPF ($p = 0.017$, adjusted hazard ratio [HR] = 1.204), recurrence ($p < 0.001$, HR = 3.119), TNM stage ($p < 0.001$), and radiation therapy ($p = 0.003$, HR = 0.429) were significantly associated with the risk of mortality (Table 3). CENPF expression ($p = 0.070$), recurrence ($p < 0.001$), TNM stage ($p < 0.001$), and radiation therapy ($p = 0.019$, HR=0.504) remained independent predictors in a multivariate model that included CENPF level, recurrence, TNM stage, and radiation therapy (Table 3). As expected, the AUC values of CENPF expression (AUC: 0.599; 95% CI: 0.550-0.647) corresponded to a better prognostic value than those of metastasis (AUC: 0.511; 95% CI, 0.461-0.560; $p = 0.028$) and recurrence (AUC: 0.502, 95% CI: 0.453-0.552, $p = 0.014$), indicating that CENPF has a good prognostic value. (Figure 3F).

3.4 | CENPF is significantly associated with clinical traits of metastasis

To construct the coexpression module, we performed a WGCNA, in which a total of 114 metastasis, primary melanoma, and nevus
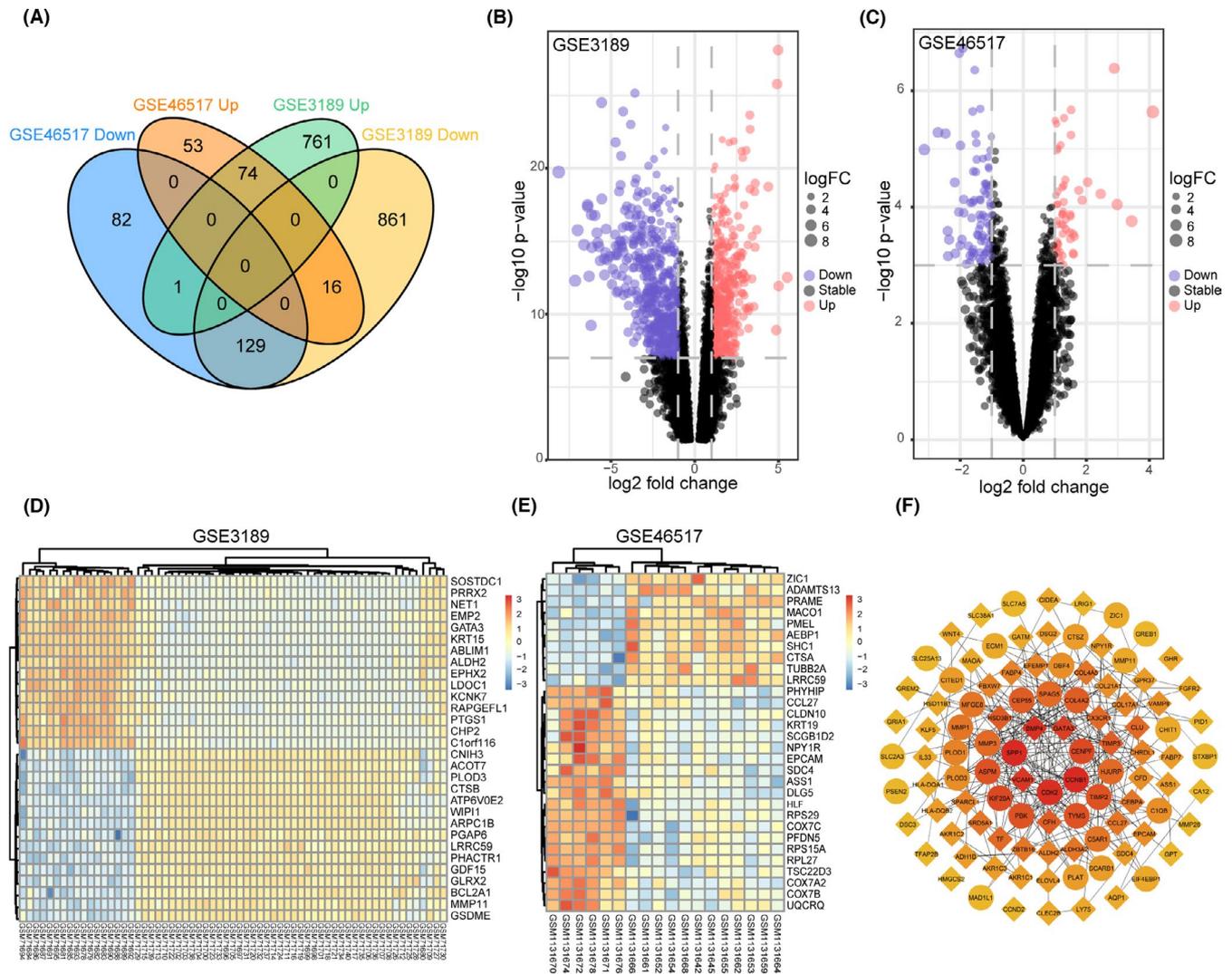
FIGURE 2 Identification of hub genes in skin cutaneous melanoma (SKCM)

TABLE 1 Hub genes identified by four ranked methods in cytoHubba

	Ranked methods in cytoHubba			
	MNC	**Degree**	**EPC**	**MCC**
Gene symbol top 10	**CCNB1**	SPP1	SPP1	**CCNB1**
	SPP1	**CCNB1**	BMP4	**CENPF**
	CDK2	BMP4	**CDK2**	ASPM
	CENPF	**CDK2**	VCAM1	PBK
	ASPM	VCAM1	**CCNB1**	**KIF20A**
	PBK	GATA3	**KIF20A**	TYMS
	TYMS	**KIF20A**	CEP55	CEP55
	KIF20A	**CENPF**	ASPM	HJURP
	TIMP2	TYMS	PBK	SPAG5
	HJURP	TIMP2	**CENPF**	**CDK2**

Note: Bold gene symbols were the overlap hub genes in top 10 by four ranked methods respectively in cytoHubba.

Abbreviations: Degree, node connect degree; EPC, edge-percolated component; MCC, maximal clique centrality; MNC, maximum neighborhood component.

samples were selected from the GSE46517 dataset, including 12,437 genes. After cluster analysis, 101 samples were selected for the subsequent analysis. The power value was 7, and the independence degree was 0.97. Therefore, 13 coexpression modules containing 6000 genes were constructed and are shown in different colors in Figure 4A. The metastasis group was strongly associated
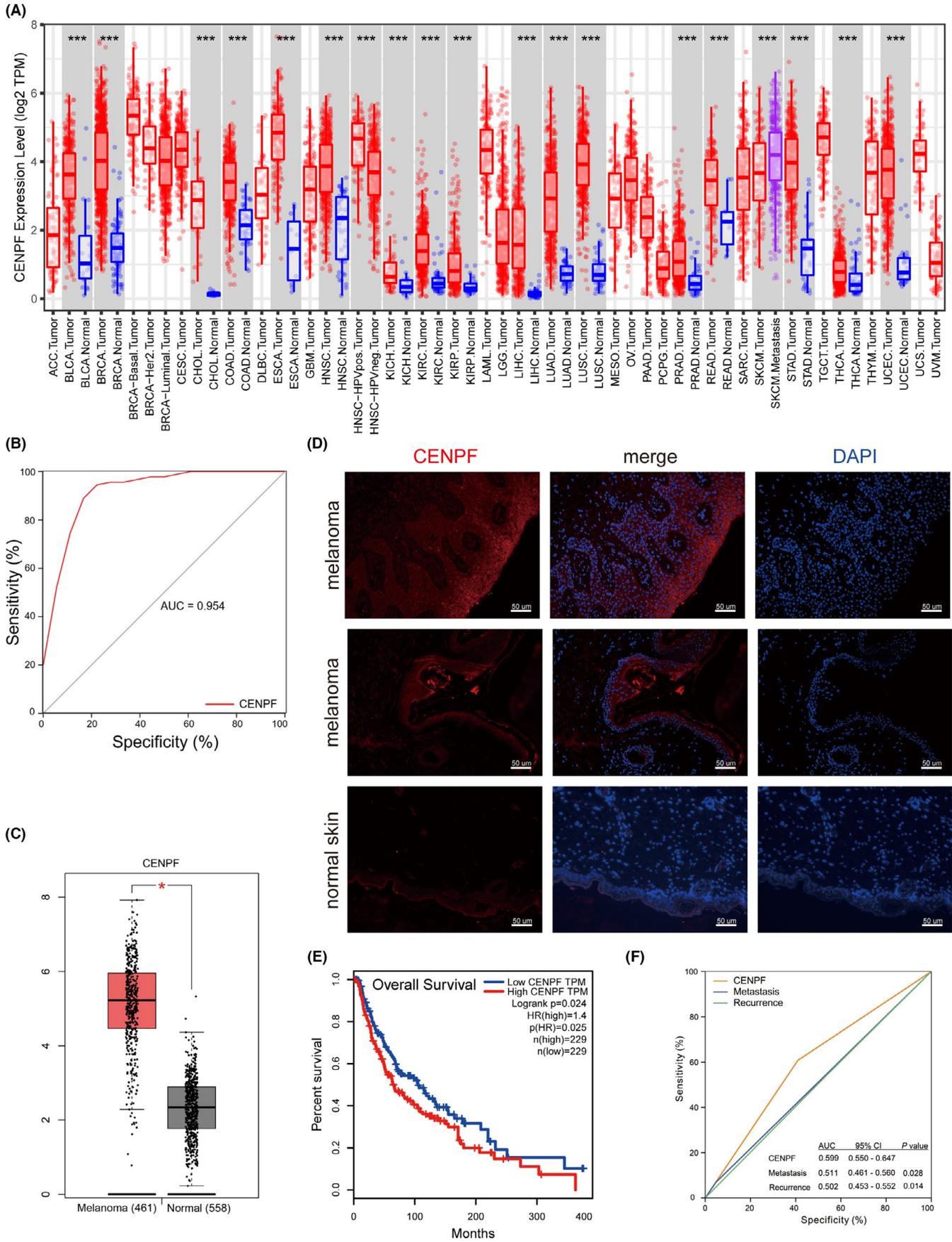
FIGURE 3 Centromere protein F (CENPF) is associated with skin cutaneous melanoma (SKCM) metastasis and poor survival

with the red module containing 287 genes ($p = $ 5e-17, $r = $ 0.71) (Figure 4B,C) (Table S3). Centromere protein F was included in this module.

3.5 | Assessment of immune cell infiltration in primary and metastatic melanoma

To further study the changes in the immune microenvironment in primary and metastatic melanoma and melanocytes, we used the CIBERSORT algorithm to classify the immune cell composition and clinical type based on the GSE46517 dataset. As shown in Figure 5A, the proportion of CD4+ memory-activated T cells and plasma cells was higher in metastatic melanoma tissues. In contrast, the proportion of CD4+ memory resting T, dendritic, and mast cells was significantly lower in metastatic and primary melanoma tissues. Next, we analyzed the correlations between CD4+ memory-activated T cells and other types of immunocytes (Figure 5B,C), which were processed and visualized using the R package "corrplot." The score and gradient of colors represent the correlation coefficient. Figure 5B shows a heatmap of 22 immune-related cells and types in the high- and low-CENPF expression groups. The correlation between activated CD4+ memory T cells and 21 immune cells is shown in Figure 5C. CD4+ memory-activated T cells were positively correlated with resting NK cells, gamma delta T cells, M2 macrophages, monocytes, neutrophils, plasma cells, activated mast cells, and eosinophils. Moreover,

TABLE 2 Baseline characteristics of skin cutaneous melanoma (SKCM) patients

Variable	N = 410 (%)
Sex	
Male	256 (62.4)
Female	154 (37.6)
Age	
<60	211 (51.5)
≥60	199 (48.5)
Recurrence	
Yes	224 (54.6)
No	186 (45.4)
TNM stage	
I	79 (19.3)
II	140 (34.1)
III	167 (40.7)
IV	24 (5.9)
Radiation therapy	
Yes	41 (10.0)
No	369 (90.0)
Event	
Alive	219 (53.4)
Dead	191 (46.6)

Spearman correlation was used to measure the linear correlation between CENPF expression and CD4+ memory-activated T cells, as shown in Figure 5D ($p = $ 8.58e-05, $r = $ 0.36).

To investigate the reliability of the CENPF-related immune cell infiltration clusters, we further analyzed the relationship between CENPF expression and immune cell infiltration in patients with SKCM based on TCGA database using the ssGSEA of the R "GSVA" package. Pearson correlation analysis demonstrated that CENPF expression was associated with activated CD4+ T cells ($p = $ 0.001, $r = $ 0.15) (Figures 5E and 6A).

Additionally, TMB (i.e., the number of somatic mutations per million bases) is widely recognized as a key driver of immunogenic mutations, as well as a predictor of the efficacy of immune checkpoint inhibitors and patient outcomes with several malignancies, including melanoma.[25,26] Using the "maftools" R package, we analyzed the somatic mutation data and transcriptome profiles of 410 patients with SKCM from TCGA database. Pearson correlation analysis demonstrated that CENPF expression was associated with TMB ($p = $ 0.0013, $r = $ 0.14) (Figure 6B).

3.6 | Centromere protein F is positively correlated with CD4+ memory T cell–related markers and negatively correlated with CD4+ memory T cell survival regulators in melanoma

We analyzed the correlation of CENPF with CD4+ memory T cell markers via pairwise gene expression correlation analysis using SKCM tumor and not–sun-exposed skin datasets of TCGA and the GTEx data provided by GEPIA. Our results showed that CENPF expression was positively correlated with phenotypic markers of CD4+ memory T cells, such as IL-23A ($p = $ 2.7e-18, $r = $ 0.32) (Figure 6C), CD62L (SELL) ($p = $ 2e-47, $r = $ 0.51) (Figure 6D), and CD28 ($p = $ 7.1e-37, $r = $ 0.46) (Figure 6E). However, CENPF was negatively correlated with survival and generation-related regulators of memory T cells, such as IL-7 ($p = $ 1.2e-24, $r = $ −0.38) (Figure 6F) and IL-15 ($p = $ 0.0012, $r = $ −0.12) (Figure 6G). Furthermore, we performed immunofluorescence analysis and found that CENPF was highly coexpressed with IL23A (Figure 6H) and CD4 (Figure 6I) in melanoma tissues. Collectively, these findings indicate an association between CENPF and the activation and depletion of CD4+ memory T cells.

3.7 | Centromere protein F nomogram and survival curve validation

To predict the OS rates of melanoma patients at 1, 3, 5, and 7 years, we designed a Nomogram melanoma diagnosis map based on TCGA. The decision curve analysis for the predictive nomogram model containing CENPF level, recurrence, radiation therapy, and TNM stage is presented in Figure 7A. This curve showed high CENPF levels, indicating a high risk of mortality. Therefore, this model can quickly determine the estimated probability of

TABLE 3 Univariate and multivariate analyses of various prognostic parameters in patients with skin cutaneous melanoma (SKCM) Cox regression analysis

	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% CI	P	Hazard ratio	95% CI	P
CENPF	1.204	1.033-1.402	0.017	1.157	0.988-1.355	0.070
Recurrence	3.119	2.142-4.543	<0.001	3.165	2.154-4.649	<0.001
TNM stage						
I	1	—	0.001			<0.001
II	1.580	1.049-2.380	0.028	1.741	1.149-2.637	0.009
III	2.077	1.408-3.062	<0.001	2.671	1.799-3.968	<0.001
IV	2.938	1.493-5.782	0.002	3.927	1.973-7.816	<0.001
Radiation therapy	0.429	0.244-0.755	0.003	0.504	0.284-0.893	0.019

Abbreviation: CENPF, centromere protein F.
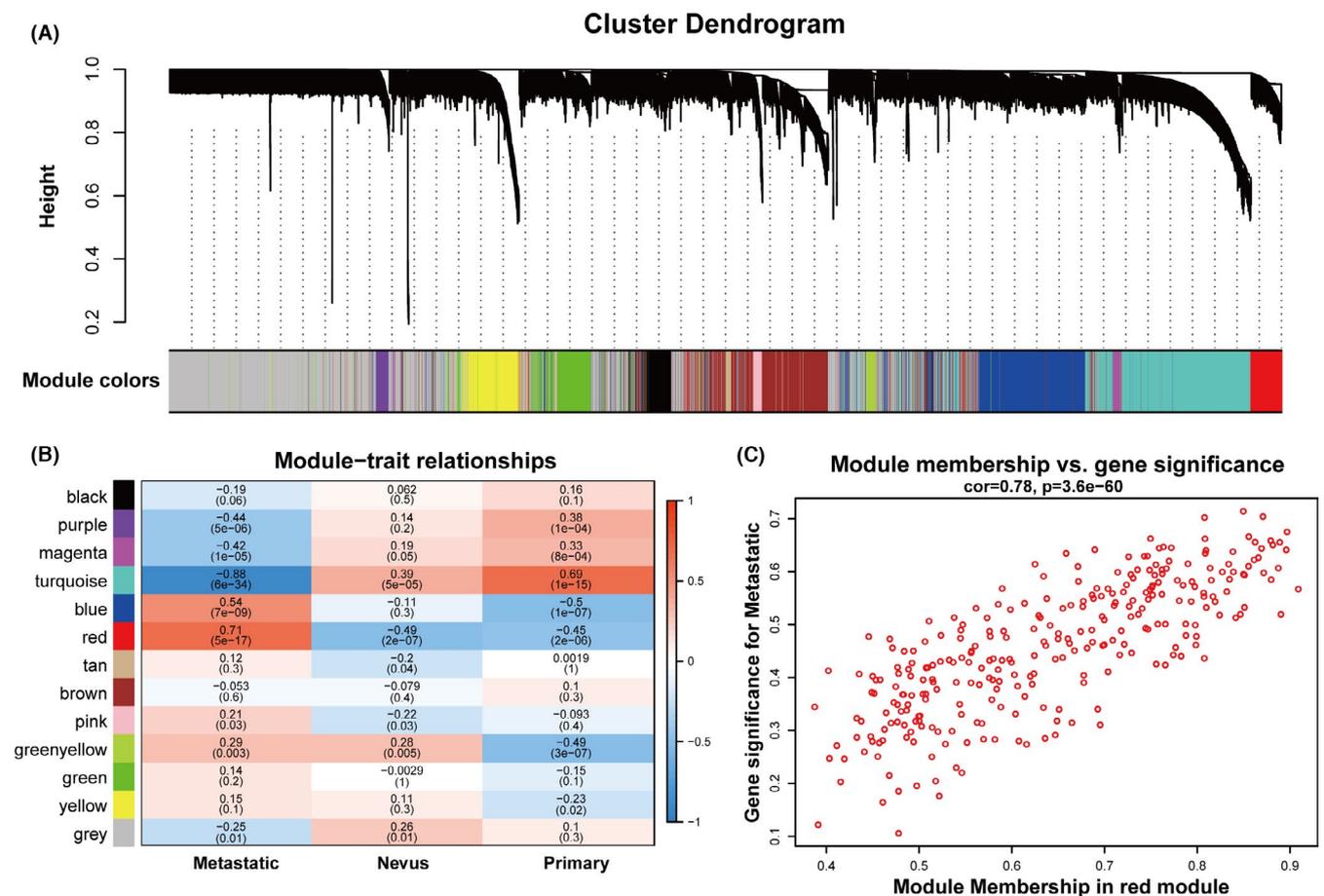
FIGURE 4 Centromere protein F (CENPF) is significantly associated with metastasis

survival by adding the total scores and finding the matching position at each time point.

To evaluate the efficiency of this predictive nomogram, we calculated the C-index of the model, which was 0.683 (with a standard error of 0.044), thereby confirming that the CENPF nomogram is moderately accurate. The 1-, 3-, 5-, and 7-year bias-corrected lines were close to the ideal curve, indicating that the CENPF nomogram

prediction results agree with the actual 1-, 3-, 5-, and 7-year overall survival rates (Figure 7B–E). Additionally, we distinguished two risk subgroups (low-risk subgroup: <234.162, high-risk subgroup: ≥234.162) using the median of nomogram scores. The KM survival curve log-rank test showed that the low-risk subgroup had better survival than the high-risk subgroup ($p < 0.001$) (Figure 7F), thereby confirming the prognostic value of the CENPF nomogram model and
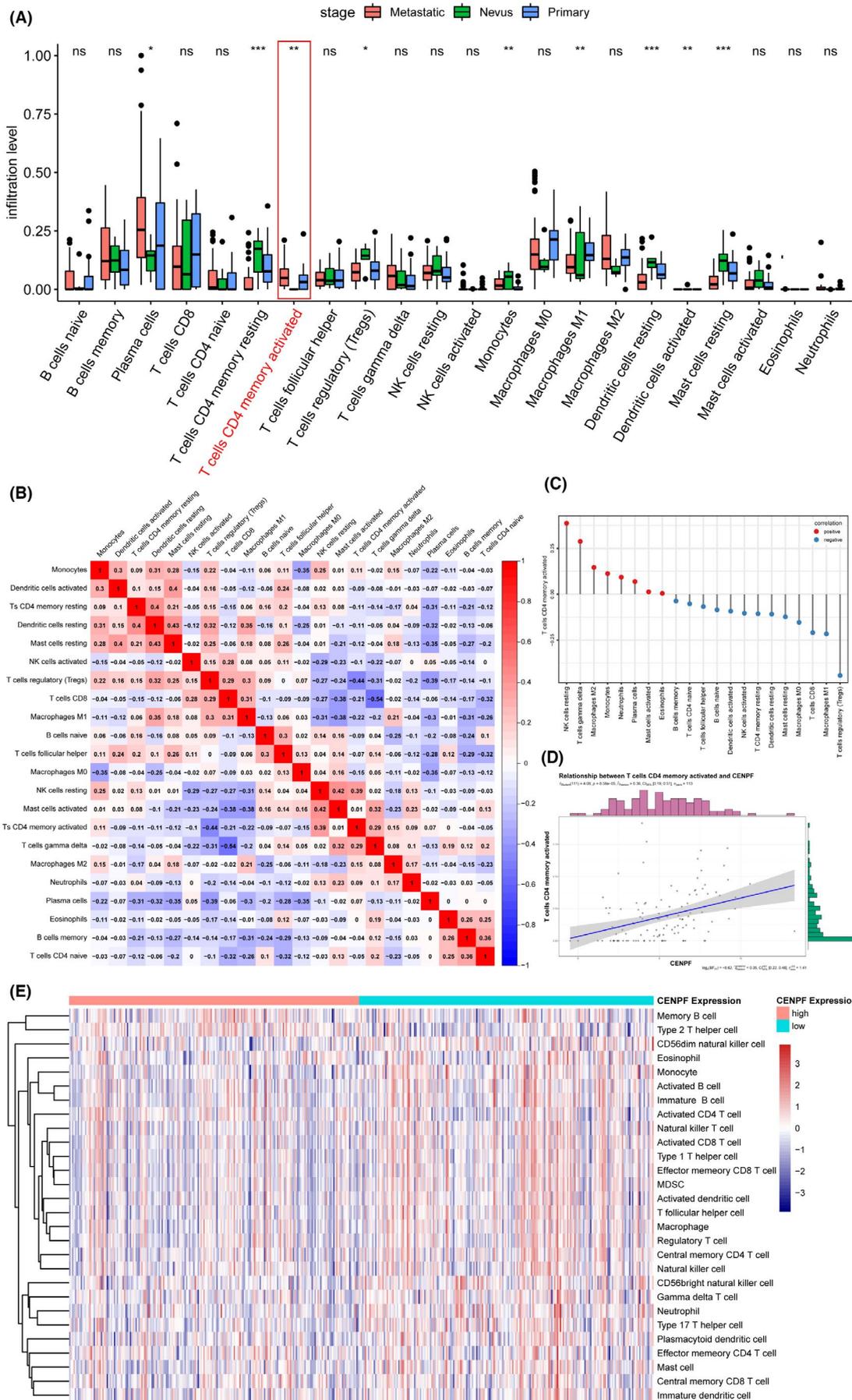
FIGURE 5 Assessment of immune cell infiltration in primary and metastatic melanoma

(A)

Relationship between CENPF and Activated CD4 T cell

$t_{Student}(469) = 3.33$, $p = 0.001$, $\hat{r}_{Pearson} = 0.15$, $CI_{95\%}$ [0.06, 0.24], $n_{pairs} = 471$

$log_e(BF_{01}) = -2.81$, $\hat{r}_{Pearson}^{posterior} = 0.15$, $CI_{95\%}^{HDI}$ [0.08, 0.22], $r_{beta}^{JZS} = 1.41$

(B)

Relationship between CENPF and TMB

$t_{Student}(463) = 2.99$, $p = 0.003$, $\hat{r}_{Pearson} = 0.14$, $CI_{95\%}$ [0.05, 0.23], $n_{pairs} = 465$

$log_e(BF_{01}) = -1.77$, $\hat{r}_{Pearson}^{posterior} = 0.14$, $CI_{95\%}^{HDI}$ [0.06, 0.20], $r_{beta}^{JZS} = 1.41$

(C) p-value = 2.7e-18, R = 0.32

(D) p-value = 2e-47, R = 0.51

(E) p-value = 7.1e-37, R = 0.46

(F) p-value = 1.2e-24, R = −0.38

(G) p-value = 0.0012, R = −0.12

(H) IL23A · CENPF · merge · DAPI
melanoma / normal skin

(I) CD4 · CENPF · merge · DAPI
melanoma / normal skin

FIGURE 6 Centromere protein F (CENPF) is associated with cluster of differentiation 4 (CD4)+ memory T cell markers and negatively associated with CD4+ memory T cell survival regulators in melanoma

(A)

(B) **(C)** **(D)** **(E)**

(F)
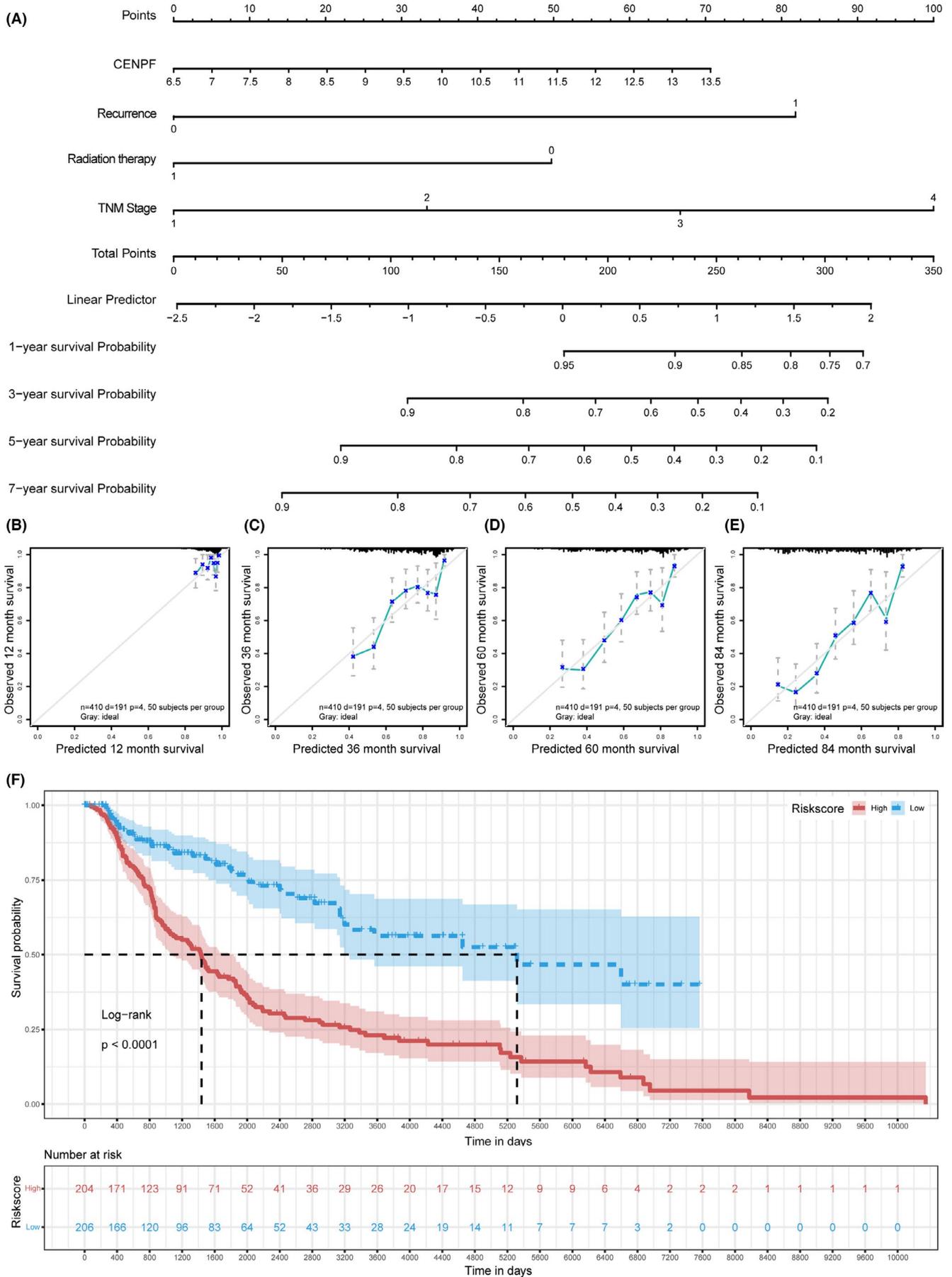
FIGURE 7 Centromere protein F (CENPF) nomogram and survival curve validation by nomogram scores

the high predictive value of CENPF mutations for survival responses in patients with SKCM.

4 | DISCUSSION

Melanomas often arise from distinctive precursor lesions, such as melanocytic nevi, intermediate lesions, or melanoma in situ, and their progression can be analyzed.[27] The malignant transformation of melanocytes into metastatic melanoma is a multifactorial process caused by complex interactions between exogenous and endogenous triggers, as well as tumor-intrinsic and immune-related factors.[28] Previous studies have identified genes that are differentially expressed in benign and malignant lesions[29] and genes that might be of prognostic value.[30] In melanoma, many genes with altered expression were found to be highly associated with the development of the disease. The gene expression changes are associated explicitly with malignant melanocyte transformation, contributing to identifying the overexpressed novel gene in melanoma precisely.[31] In this study, we systematically analyzed GEO and TCGA SKCM databases, adopted various prediction algorithms, and found that CENPF was highly expressed in SKCM and significantly affected the metastasis and survival of patients with SKCM. Additionally, CIBERSORT, ssGSEA, and immunofluorescence analyses showed a potential association between CENPF and CD4+ memory-activated T cell infiltration, supporting the immunotherapeutic value of CENPF in patients with SKCM.

Centromere protein F, a kinetochore-associated protein that controls the mitotic progression and regulates the cell cycle, is involved in the recruitment of the spindle checkpoint proteins to maintain checkpoint response.[15,32,33] Centromere protein F promotes bone metastasis of breast cancer by activating the PI3K-AKT-mTORC1 signaling pathway, increasing the parathyroid hormone–related protein.[21] Additionally, overexpression of the COUP-TFII-FOXM1-CENPF axis contributes to the metastasis and drug resistance of prostate cancer by promoting cell migration.[34] Here, we showed that high CENPF expression is strongly associated with melanoma metastasis. We identified the prognostic value of CENPF using multiple methods, including immunofluorescence, survival analysis, univariate and multivariate Cox proportional hazard models, and ROC curves. CENPF was positively correlated with TMB, a predictive biomarker and a predictor of immunotherapy accuracy.[25] Weighted gene coexpression network analysis showed a significant correlation between CENPF and metastasis. Additionally, our CIBERSORT results showed that CD4+ memory-activated T cells promote melanoma metastasis. Centromere protein F demonstrated a positive association with activated CD4+ T cells.

The function of naïve and memory T cell facets in immunological memory was first reported in 1999.[35] Memory T cells can optimize tumor immunotherapy strategies.[4] CENPs, such as CENPA and CENPB, have been proven to be specific immune autoantigens for scleroderma; therefore, CENPF is possibly associated with immune

regulation in patients with melanoma.[36] It is worth investigating whether CENPF expression affects immune cell infiltration and specific cell types.

The phenotype of memory T cells is affected by their microenvironment; however, the mechanism of the activation of resident memory CD4+ T cells remains unclear.[37] A study showed that CD4+ memory-activated T cells infiltrated the inflamed region in a patient with brain metastatic melanoma who suffered from pembrolizumab-induced encephalitis.[38] Additionally, in-depth analysis of the activation and regulation of memory T cells holds notable potential for improving vaccines against infections, cancer immunotherapies, and the treatment of autoimmunity.[39] Hence, we performed a pairwise gene expression correlation analysis using the SKCM tumor and not–sun-exposed skin datasets of TCGA and the GTEx data provided by GEPIA. We found CENPF to be positively correlated with CD4+ memory T cell phenotypic markers,[4] such as IL-23A,[8] CD28, and CD62L (SELL),[6] and negatively correlated with memory T cell survival and regulators, such as IL-7 and IL-15.[4] In addition, we performed immunofluorescence and found that CENPF was highly coexpressed with IL23A in melanoma tissues. Given these findings, we hypothesized that CENPF might be associated with CD4+ memory T cell activation. Moreover, an increase in CENPF expression level inhibited the maintenance and homeostasis of CD4+ memory T cells, leading to immunosuppression. We also confirmed the high expression of CENPF and CD4 in melanoma tissues via a fluorescent multiplex immunofluorescence analysis. This immunofluorescence technique avoids false positives caused by the pigmentation of melanoma tissues and provides a more reliable confirmation of target gene expression. Collectively, these results confirmed that CENPF promotes immune escape in melanoma.

Furthermore, clinical nomograms were shown to have an accurate prognostic predictive ability in cancer prognosis.[40,41] Recently, nomograms have made significant progress in the development of prognostic models in several diseases. The accuracy of the models can be verified by a calibration curve; hence, nomograms are of high clinical value.[42] In this study, using univariate and multivariate analyses, we confirmed that CENPF level, recurrence, radiation therapy, and TNM stage were significant independent factors affecting the OS. The predictive value of the CENPF nomogram was consistent with the calibration curves of 1-, 3-, 5-, and 7-year OS rates. The KM survival curve log-rank test showed that the low-risk subgroup had a better survival rate ($p < 0.001$), thereby confirming the accuracy of the CENPF nomogram survival assessment in patients with SKCM.

In conclusion, we confirmed the role of CENPF in the carcinogenesis of SKCM with a risk of distant metastases. Centromere protein F had a clinical predictive value for 1-, 3-, 5-, and 7-year OS. Notably, CENPF was associated with CD4+ memory T cell markers in melanoma; upregulation of CENPF expression was negatively correlated with CD4+ memory T cell survival regulators, leading to premature depletion of CD4+ memory T cells and immunosuppression. Thus, CENPF is a potential clinical prognostic marker for skin melanoma metastasis and may be an effective therapeutic target. Further investigation is required to validate these findings.

ORCID

Shaohai Qi https://orcid.org/0000-0001-5962-4203

REFERENCES

1. Liu Q, Das M, Liu Y, Huang L. Targeted drug delivery to melanoma. *Adv Drug Deliv Rev*. 2018;127:208-221.
2. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*. 2019;381(16):1535-1546.
3. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA: A Cancer J Clin* 2019;69(5):363-385.
4. Liu Q, Sun Z, Chen L. Memory T cells: strategies for optimizing tumor immunotherapy. *Protein Cell*. 2020;11(8):549-564.
5. Lanzavecchia A, Sallusto F. Progressive differentiation and selection of the fittest in the immune response. *Nat Rev Immunol*. 2002;2(12):982-987.
6. Lanzavecchia A, Sallusto F. Dynamics of T lymphocyte responses: intermediates, effectors, and memory cells. *Science*. 2000;290(5489):92-97.
7. Jenkins MK, Khoruts A, Ingulli E, et al. In vivo activation of antigen-specific CD4 T cells. *Annu Rev Immunol*. 2001;19:23-45.
8. Ge P, Wang W, Li L, et al. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed Pharmacoth*. 2019;118:109228.
9. Kuwahara T, Hazama S, Suzuki N, et al. Correction: Intratumoural-infiltrating CD4 + and FOXP3 + T cells as strong positive predictive markers for the prognosis of resectable colorectal cancer. *Br J Cancer*. 2019;121(11):983-984.
10. Bromwich EJ, McArdle PA, Canna K, et al. The relationship between T-lymphocyte infiltration, stage, tumour grade and survival in patients undergoing curative surgery for renal cell cancer. *Br J Cancer*. 2003;89(10):1906-1908.
11. Kinoshita T, Muramatsu R, Fujita T, et al. Prognostic value of tumor-infiltrating lymphocytes differs depending on histological type and smoking habit in completely resected non-small-cell lung cancer. *Ann Oncol*. 2016;27(11):2117-2123.
12. McArdle PA, Canna K, McMillan DC, McNicol AM, Campbell R, Underwood MA. The relationship between T-lymphocyte subset infiltration and survival in patients with prostate cancer. *Br J Cancer*. 2004;91(3):541-543.
13. Droeser R, Zlobec I, Kilic E, et al. Differential pattern and prognostic significance of CD4+, FOXP3+ and IL-17+ tumor infiltrating lymphocytes in ductal and lobular breast cancers. *BMC Cancer*. 2012;12:134.
14. Yang ZY, Guo J, Li N, Qian M, Wang SN, Zhu XL. Mitosin/CENP-F is a conserved kinetochore protein subjected to cytoplasmic dynein-mediated poleward transport. *Cell Res*. 2003;13(4):275-283.
15. Liao H, Winkfein RJ, Mack G, Rattner JB, Yen TJ. CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis. *J Cell Biol*. 1995;130(3):507-518.
16. Yang X, Miao BS, Wei CY, et al. Lymphoid-specific helicase promotes the growth and invasion of hepatocellular carcinoma by transcriptional regulation of centromere protein F expression. *Cancer Sci*. 2019;110(7):2133-2144.
17. Li X, Li Y, Xu A, et al. Apoptosis-induced translocation of centromere protein F in its corresponding autoantibody production in hepatocellular carcinoma. *Oncoimmunology*. 2021;10(1):1992104.
18. Brendle A, Brandt A, Johansson R, et al. Single nucleotide polymorphisms in chromosomal instability genes and risk and clinical outcome of breast cancer: a Swedish prospective case-control study. *Eur J Cancer*. 2009;45(3):435-442.
19. Erlanson M, Casiano CA, Tan EM, Lindh J, Roos G, Landberg G. Immunohistochemical analysis of the proliferation associated nuclear antigen CENP-F in non-Hodgkin's lymphoma. *Mod Pathol*. 1999;12(1):69-74.
20. Li MX, Zhang MY, Dong HH, et al. Overexpression of CENPF is associated with progression and poor prognosis of lung adenocarcinoma. *Int J Med Sci*. 2021;18(2):494-504.
21. Sun J, Huang J, Lan J, et al. Overexpression of CENPF correlates with poor prognosis and tumor bone metastasis in breast cancer. *Cancer Cell Int*. 2019;19:264.
22. Chen Q, Wang L, Jiang M, et al. E2F1 interactive with BRCA1 pathway induces HCC two different small molecule metabolism or cell cycle regulation via mitochondrion or CD4+T to cytosol. *J Cell Physiol*. 2018;233(2):1213-1221.
23. Si T, Huang Z, Jiang Y, et al. Expression levels of three key genes CCNB1, CDC20, and CENPF in HCC are associated with antitumor immunity. *Front Oncol*. 2021;11:738841.
24. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48(W1):W509-W514.
25. Sha D, Jin Z, Budczies J, Kluck K, Stenzinger A, Sinicrope FA. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov*. 2020;10(12):1808-1825.
26. Kang K, Xie F, Mao J, Bai Y, Wang X. Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma. *Front Oncol*. 2020;10:573141.
27. Shain AH, Yeh I, Kovalyshyn I, et al. The genetic evolution of melanoma from precursor lesions. *N Engl J Med*. 2015;373(20):1926-1936.
28. Schadendorf D, van Akkooi ACJ, Berking C, et al. Melanoma. *Lancet*. 2018;392(10151):971-984.
29. Luo J, Duggan DJ, Chen Y, et al. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Can Res*. 2001;61(12):4683-4688.
30. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol*. 2004;22(9):1564-1571.
31. Talantov D, Mazumder A, Yu JX, et al. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin Cancer Res*. 2005;11(20):7234-7242.

32. Zhu X, Mancini MA, Chang KH, et al. Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression. *Mol Cell Biol.* 1995;15(9):5017-5029.

33. Laoukili J, Kooistra MR, Bras A, et al. FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat Cell Biol.* 2005;7(2):126-136.

34. Lin SC, Kao CY, Lee HJ, et al. Dysregulation of miRNAs-COUP-TFII-FOXM1-CENPF axis contributes to the metastasis of prostate cancer. *Nat Commun.* 2016;7:11418.

35. Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature.* 1999;401(6754):708-712.

36. Nunes JPL, Cunha AC, Meirinhos T, et al. Prevalence of auto-antibodies associated to pulmonary arterial hypertension in scleroderma - A review. *Autoimmun Rev.* 2018;17(12):1186-1201.

37. Beura LK, Fares-Frederickson NJ, Steinert EM, et al. CD4(+) resident memory T cells dominate immunosurveillance and orchestrate local recall responses. *J Exp Med.* 2019;216(5):1214-1229.

38. Johnson DB, McDonnell WJ, Gonzalez-Ericsson PI, et al. A case report of clonal EBV-like memory CD4(+) T cell activation in fatal checkpoint inhibitor-induced encephalitis. *Nat Med.* 2019;25(8):1243-1250.

39. Masopust D, Kaech SM, Wherry EJ, Ahmed R. The role of programming in memory T-cell development. *Curr Opin Immunol.* 2004;16(2):217-225.

40. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* 2016;34(18):2157-2164.

41. Oh B, Gransar H, Callister T, et al. Development and validation of a simple-to-use nomogram for predicting 5-, 10-, and 15-year survival in asymptomatic adults undergoing coronary artery calcium scoring. *JACC Cardiovasc Imaging.* 2018;11(3):450-458.

42. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* 2015;16(4):e173-180.

43. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D613.

44. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504.

45. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.

46. Sang L, Wang XM, Xu DY, Zhao WJ. Bioinformatics analysis of aberrantly methylated-differentially expressed genes and pathways in hepatocellular carcinoma. *World J Gastroenterol.* 2018;24(24):2605-2616.

47. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;45(W1):W98-W102.

48. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.

49. Langfelder P, Fast S, Horvath R. Functions for robust correlations and hierarchical clustering. *J Stat Softw.* 2012;46(11):i11.

50. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453-457.

51. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.

52. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747-1756.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.