Original Research

PAD4-dependent citrullination of nuclear translocation of GSK3β promotes colorectal cancer progression via the degradation of nuclear CDKN1A

Xiaonuan Luo [a]; Shanshan Chang [a]; Siyu Xiao [a]; Yin Peng [a]; Yuli Gao [a]; Fan Hu [a]; Jianxue Liang [a]; Yidan Xu [a]; Kaining Du [a]; Yang Chen [a]; Jiequan Qin [a]; Stephen J. Meltzer [b]; Shiqi Deng [a]; Xianling Feng [a]; Xinmin Fan [a]; Gangqiang Hou [c]; Zhe Jin [a]; Xiaojing Zhang [a,*]

[a] Guangdong Provincial Key Laboratory of Genome Stability and Disease Prevention and Regional Immunity and Diseases, Department of Pathology, Shenzhen University School of Medicine, Shenzhen, Guangdong, People's Republic of China
[b] Department of Medicine/GI Division, Johns Hopkins University School of Medicine and Sidney Ki-mmel Comprehensive Cancer Center, Baltimore, MD, USA.
[c] Department of Medical Image Center, Kangning Hospital of Shenzhen, Shenzhen, Guangdong Province, People's Republic of China

Abstract

Peptidylarginine deiminase 4 (PAD4), a Ca^{2+}-dependent enzyme, catalyzes the conversion of arginine to citrulline and has been strongly associated with many malignant tumors. However, the molecular mechanisms of PAD4 in the development and progression of colorectal cancer (CRC) remain unclearly defined. In our study, PAD4 expression was increased in CRC tissues and cells, and was closely related to tumor size, lymph node metastasis. Moreover, the transcription factor KLF9 directly bound to *PADI4* gene promoter, leading to overexpression of PAD4 in CRC cells, which augmented cell growth and migration. We revealed that PAD4 interacted with and citrullinated glycogen synthase kinase-3β (GSK3β) in CRC cells, and GSK3β Arg-344 was the dominating PAD4-citrullination site. Furthermore, IgL2 and catalytic domains of PAD4 directly bound to the kinase domain of GSK3β in CRC cells. Mechanistically, PAD4 promoted the transport of GSK3β from the cytoplasm to the nucleus, thereby increasing the ubiquitin-dependent proteasome degradation of nuclear cyclin-dependent kinase inhibitor 1 (CDKN1A). Our study is the first to reveal the details of a critical PAD4/GSK3β/CDKN1A signaling axis for CRC progression, and provides evidence that PAD4 is a potential diagnosis biomarker and therapeutic target in CRC.

Introduction

Colorectal cancer (CRC) is one of the most prevalent cancers worldwide [1]. The majority of CRC cases (88-94%) are sporadic, while roughly 5-10% of CRC develops from hereditary factors [2–4]. CRC is characterized by high heterogeneity and genetic instability, and gene mutations, such as in *APC*, *TP53, KRAS*, and *PIK3CA*, are the basis for the development of CRC [5,6]. Therefore, a more comprehensive understanding of the mechanisms of CRC progression and metastasis is important for improving the prognosis of CRC patients.

Citrullination is a post-transcriptional modification that catalyzes the conversion of positively charged arginine into electrically neutral citrulline residues in the presence of Ca^{2+} [7]. This reaction is catalyzed by peptidylarginine deiminases (PADs) and five highly conserved PAD enzymes (*I-IV* and *VI*) have been reported in mammals so far [8,9] and each exhibits specific tissue distribution and substrate specificities [10]. Although PAD4 is the only member of the PAD family that contains a distinct nuclear localization sequence [11], all the PAD members can enter into nucleus, and PAD1, PAD2 and PAD4 can citrullinate histone [12–17]. PAD4, a 663-

amino-acid protein with a molecular weigh of 74 kDa, is the most widely studied deiminase [18]. It has five Ca^{2+}-binding sites and Ca^{2+}-binding is capable of inducing conformational changes to generate the active site [18]. PAD4 binds and converts arginine and monomethylarginine to citrulline in histones H2A, H3, and H4. Arginine deimination of histones is central for transcriptional regulation [8] and formation of neutrophil extracellular traps (NETs) [19]. Moreover, researches show that non-histone proteins, such as ING4 [20], p300 [21] and p53 [22] are also targets of PAD4.

Reports have also confirmed that PAD4 shows ectopic expression in various tumors and plays a critical role in cancer cells proliferation or migration [23–30]. In gastric cancer (GC), PAD4 acceblerates migration and metastasis through promoting interleukin 8 (IL-8) expression in GC cells [23]. Cui et al. [24] reveal that, PAD4 promotes tumorigenesis through regulating the expression of WAS/WASL-interacting protein family member 1 and insulin-like growth factor 1 in ovarian cancer. PAD4 stimulates esophageal squamous cell carcinoma cells proliferation and up-regulates CA9 expression [27]. In addition, PAD4 mediates epithelial–mesenchymal transition (EMT) and promotes tumor progression in lung cancer [28]. In CRC, citrullination of the extracellular matrix by tumor cell derived PAD4 is essential for the growth of liver metastases [31]. Therefore, PAD4 is a potential therapeutic target in various cancers. Pan PAD inhibitor Cl-amidine [32–35] and specific PAD4 inhibitor GSK199 [36] and GSK484 [37,38] have been developed and applied to inhibit the deiminase activity of PAD4 [16]. So far, Cl-amidine, GSK199 and GSK484 are the recognized PAD inhibitors commonly. Cl-amidine is a irreversible pan PAD-inhibitor which covalently binds to the active site cysteine of PAD4 and decreases the binding between PAD4 and natural substrates [9,16]. GSK199 and GSK484 are reversible specific-PAD4 inhibitor. These two compounds preferentially bind to PAD4 through competing with calcium and decrease the activity of PAD4 [16]. However, irreversible inhibition should make a better repeating result, when compared with the reversible inhibition. In the other hand, concentration of the irreversible inhibition treatment should be controlled better. Therefore, Cl-amidine have been used to inhibit PAD4 activity *in vivo* and *in vitro* in many scientific reports [32–35].

In our study, we demonstrated that PAD4 expression was increased in CRC tissues and cells, and promoted CRC cell growth and migration. Glycogen synthase kinase-3β (GSK3β) Arg-344 was identified as a PAD4-dependent citrullinated site in CRC cells. Mechanically, PAD4-driven citrullination of GSK3β contributed to the nuclear translocation of GSK3β in HCT116 cells, which increased the ubiquitin-dependent proteasome degradation of nuclear cyclin-dependent kinase inhibitor 1 (CDKN1A). Therefore, we hypothesize that the biological function of PAD4 in CRC cells depends on the PAD4-induced nuclear accumulation of GSK3β.

Materials and methods

Cell culture, patient samples, treatments with Cl-amidine and MG-132 and A23187

NCM460 cells were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). NCM460 cell line is an epithelial cell line derived from the normal colon and is therefore used as the control cell line [39]. HCT116, SW480, HT29, and SW620 cells were obtained from the American Type Culture Collection. HCT116 [40] and HT29 [41] cells are separated isolated from primary tumors obtained from colorectal adenocarcinoma patients. SW480 cells are obtained from the primary lesion of colon cancer patient, and SW620 cells are isolated from lymph node metastasis of the same patient [42]. All cells were grown in DMEM media supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37°C in a 5% CO_2 incubator.

Clinical samples, including 24 primary human CRC tissues and 24 corresponding adjacent non-cancerous tissues, were obtained by surgical resection from patients diagnosed with CRC at the first Affiliated Hospital of Shenzhen University. All the clinical-pathology data was shown in Table S1. All processes were consistent with the requirement of the Institutional Review Board of The School Medicine of Shenzhen University.

In our study, we performed the drug concentration assays. The plasmid FLAG-PAD4 was transfected into HCT116 cells, and 24 h after transfection, fresh medium containing different concentrations of Cl-amidine (HY-100574, MCE) was added to the transfected cells and incubated for 48 h. Western blot showed that the citrullination level of H3 was decreased to 20% with 200 μM Cl-amidine treatment, compared with the 100 μM Cl-amidine or untreated HCT116 cells (Fig. S1). Thus, we chose 200 μM Cl-amidine treatment in the follow-up experiments. To inhibit the deiminase activity of PAD4, the plasmid FLAG-PAD4 was transfected into HCT116 cells, and 24 h after transfection, fresh medium containing 200 μM Cl-amidine was added to the transfected cells and incubated for 48 h. FLAG-PAD4 were transiently transfected into HCT116 cells [43]. At 24 h after transfection, the cells were treated with a calcium ionophore 5 uM A23187 (HY-N6678, MCE) for 2 h. To inhibit proteasome activity, HCT116 cells were incubated with DMEM medium containing the proteasome inhibitor MG132 (HY-13259, MCE), at a concentration of 10 μM for 24 h at 37°C.

Tissue microarrays and immunohistochemistry

Tissue microarrays (HColA180Su19) were purchased from Outdo Biotech (Shanghai), containing eighty-six CRC tissues and para-cancer tissues, and eight CRC tissues. All patients were followed up for 5-6 years, and their complete clinic-pathologic data were collected for further analysis. Immunohistochemistry (IHC) was performed to determine the relative expression of PAD4 according to the protocol. In brief, tissue sections were incubated at 4°C with anti-PAD4 antibody (1:2000, ab128086, Abcam) overnight. After washing three times with PBS, the sections were incubated with biotinylated anti-mouse IgG (1:1000, 58802, Cell Signaling). Finally, the DAB system was used to visualize the signal, and hematoxylin was used to stain the nucleus. The immunostaining images were captured using an upright Zeiss LSM 510 confocal laser scanning microscope. The IHC score was calculated according to the intensity and extent of the IHC staining and was independently determined by two pathologists who were blinded to the patient characteristics (Table S2).

Quantitative reverse-transcription PCR

Total RNA from tissue samples and cells was extracted using TRIzol reagent (Invitrogen) according to established protocols. Total RNA was reversely transcribed using the PrimeScript RT reagent Kit (RR037A, TaKaRa) to generate cDNA. The qPCR was performed using GoTaq qPCR Master Mix (A6001, Promega) in triplicate independent biological experiments. All samples were normalized to the internal control gene human *ACTB*, and relative changes in gene expression were calculated using the comparative Ct method ($2^{-\Delta\Delta Ct}$) (Table S3).

Cell proliferation and migration assays

The 5-ethynyl-2′-deoxyuridine (EdU) and wound-healing assays were performed according to established protocols [44,45]. For the Transwell assay, HCT116 and SW480 cells in 200 μl of DMEM with 2% FBS were seeded into the upper chamber. The bottom chamber was filled with 600 μl culture media with 20% FBS. After incubation for 24 h, the migrated cells were fixed with 4% paraformaldehyde and stained with 0.5% crystal violet. For the growth curve assay, cells were seeded into a 96-well plate, and MTS (ab197010, Abcam) was added to each well (10 μl/ well) at 24 h, 48 h, 72 h, and 96 h. Then the plates were incubated with MTS reagent for 4 h, and cell viability was evaluated by measuring the absorbance at 490 nm.

Plasmid constructs and RNA interference

GSK3β genes 1-1029bp and 1010-1263bp (1030bp C-A and 1031bp G-A) were amplified from the cDNA of HCT116 cells, then these two genes are both cloned into the vector pIRESneo-FLAG/HA-EYFP, generating vector GSK3β-R344K. *FLAG-PADI4* and *GSK3β* genes were amplified from the cDNA of HCT116 cells; These PCR sequences were separately cloned into the vector pIRESneo-FLAG/HA-EYFP.

PADI4 gene was amplified from the cDNA of HCT116 cells, *mCherry* gene was amplified from plasmid pcDNA3.1-mCherry, *P2A* gene was amplified from plasmid lentiCRISPR-v2 and *GFP* gene was amplified from plasmid pcDNA3.1-EGFP. *PADI4, mCherry, P2A, GFP* and *GSK3β* or *GSK3β-R344K* were all cloned into the vector pIRESneo-FLAG/HA-EYFP, generating vector mCherry-PAD4-P2A-GFP-GSK3β and mCherry-PAD4-P2A-GFP-GSK3β-R344K.

GSK3β genes 1021-1169bp (1148bp C-A and 1149bp G-A) and 1127-1263bp (1148bp C-A and 1149bp G-A) were amplified from the cDNA of HCT116 cells, then these two genes are both cloned into the vector pGEX-4T1 vector, generating vector GSK3β-C domian-R383K. The *GSK3β, GSK3β-N domian, GSK3β-kinase domain, GSK3β-C domain, GSK3β-C domain-R344K, GSK3β-C domian-R354K, GSK3β-C domain-R405K, PADI4, PADI4-IgL1 domain, PADI4-IgL2 domain, and PADI4-catalytic domain* sequences were amplified from cDNA of HCT116 cells and separately cloned into the pGEX-4T1 vector.

The promoter of human *PADI4*, a series of 5′-unidirectional truncation mutants, and point mutants of *PADI4* promoter were obtained from the DNA of HCT116 cells and separately cloned into pGL3-Basic. All vectors were constructed using the ClonExpress II One-Step Cloning kit (Vazyme Biotech Co., Ltd.). All constructed vectors were validated by DNA sequencing by Shanghai Sangon (Shanghai, China). siRNAs targeting human *GSK3β* (si1-GSK3β, si2-GSK3β, si3-GSK3β) and siRNA-NC were synthesized by Ribobio Co. (Guangzhou, China). The mix of the three siRNAs against *GSK3β* was termed as simix-GSK3β. All primer sequences and siRNA sequences are listed in Table S3.

Recombinant protein purification

Plasmids were separately transformed into *E. coli* BL21 (DE3). The transformed bacteria were grown overnight in 2 ml LB medium with kanamycin 50 μg/mL at 37°C and shaking at 200 rpm. The following day, 2 ml of bacterial culture was used to inoculate 100 ml of LB medium, which was incubated until the culture reached an absorbance approximately 0.3-0.6 at 600 nm. Then 0.4 mM IPTG was added to induce the production of recombinant protein at 18°C for 16 h with shaking at 180 rpm. Next, 100 ml of the culture was centrifuged at 8000 × *g* at 4°C for 10 min, and the pellet was resuspended in 5 ml PBS. Subsequently, the media was sonicated for 20 cycles (10s on and 10s off) on ice and centrifuged at 12,000 × *g* at 4°C for 30 min. The protein in the supernatant was purified according to the manufacturer's instructions with GST Resin (Beyotime).

GST pull down assay, immunoprecipitation, and western blot assays

GST pull down experiments were performed using purified proteins and cell extraction. Briefly, the purified GST fusion protein was immobilized on GST resin (Beyotime), and GST fusion protein was incubated with cell extraction at 4°C for 1 h with constant shaking. After incubation, the resin was washed thoroughly four times with PBS. Any retained proteins were eluted by the addition of 10 mM L-glutathione reduced in 50 mM Tris-HCl (pH 8.0).

For immunoprecipitation (IP) and co-immunoprecipitation (co-IP), cells were lysed in lysis buffer (PH 7.6, 25 mM Tris-HCl, 150 mM NaCl, 1% NP40) and incubated with anti-FLAG (1:1000, F1804, Sigma) or anti-GSK3β (1:1000, 12456, Cell Signaling) at 4°C overnight with shaking. The immune complex was incubated with protein A/G magnetic beads (Thermo) at room temperature for 1 h with shaking and then washed with lysis buffer. The bound immune complex was dissociated from the beads by incubating with SDS-PAGE sample loading buffer for 10 min at room temperature. Samples were then subjected to electrophoresis and western blotting. All the primary antibodies used for the western blot analysis were anti-PAD4 (1: 2000, ab128086, Abcam), anti-GSK3β (1:1000, 12456, Cell Signaling), anti-GAPDH (1:3000, 2118, Cell Signaling), anti-FLAG (1:1000, F1804, Sigma), anti-H3-cit (1:5000, ab281584, Abcam), anti-H3 (1:5000, ab1791, Abcam), anti-modified-citrulline (1:1000, 17-347, Millipore), anti-HDAC1 (1:1000, 5356, Cell Signaling), anti-KLF9 (1:1000, ab227920, Abcam), anti-CDKN1A (1:1000, 2947, Cell Signaling), and anti-ubiquitin (1:1000, 3936, Cell Signaling) antibodies.

Protein citrullination assay

Protein citrullination assays were performed by incubating recombinant PAD4 (ab196393, Abcam) with purified proteins at 37°C for 1 h in 100 mM Tris-HCl, pH 7.4, containing 5 mM DTT and 10 mM $CaCl_2$. Reactions were stopped by the addition of SDS-PAGE sample loading buffer. Samples were then subjected to electrophoresis and western blotting. Citrulline was detected with an anti-modified-citrulline detection kit (17-347, Millipore) according to the manufacturer's instructions.

Immunofluorescence (IF) staining

HCT116 cells at 3×10^5 were plated into a confocal dish. The next day, they were fixed with 4% formaldehyde for 20 min and blocked with blocking buffer (5% BSA and 0.3% Triton X-100) for 1 h. The cells were incubated with anti-FLAG (1:1000, F1804, Sigma) or anti-PAD4 (1: 2000, ab128086, Abcam) or anti-GSK3β (1:1000, 12456, Cell Signaling) overnight at 4°C and then with secondary antibodies conjugated with Alexa Fluor for 1 h. Then the cells were incubated with DAPI (1:1000, C10310, RiboBio Inc.) for 30 min and photographed using an upright Zeiss LSM 510 confocal laser scanning microscope.

Luciferase activity assay

HCT116 cells (1×10^5) were seeded into 24-well plates and allowed to settle overnight. Firefly plasmid (0.5 μg) and renilla plasmid (0.01 μg) were co-transfected into HCT116 cells using lipofectamine 3000 (Thermo). At 48 h post-transfection, the firefly luciferase and renilla luciferase luminescence was measured using a dual luciferase reporter assay kit (E1960, Promega) according to the manufacturer's instructions. The firefly luciferase activity was normalized for transfection efficiency using the corresponding renilla luciferase activity.

Chromatin immunoprecipitation

HCT116 cells were harvested, and the Chromatin immunoprecipitation (ChIP) experiments were performed using Magna ChIP protein G Kit (17-611, Millipore) according to the manufacturer's protocol. In brief, the cross-linked chromatin was sonicated into fragments, and cell lysates were immunoprecipitated using a KLF9 antibody. Finally, the precipitated DNA was analyzed by PCR with the primers listed in Table S3.

RNA sequencing

RNA isolation and sequencing were performed by LC Sciences. Briefly, total RNA was extracted using the RNeasy mini kit (Qiagen), and RNA

concentrations were measured using Quant-ITTM RiboGreen RNA Reagent (Invitrogen, USA). The Agilent RNA 6000 Nano kit (Agilent 2100 Bioanalyzer) was used to assess the integrity of the total RNA. High-quality mRNA was fragmented into small pieces using RNA Fragmentation Reagents (AM8740, Invitrogen), then the cleaved RNA fragments were reverse transcribed to produce the final cDNA library using the mRNA-seq sample preparation kit (Illumina, USA). The library products were sequenced on the Illumina Hiseq 2500 platform (Illumina) following the vendor's recommended protocols.

Statistical analysis

Statistical analyses were carried out using GraphPad Prism. All experiments were conducted in three biological replicates and the data are presented as mean ± SD. Two-tailed Student's t-test was used to assess the statistical significance of differences among groups. The chi-square test was employed to evaluate clinicopathological features. Survival rate was evaluated using the Kaplan-Meier method, and differences between survival curves were tested with the log-rank test. For all tests, the significance level for statistical analysis was set at p value < 0.05.

Results

PAD4 is highly expressed in CRC cells and correlated with clinicopathologic parameters

To reveal the role of PAD4 in CRC, we investigated the expression level of *PADI4* in 24 paired tumor tissue specimens from CRC patients by RT-qPCR. The results showed that the expression of *PADI4* was higher in 14 of the 24 tumor tissues compared with their corresponding non-cancerous controls. *PADI4* expression was significantly increased in the overall statistical analysis of the 24 paired tumor tissue specimens (Fig. 1A). *PADI4* expression levels were also measured in normal human colon mucosal epithelial cells NCM460 cells and CRC cells HCT116, HT29, SW480, and SW620 cells. *PADI4* expression was markedly up-regulated in HT29 and SW620 cells compared with normal colon mucosal epithelial cells in the RT-qPCR and western blot analysis (Fig. 1B). Consistent with these results, western blotting analysis confirmed that PAD4 was overexpressed in human CRC specimens (Fig. 1C). IHC showed that PAD4 was expressed in both the nucleus and cytoplasm, but mainly in the nucleus in CRC tissue (Fig. 1D). In a microarray of 73 paired CRC tissue samples, PAD4 expression was significantly increased, and more intense staining was observed in cancer tissues than the matched adjacent normal tissues. Sixty-four (87.7%) of seventy-three CRC specimens exhibited high levels of PAD4 protein, while only twenty nine (39.7%) normal mucose specimens expressed abundant PAD4 (p < 0.0001) (Fig. 1E and F). Similar results were observed in human CRC samples compared with matched non-CRC tumors from the clinical proteomic tumor analysis consortium (CPTAC) database (Fig. 1G). Additionally, ROC curve was generated based on the data from a microarray of CRC tissues (91 CRC patients and 73 healthy subjects), and the area under the ROC curve (AUC) PAD4 was 0.7457 (95%, CI=0.6684–0.8229) (Fig. 1H). Together, these results demonstrated that PAD4 was over-expressed in CRC tissues and cells, and could serve as a new diagnosis biomarker of CRC.

Then we divided 91 CRC patients into two groups: high-expression group (n = 78) and low-expression group (n = 13) according to PAD4 expression level from IHC staining. Kaplan-Meier analysis showed that PAD4 expression had no significant correlation with overall survival of CRC patients (Fig. S2). As presented in the statistical data Table 1, high PAD4 expression in CRC tissues was strongly correlated with increased tumor size and positive lymphatic metastasis (p < 0.05), although there were no significant differences in the expression level of PAD4 according to

Table 1

The relationship between PAD4 and clinical characteristics of CRC patients.

Features	n	PAD4 expression		p value
		High	Low	
Number	91	78	13	
Gender				0.145
Female	45	36	9	
Male	46	42	4	
Age				0.255
<62	36	29	7	
>=62	55	49	6	
Tumor size				*0.032**
<5 cm	40	27	13	
>=5 cm	51	44	7	
Clinical staging				0.596
I+II	57	48	9	
III+IV	34	30	4	
Lymphatic metastasis				*0.039**
No	59	40	19	
Yes	32	28	4	
Remote metastasis				0.348
No	86	73	13	
Yes	5	5	0	

* p value< 0.05

gender, age, clinical stage, or distant metastasis (all p > 0.05). Collectively, these findings demonstrated that there is a possible link between increased expression of PAD4 and CRC progression.

PAD4 promotes cell proliferation and migration of CRC in vitro

The frequent overexpression of PAD4 in CRC tumors and cell lines prompted us to investigate the biological role of PAD4 in CRC cells. Subsequent western blot analysis revealed PAD4 was ectopically expressed, and PAD4 induced the citrullination of histone 3 in FLAG-PAD4-transfected SW480 and HCT116 cells (Fig. 2A), suggesting that PAD4 has deiminase activity. The effect of PAD4 on cell metastasis was assessed using transwell and wound-healing assays. PAD4 expression led to the promotion of the migratory capacity of SW480 and HCT116 cells (Fig. 2B, D). To determine whether the deiminase activity of PAD4 is required for the up-regulated migration of CRC cells, PAD4 inhibitor Cl-amidine was added to FLAG-PAD4-transfected HCT116 cells to inhibit deiminase activity. The wound-healing assay showed that Cl-amidine treatment effectively reduced the up-regulated migration capacity of the PAD4-overexpressed HCT116 cells, indicating that deiminase activity PAD4 is essential for the PAD4-mediated the up-regulation of cell migration (Fig. 2E). Collectively, these results indicated that PAD4 can promote the migration of CRC cells *in vitro*.

To detect the function of PAD4 in cell growth, EdU assay was employed. The ectopic expression of PAD4 significantly promoted the growth of SW480 and HCT116 cells (Fig. 2F). However, Cl-amidine treatment decreased the proliferation ability of FLAG-PAD4-transfected HCT116 cells (Fig. 2G). Consistent with this, the cell growth curve assay showed that the forced expression of PAD4 significantly increased SW480 cells proliferation. However, PAD4 expression in HCT116 cells did not change their proliferation ability in any obvious way in the cell growth curve assay (Fig. 2C). Taken together, it suggested that PAD4 promotes CRC cells growth *in vitro*.
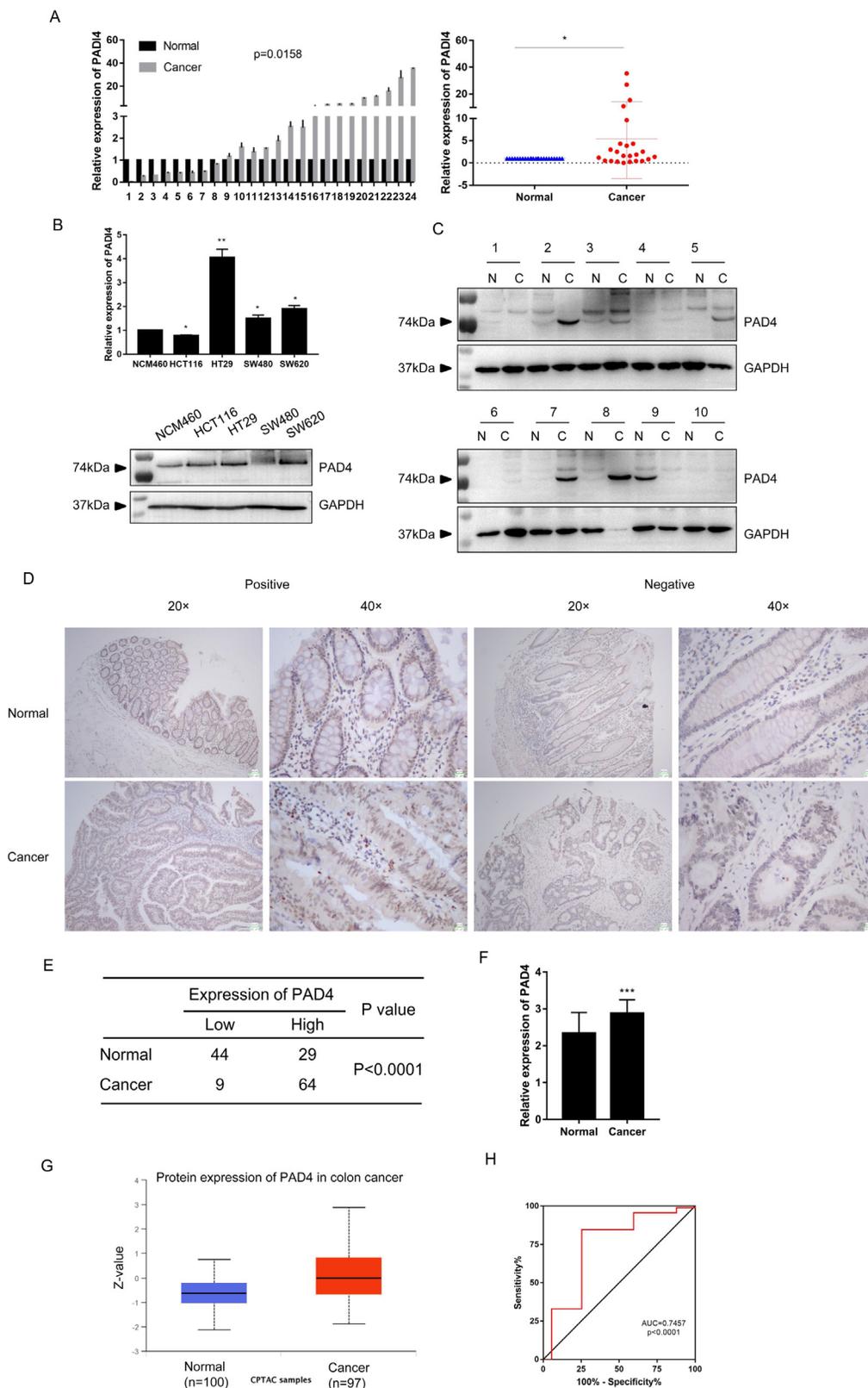
Fig. 1. PAD4 is highly expressed in CRC. (A) RT-qPCR analysis of PAD4 expression in 24 pairs of CRC tissues. The values of controls were normalized to 1. (B) Western blot and RT-qPCR analyses of PAD4 expression in CRC cells. (C) Western blot analyses of PAD4 expression in 10 pairs of CRC tissues. (D) IHC analysis of PAD4 expression in 73 pairs of CRC microarray tissues. (E) Statistical analysis for PAD4 proteins visualized by IHC in CRC tissues. (F) Expression levels of PAD4 proteins visualized by IHC in 73 pairs of CRC tissues. (G) PAD4 expression in CRC patients from CPTAC database. (H) ROC curves were used to determine the diagnostic value of PAD4 in CRC tissues. All the experiments are conducted in three biological replicates, and error bars represent SD. *** p value < 0.001, ** p value < 0.01, * p value < 0.05.
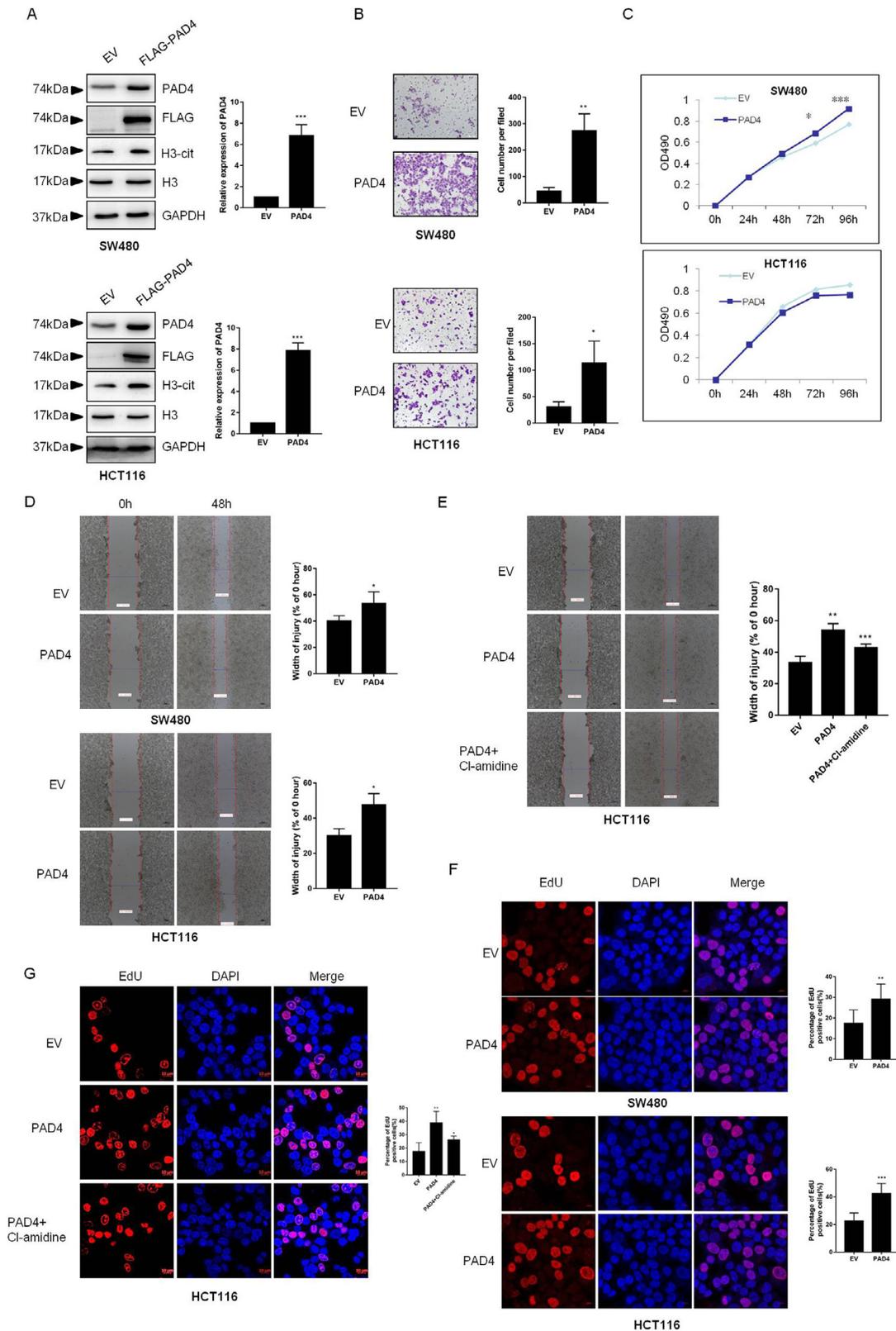
Fig. 2. Overexpression of PAD4 promotes cell proliferation and migration in CRC cells. (A) The expression of PAD4 in SW480 and HCT116 cells transfected with FLAG-PAD4 was confirmed by western blotting. GAPDH was used as a loading control. (B) Transwell assay for PAD4-overexpressed SW480 and HCT116 cells to determine the effect on cell migration, invasive cells were stained and counted under microscope at 24 h after reseeding. (C) Cell growth curve assay was conducted to analyze the proliferation of SW480 and HCT116 cells transfected with FLAG-PAD4. (D) and (E) Scratch test was produced to determine the effect of PAD4 overexpression on the migration rate based on width of SW480 and HCT116 cells with without Cl-amidine. (F) and (G) EdU proliferation assay was conducted on FLAG-PAD4-transfected HCT116 and SW480 cells treated with or without Cl-amidine. All the experiments are conducted in three biological replicates, and error bars represent SD. *** p value < 0.001, ** p value < 0.01, * p value < 0.05.
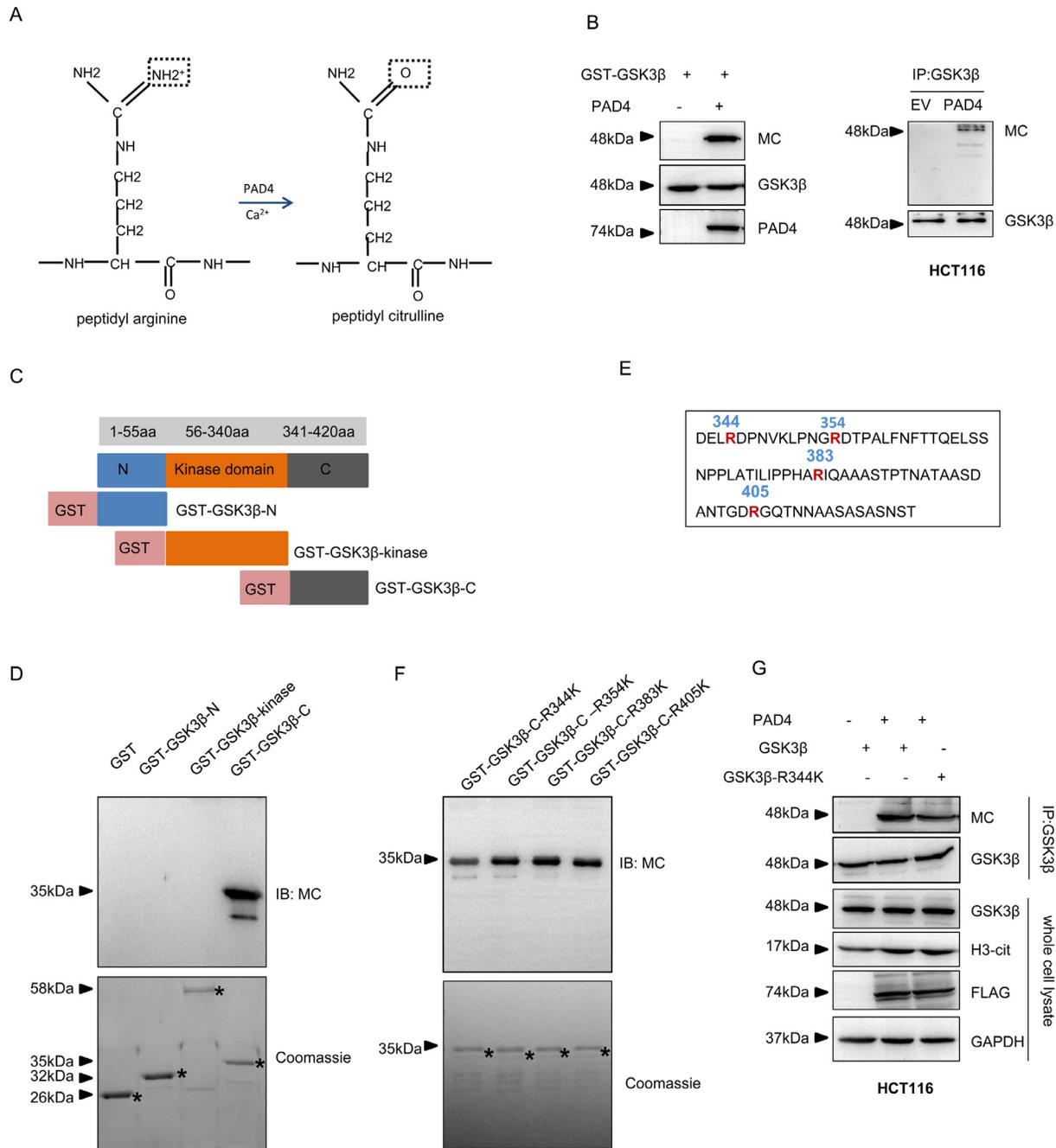
Fig. 3. PAD4 citrullinates R344 of GSK3β C-domain in CRC cells. (A) Schematic representation of PAD4-mediated citrullination reaction. (B) PAD4 citrullinates GSK3β *in vitro* (left). PAD4 was treated with GSK3β at 37°C for 1 h. Samples were analyzed on a western blot using an anti-modified-citrulline antibody. PAD4 citrullinates GSK3β in HCT116 cells (right). FLAG-PAD4 was transfected into HCT116 cell, GSK3β was immunoprecipitated to detect citrulline levels. (C) Schematic of GSK3β truncated sequences. (D) Citrullinated domain (s) of GSK3β was mapped. (E) Arginine residue in GSK3β C domain was mapped. (F) Citrullination assays identified citrulline site (s) on GSK3β C domain. (G) Citrullination by PAD4 was much weaker when GSK3β Arginine 344 was mutated in HCT16 cells. All the experiments are conducted in three biological replicates.

PAD4 citrullinates GSK3β R344 in CRC cells

PAD4 catalyzes a guanidino group of arginine into a urea group in a post-translational modification process (Fig. 3A) [46]. To explore the potential targets of PAD4 in CRC, we searched for the proteins beginning with MSGR from the UniProt database. Among the identified proteins, GSK3β attracted our attention, as it is a serine/threonine protein kinase that plays vital roles in multiple tumor types (Table 2).

To identify whether PAD4 citrullinates GSK3β, the western blotting showed that PAD4 citrullinated GSK3β (Fig. 3B left). In PAD4-overexpressed HCT116 cells, GSK3β was immunoprecipitated and probed for citrulline content. Western blotting showed that GSK3β was clearly citrullinated in the presence of PAD4 in HCT116 cells (Fig. 3B right). These result indicated that PAD4 citrullinates GSK3β in CRC cells.

To determine which domain(s) of GSK3β is preferentially citrullinated, GST-GSK3β-N domain, kinase domain, and C domain were purified and

Table 2

List of 10 proteins containing an amino-terminal MSGR sequence

Uniprot ID	Protein	Full name
P49841	*GSK3β*	*Glycogen synthase kinase-3 beta*
A0A578	TRBV5-1	T cell receptor beta variable 5-1
F5GX29	CHP1	Calcineurin B homologous protein 1
Q96GE9	DMAC1	Distal membrane-arm assembly complex protein 1
Q99944	EGFL8	Epidermal growth factor-like protein 8
B4E3F4	ITGA5	Integrin alpha-5
E5RHR4	CCDC121	Coiled-coil domain-containing protein 121
C9JXP5	NME9	Thioredoxin domain-containing protein 6
Q6FGU7	RAB7L1	Ras-related protein Rab
Q86SQ3	ADGRE4P	Putative adhesion G protein-coupled receptor E4P

incubated with the activated recombinant PAD4 (Fig. 3C). Western blotting showed that a strong citrullination signal existed in the C domain but not in the N domain or kinase domain of GSK3β (Fig. 3D), indicating that the GSK3β C domain was preferentially citrullinated by PAD4.

To more precisely identify the citrullination sites in the C domain of GSK3β, GSK3β-C domain R344K, R354K, R383K, and R405K were purified (Fig. 3E). The four mutants were separately incubated with the activated recombinant PAD4. Western blotting showed that the R344K mutant displayed a significant reduction of citrullination signal compared with the R354K, R383K, and R405K mutants (Fig. 3F). It indicated that R344 is the major site of citrullination by PAD4. Then, plasmid GSK3β and mutant GSK3β-R344K were constructed and separately transfected into PAD4-overexpressing HCT116 cells. The western blot showed that GSK3β and GSK3β-R344K were separately overexpressed in HCT116 cells. Co-IP assay showed that GSK3β citrullination was significantly increased in the presence of PAD4. However, citrullination by PAD4 was much weaker when GSK3β was mutated at R344 (Fig. 3G). These results demonstrated that GSK3β is a new substrate of PAD4 in CRC, and the major citrullination site is R344 in GSK3β.

PAD4-induced nuclear accumulation of GSK3β promotes proliferation and migration of CRC cells

The expression of GSK3β in CRC cells was detected by RT-qPCR and western blot analyses. Results showed that the expression of GSK3β was markedly up-regulated in HCT116, HT29, and SW480 cells compared with the normal colon mucosal epithelial cell line NCM460 (Fig. S3A). A wound-healing assay demonstrated that the down-regulation of GSK3β by simix-GSK3β (si1-GSK3β + si2-GSK3β + si3-GSK3β) (Fig. 4A) significantly inhibited HCT116 cell migratory viability compared with control cells. However, the over-expression of GSK3β led to no significant effect on migratory viability of HCT116 cells (Fig. S3B). These results indicated that GSK3β is involved in CRC progression.

As our results showed that GSK3β is involved in CRC progression and PAD4 induces GSK3β citrullination in CRC, we aimed to determine whether PAD4-mediated GSK3β citrullination is involved in the proliferation and migration of CRC cells. Simix-GSK3β was transfected into PAD4-overexpressing HCT116 cells to silence their expression of GSK3β (Fig. 4B). A transwell assay showed that the forced expression of PAD4 promoted the migration, whereas GSK3β knockdown inhibited the cell migration caused by the over-expression of PAD4 in HCT116 cells (Fig. 4C). In addition, an EdU assay showed that GSK3β down-regulation decreased the proliferation potential of PAD4-overexpressing HCT116 cells (Fig. 4D). However, the wound-healing assay showed knockdown of GSK3β led to no difference in migratory potential of PAD4-overexpressing HCT116 cells (Fig. S3C). In conclusion, these results demonstrated that PAD4-mediated

promotion of proliferation and migration of HCT116 cells was significantly reversed by knockdown of GSK3β, and the PAD4–GSK3β signaling axis has an important role in CRC progression.

To explore how PAD4–GSK3β axis influences CRC progression, we examined the expression level of GSK3β in HCT116 and SW480 cells transfected with FLAG-PAD4. According to the western blot and RT-qPCR analyses, the expression of total GSK3β was not obviously changed after PAD4 overexpression in HCT116 cells compared with control cells (Fig. S3D, E). Nuclear and cytoplasmic fractionation assays showed that overexpression of PAD4 significantly increased the nuclear GSK3β in HCT116 cells (Fig. 4E). Consistent with this, IF analysis showed that the positive signal of GSK3β was increased in the nucleus after the overexpression of PAD4, which demonstrated that PAD4 enhances nuclear GSK3β in HCT116 cells (Fig. 4F).

To establish whether PAD4-mediated GSK3β citrullination regulates the location of GSK3β, the plasmid cherry-PAD4-P2A-GFP-GSK3β (C-GSK3β) and the mutant plasmid cherry-PAD4-P2A-GFP-GSK3β-R344K (C-GSK3β-R344K) were separately transfected into HCT116 cells (Fig. 4G). IF assays showed that nuclear GFP intensity decreased when GSK3β was mutated at R344 (Fig. 4H). Collectively, these results indicated that the PAD4-induced nuclear accumulation of GSK3β promotes the proliferation and migration of CRC cells.

GSK3β-kinase domain and PAD4-IgL2 and catalytic domains directly bind in HCT116 cells

We used IF staining to validate the physical interactions between PAD4 and GSK3β in HCT116 cells, which revealed that PAD4 was co-localized extensively with GSK3β (Fig. 5A). A GST pull-down assay showed that GST-PAD4 was able to bind to GSK3β from total cell lysate, while GST failed to bind (Fig. 5B), suggesting that PAD4 binds to GSK3β. Similarly, in a co-IP assay, PAD4 co-immunoprecipitated with GSK3β in HCT116 cells (Fig. 5C upper). Reciprocally, GSK3β was shown to co-immunoprecipitate with PAD4 in HCT116 cells (Fig. 5C down). To map the binding between PAD4 and GSK3β in more detail, GST-PAD4 IgL1 domain, IgL2 domain, and catalytic domain (Fig. 5D), were purified. GST pull-down assay showed that IgL2 domain, and catalytic domain but not IgL1 domain of PAD4 were able to pull down GSK3β from HCT116 cells extracts (Fig. 5E), suggesting that the IgL2 and catalytic domains of PAD4 are sufficient to mediate its interaction with GSK3β. To map the region(s) of GSK3β that specifically interacts with PAD4, GST pull-down experiment showed that the kinase domain, but not the N domain or C domain of GSK3β, bound to PAD4 (Fig. 5F), suggesting that the kinase domain of GSK3β is sufficient for binding to PAD4. The above GST pull-down experiments showed there were physical interactions between the kinase domain of GSK3β and the IgL2 and catalytic domains of PAD4 (Fig. 5G).
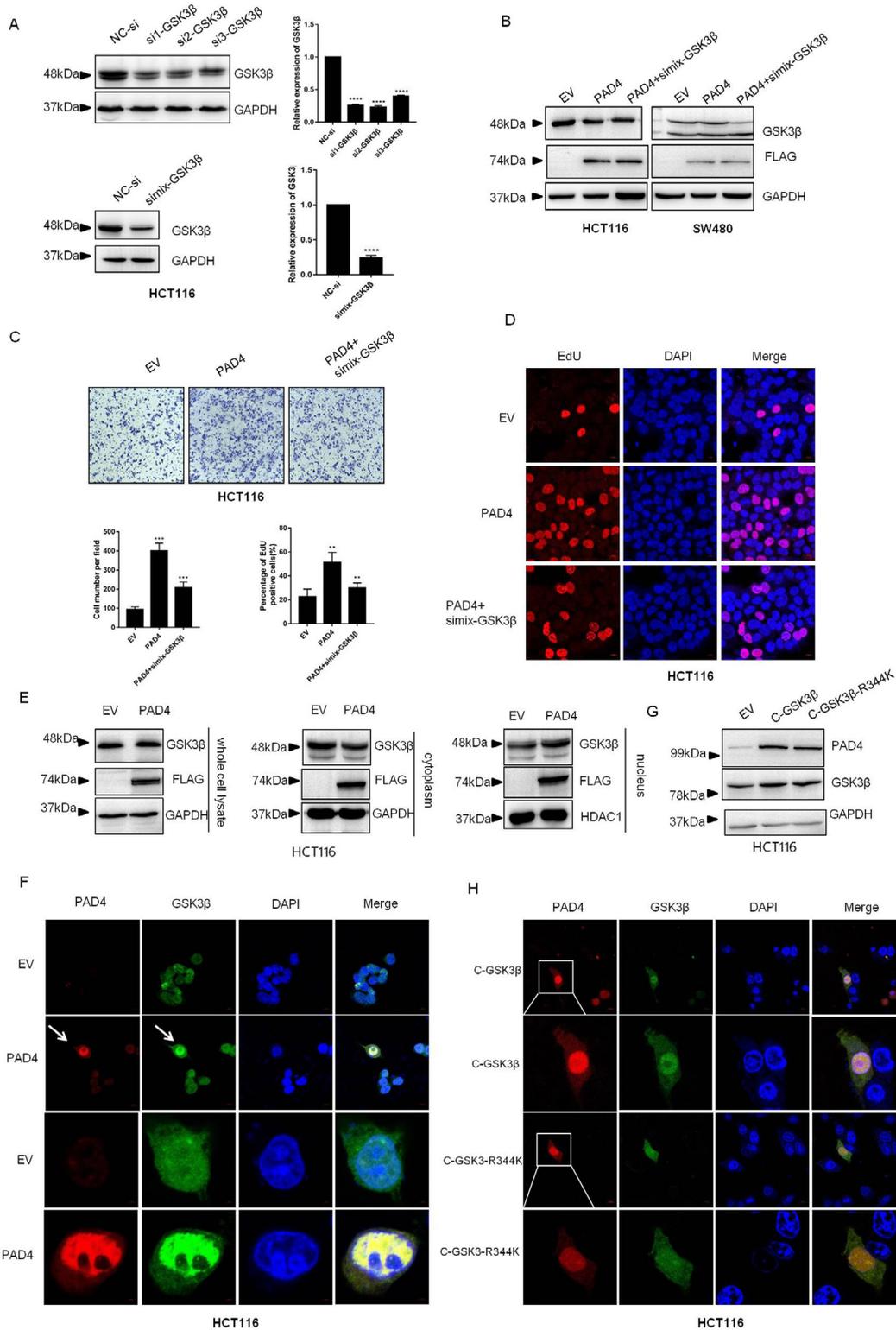
Fig. 4. PAD4-induced nuclear accumulation of GSK3β promotes proliferation and migration of CRC cells. (A) HCT116 cells were transfected with si1-GSK3β, si2-GSK3β, si3-GSK3β and simix-GSK3β, and expression of GSK3β was confirmed by western blot and RT-qPCR. (B) Expression of PAD4 and GSK3β in HCT116 cells co-transfected with FLAG-PAD4 and simix-GSK3β. (C) Transwell assay was conducted on HCT116 cells co-transfected with FLAG-PAD4 and simix-GSK3β. (D) EdU assay was conducted on HCT116 cells co-transfected with FLAG-PAD4 and simix-GSK3β. (E) Expression of GSK3β in whole cell lysate, cytoplasm, and nucleus from FLAG-PAD4 transfected HCT116 cells was identified by western blot. (F) IF analysis of cellular location of GSK3β in FLAG-PAD4 transfected HCT116 cells. (G) Expression of PAD4 and GSK3β in C-GSK3β or C-GSK3β-R344K transfected HCT116 cells. (H) IF analysis of location of GSK3β in C-GSK3β or C-GSK3β-R344K transfected HCT116 cells. All the experiments are conducted in three biological replicates, and error bars represent SD. *** p value < 0.001, ** p value < 0.01, * p value < 0.05.
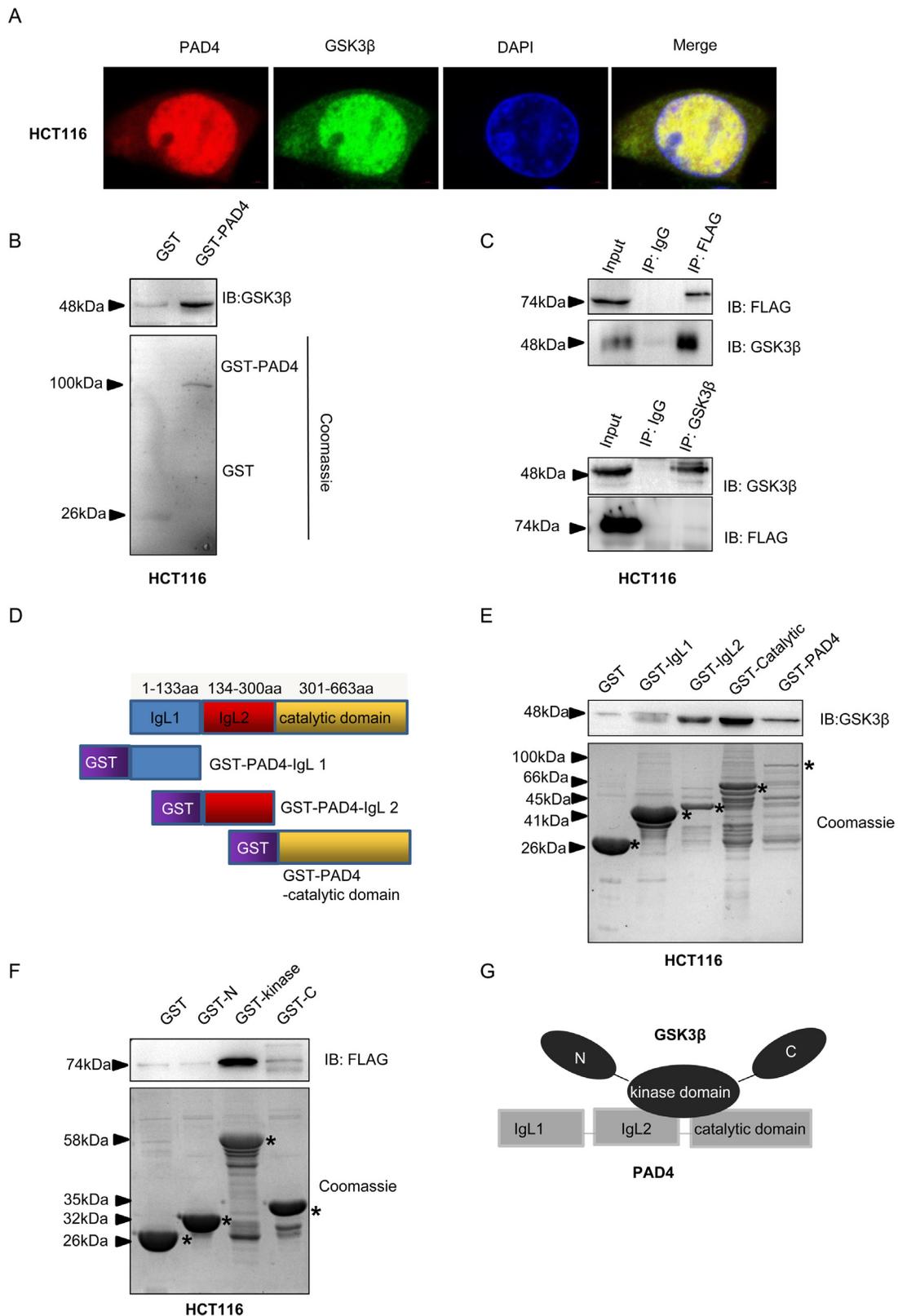
Fig. 5. GSK3β kinase-domain binds to IgL2 and catalytic-domains of PAD4 in CRC cells. (A) IF analysis of locations of GSK3β and PAD4 in FLAG-PAD4 transfected HCT116 cells. (B) GST pull-down indicated that PAD4 associated with GSK3β. (C) Immunoprecipitation showed that PAD4 associated with GSK3β in FLAG-PAD4 transfected HCT116 cells. (D) Schematic of PAD4 truncated sequences. (E) GST pull-down indicated the IgL2 and catalytic domains of PAD4 were sufficient for interaction with GSK3β. (F) GST pull-down indicated the kinase domain of GSK3β was sufficient for interaction with PAD4. (G) There was physical interaction between the kinase domain of GSK3β and the GST-IgL2 and GST-catalytic domains of PAD4. All the experiments are conducted in three biological replicates.

KLF9 upregulates the transcription of PADI4 *gene in CRC cells*

PAD4 is highly expressed in CRC cells and promotes their proliferation and migration. However, little is known about the mechanism that regulates the transcription of *PADI4* in CRC. To understand the upstream regulatory elements controlling *PADI4* expression, we amplified a 2120 bp DNA fragment (-2000 to +120) containing the proximal 5′-flanking region (-2000/-1), transcription start site (+1) and exon 1 (+1/+120) of *PADI4* gene. A series of 5′-truncated sequences of *PADI4* gene were cloned into the luciferase reporter vector pGL3-basic. These constructs were then transiently transfected into HCT116 cells, and luciferase activity was measured. The -2000/+120 region was sufficient to drive the expression of luciferase, which was increased compared with HCT116 cells transfected with pGL3-basic, and −146/+120 region conferred no significant reduction of luciferase activity compared with those in the −2000/+120 region. However, the −78/+120 and −12/+120 regions gradually decreased luciferase activity (Fig. 6A). Therefore, −146/−12 region contains the core promoter for *PADI4* transcription.

When we then analyzed the −146/−12 sequence of *PADI4* using the databases JSAPAR (https://jaspar.genereg.net/), UCSC (http://genome.ucsc.edu/), and PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3#opennewwindow). We predicted four cis-elements for binding with the transcription factors NF-kB, TFAP, KLF9, and Sp3 in −146 to −12 sequence of *PADI4* (Fig. 6B), suggesting that these transcription factors may be involved in regulating the transcription of *PADI4*. Furthermore, mutations in the NF-kB, TFAP, KLF9, or Sp3 binding site of −146/+120 region was separated cloned into pGL3-basic and there were significantly reduced luciferase activity compared with control, and mutation of the KLF9 binding site of −146/+120 region showed the lowest luciferase activity (Fig. 6C), suggesting that KLF9 palys a vital role in regulating of *PADI4* transcription. To test whether KLF9 actually binds to the *PADI4* promoter in HCT116 cells, a ChIP assay were performed. The KLF9 binding site was amplified with a pair of primers, which resulted in a band of 130 bp. No bands were amplified when the control IgG was used (Fig. 6D). Collectively, these findings suggested that KLF9 plays a prominent role in the transcription of the *PADI4* gene.

PAD4 promotes GSK3β-mediated nuclear CDKN1A ubiquitination-degradation in CRC cells

To inspect the downstream molecular mechanisms by which PAD4-GSK3β axis contribute to CRC progression, differentially expressed genes (DEGs) were identified in HCT116 cells and PAD4-overexpressing HCT116 cells using RNA-sequencing. Using fold change > 2 and p value < 0.05 as statistical cutoffs, we detected 290 genes that were up-regulated and 152 genes that were down-regulated after PAD4 over-expression in HCT116 cells. The DEGs are presented in a heat map plot and volcano plot (Fig. 7A and S3A) (Table S4). The significantly up-regulated genes were analyzed for pathway enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The top 15 enriched pathways are presented in Fig. 7B, which shows that these genes are involved in several important pathways, such as pathways in cancer (47 DEGs), PI3K-Akt (20 DEGs), MAPK (18 DEGs), and the Hippo signaling pathway (18 DEGs). Among the upregulated genes, *CDKN1A* existed in 11 of the top 15 enriched pathways. And GSK3β can phosphorylate CDKN1A to contribute its degradation [47]. Therefore, CDKN1A was subjected for further analysis.

Western blot showed that there was no change in the total expression of CDKN1A in the PAD4-overexpressing HCT116 cells (Fig. 7C). However, the expression of nuclear CDKN1A was significantly decreased (Fig. 7D). Moreover, silencing GSK3β increased the expression of nuclear CDKN1A in the PAD4-overexpressing HCT116 cells, suggesting that PAD4 decreased the expression of nuclear CDKN1A through GSK3β. To determine whether ubiquitination is involved in nuclear CDKN1A degradation, the nuclear CDKN1A protein was immunoprecipitated using an anti-CDKN1A antibody from the nuclear extract of PAD4-overexpressing HCT116 cells. Then, the ubiquitination level of CDKN1A was analyzed by immunoblotting with an anti-ubiquitin antibody. The western blotting showed that the ectopic expression of PAD4 increased ubiquitination level of nuclear CDKN1A. However, GSK3β-silencing significantly decreased the up-regulated nuclear CDKN1A ubiquitination due to the high expression of PAD4 (Fig. 7E), suggesting that PAD4 degrades CDKN1A through GSK3β in an ubiquitin-dependent proteasome manner in HCT116 cells. Taken together, our findings demonstrated that PAD4 interacts with and citrullinates GSK3β, which increases the nuclear localization of GSK3β, and nuclear GSK3β degrades CDKN1A in a ubiquitin-dependent way to induce the progression of CRC (Fig. 7F).

Discussion

The family of PADs has five highly conserved PAD enzymes (*I-IV and VI*) [7,8] and each exhibits specific tissue distribution and substrate specificities [10]. PAD1 mainly exists in human skin epidermis [48] which is involved in epidermal cornification [49]. PAD3 is primarily localized in the hair follicles [50] and mutations in *PADI3* gene are associated with central centrifugal cicatricial alopecia (CCCA) [51]. PAD6 is preferentially expressed in the adult ovary [52,53]. The absence of the PAD6 protein or mutations in *PADI6* gene contribute to female fertility [54,55]. There are few reports on the relationship between PAD1/PAD3/PAD6 and cancers. PAD2 and PAD4 play vital roles in several cancers. They regulate cancer development through citrullinating histone and regulating genetic transcription [56,57]. In CRC, we had detected the expression of *PADI1, PADI2, PADI3 PADI4* and *PADI6* in 24 pairs of CRC tissues and control tissues. The expression of *PADI1* and *PADI3* are low in CRC and adjacent tissues *PADI1* can only be detected in 15 pairs of 24 pairs of CRC tissues, and *PADI3* can be detected in 12 pairs of 24 pairs of CRC tissues. RT-qPCR results showed that *PADI1* (11 of 15, $p = 0.0241$) and *PADI3* (7 of 12, $p = 0.0296$) were significantly increased in CRC tissues compared with the control (Fig. S5 A and S5 C). Similar results were observed in human CRC samples compared with matched non-CRC tumors from The Cancer Genome Atlas (TCGA) database (Fig. S5 E and F); *PADI6* (18 of 24, $p = 0.0312$) was significantly decreased in CRC tissues compared with the control (Fig. S5D); *PADI2* (18 of 24, $p < 0.0001$) was significantly decreased in CRC tissues compared with the control (Fig. S5B). Furthermore, PAD2 expression was also obvious declined in CRC samples compared with matched controls from TCGA database and the CPTAC database (Fig. S5G and S5H). *PADI4* was significantly higher in 14 of the 24 tumor tissues compared with their corresponding non-cancerous controls (Fig. 1A). Similar results were observed in human CRC samples compared with matched non-CRC tumors from the CPTAC database (Fig. 1G). Collectively, PAD2 and PAD4 are related to cancers, and among which only PAD4 is highly expressed in CRC. So we selected PAD4 as the target in our study.

Up to now, little is known on the role of PAD1, PAD3 or PAD6 in CRC. However, there are several reports about PAD2 in CRC. Funayama et al. [58] reported that PAD2 was decreased in CRC tissues and inhibited the proliferation of CRC cells. Furthermore, Qu et al. [59] reported that small molecule nitazoxanide inhibits Wnt signaling through targeting PAD2 in CRC cells. NTZ increased the activity of PAD2 which leads to the citrullination and degradating of β-catenin. However, Chen et al. [60] reported that lncRNA HOXA11-AS promoted liver metastasis of CRC patients through sponging miR-125a-5p and increasing PAD2 expression.

Increasing amounts of evidences from basic and clinical studies support the crucial roles of PAD4 in the onset and progression of cancers. PAD4 is highly expressed in multiple cancer types, including CRC, lung cancer, pancreatic cancer, osteosarcoma, esophageal squamous cell carcinoma, GC,
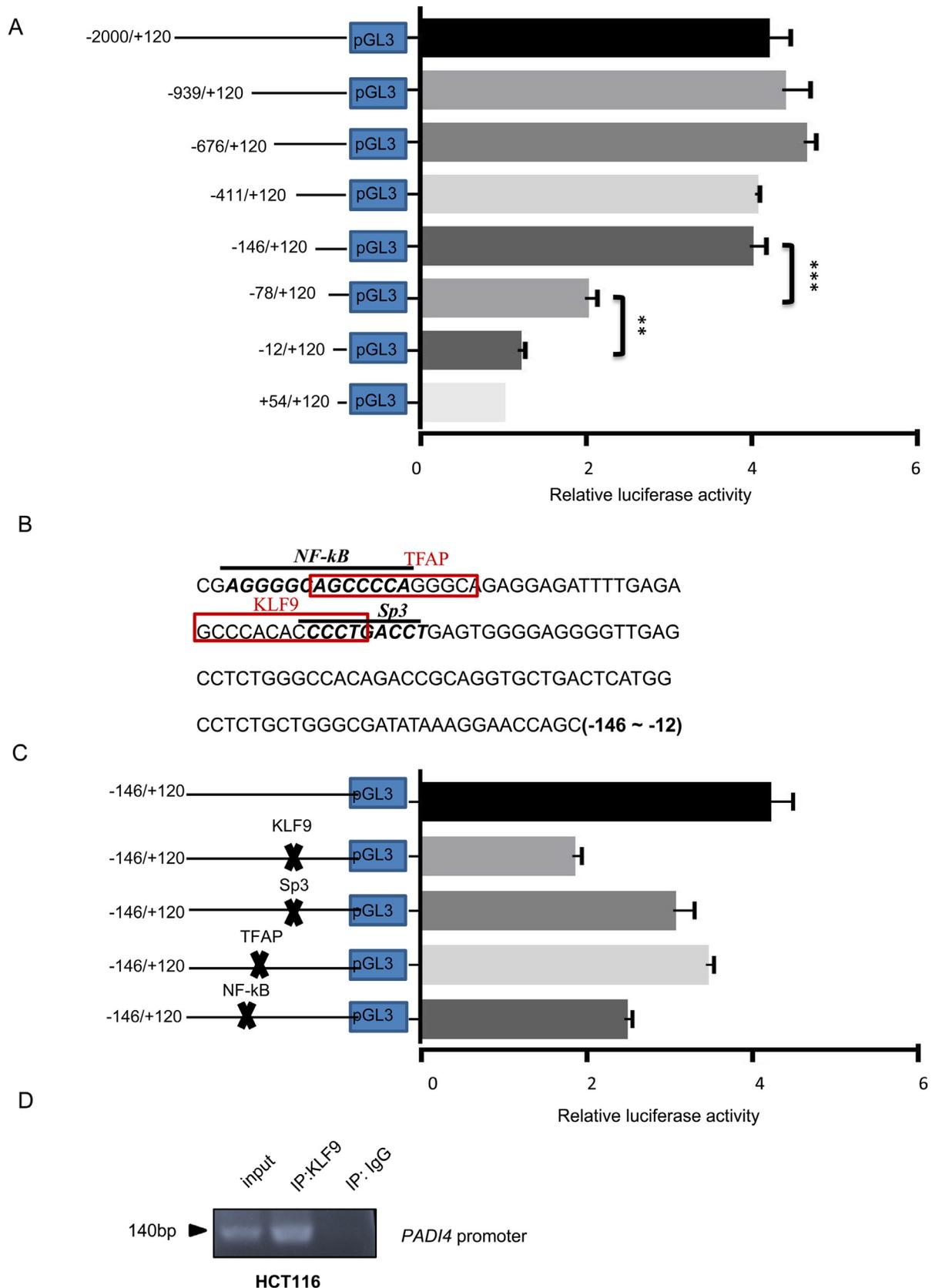
Fig. 6. KLF9 contributes transcription of *PADI4* gene in CRC cells. (A) Identification of minimum promoter sequence of *PADI4*. HCT116 cells were transfected with the indicated constructs and assayed for luciferase activity after 48 h. Luciferase activity is expressed as fold increase over a control vector pGL3-Basic. (B) Putative transcription factor-binding sites of the minimal promoter of *PADI4*. (C) Site-directed mutation was carried out on the putative transcription factor-binding sites spanning the -146/-12 region. (D) ChIP assay using anti-KLF9 antibody or IgG was performed on chromatin from HCT116 cells to identify the association of KLF9 and *PADI4* promoter. All the experiments are conducted in three biological replicates.
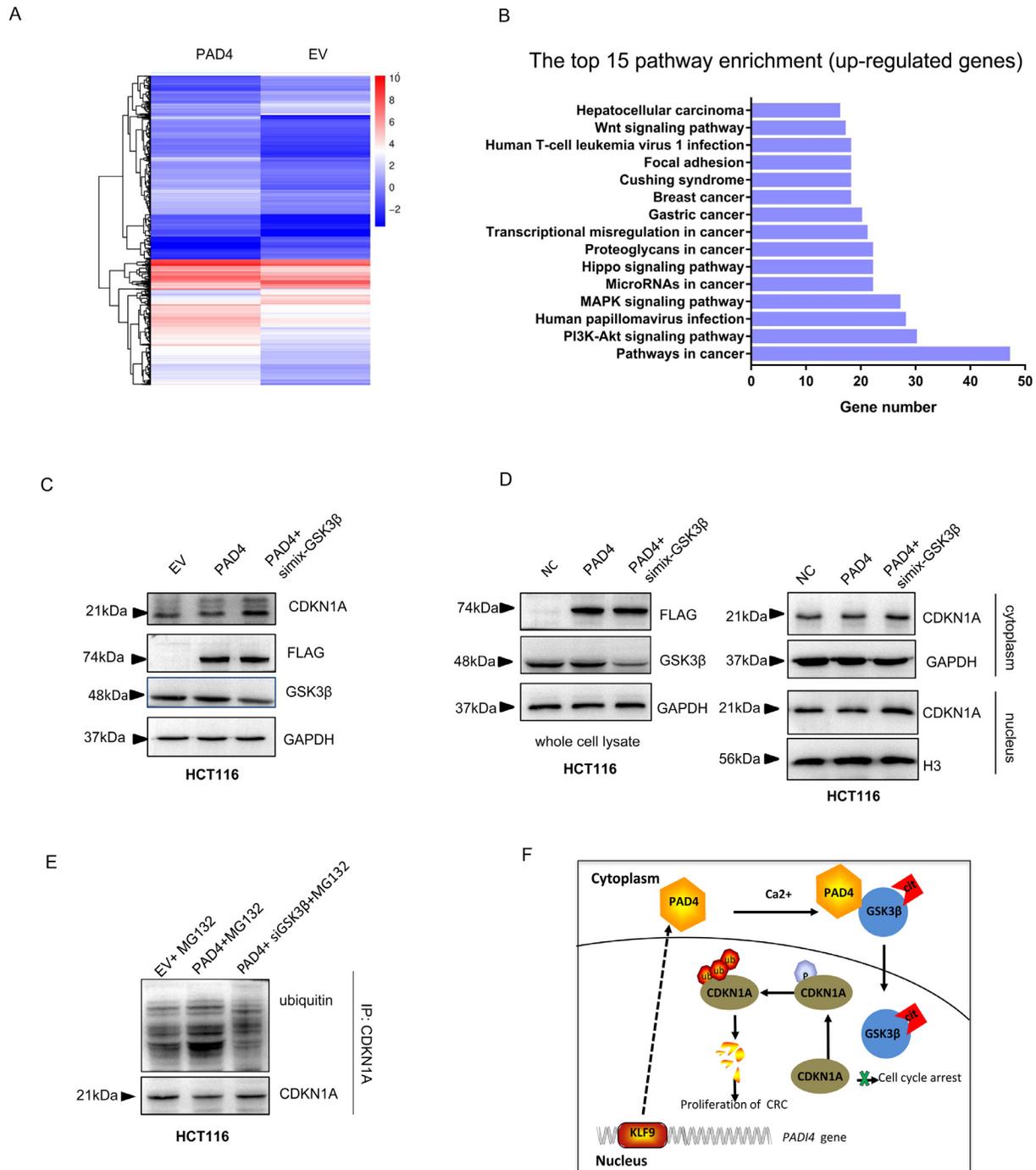
Fig. 7. PAD4-dependent citrullination of GSK3β promotes degradation of CDKN1A in CRC cells. (A) Heat map showing DEGs in HCT116 cells transfected with FLAG-PAD4. (B) KEGG pathway enrichment analysis of up-regulated genes of FLAG-PAD4 transfected HCT116 cells. (C) Western blot analysis of CDKN1A expression in simix-GSK3β and FLAG-PAD4 co-transfected HCT116 cells. (D) Expression of CDKN1A in cytoplasmic and nuclear extracts from simix-GSK3β and FLAG-PAD4 co-transfected HCT116 cells. (E) Ubiquitin levels of CDKN1A from cell nucleus in simix-GSK3β and FLAG-PAD4 co-transfected HCT116 cells. (F) Schematic diagram of PAD4/GSK3β/CDKN1A signaling axis for CRC progression. All the experiments are conducted in three biological replicates.

and nasopharyngeal carcinoma [23-25,61,62]. Moreover, there is also a significant increase in PAD4 expression in the blood of patients with various tumors [63]. PAD4 is also frequently correlated with clinicopathological parameters and the prognosis of various tumors, e.g., higher PAD4 expression corresponds with a poorer prognosis for patients with human pancreatic cancer [64]. Guo et al. reported that PAD4 is increased in osteosarcoma

and upregulated PAD4 is connected with larger tumor size [26]. Zhai et al. identified that higher PAD4 is related to positive pulmonary metastasis in osteosarcoma [30]. In our study, PAD4 expression was significantly higher expressed in CRC tissues and cells. ROC curve showed the area under the ROC curve (AUC) PAD4 was 0.7457. Furthermore, high PAD4 expression in CRC tissue was strongly correlated with increased tumor size and positive

lymphatic metastasis ($p < 0.05$). Taken together, these results demonstrated that there is a possible link between increased expression of PAD4 and cancer progression, and PAD4 could serve as a new diagnosis biomarker and therapeutic target of various cancers.

PAD family members are highly conserved and the conservation of the primary structure is approximately 50% [65], and it is necessary to identify whether over-expression of *PADI4* influence the transcription of other PAD family members. RT-qPCR was carried out, and the result showed that *PADI4* expression was significantly increased in PAD4-overexpressed cells compared with the control HCT116 and SW480 cells. However, expression of *PADI1, PADI2, PADI3* and *PADI6* showed little changes compared with the controls, suggesting that over-expression of *PADI4* did not affect the transcription of *PADI1, PADI2, PADI3*, and *PADI6* in CRC cells HCT116 and SW480 cells (Fig. S5 I and S5J).

PAD4 is involved in tumor development by citrullinating histone substrates and non-histone proteins. Histones H4 and H2A are the most extensively studied target proteins, and both have identical N-terminal MSGR sequences, the arginine of which is citrullinated by PAD4. To explore the potential targets of PAD4 interaction, we obtain all human protein sequences from the UniProt database, and then use the python programming language to search the proteins beginning with MSGR. Among the identified proteins, GSK3β is a serine/threonine protein kinase that plays vital roles in multiple tumor types. PAD4 citrullinated GSK3β and promoted the nuclear location of GSK3β. To detect the location of PAD4 in CRC, confocal assay was conducted in normal colon cells NCM460 and CRC cells. Results indicated that PAD4 was expressed in both the nucleus and cytoplasm, but mainly in the nucleus of cells. There was no significantly difference of PAD4 location in CRC cells compared with normal colon cells NCM460 (Fig. S6).

Aberrant nuclear accumulation of GSK-3β has been identified as a hallmark of cancer cells in malignant tumors, and the NLS motif, N-terminal tail, and tyrosine 216 phosphorylation have been found to be involved in the nuclear localization of GSK3β [11,66]. Sonja et al. [11] demonstrated that PAD4 is an important post-translational modifier of GSK3β that regulates nuclear location of GSK3β by citrullinating its N-terminal domain. Both of us found that PAD4 citrullinates GSK3β and promotes nuclear translocation of GSK3β, which is consistent with each other. However, Sonja reported that PAD4-mediated citrullination of nuclear GSK3β is found to inhibit EMT and suppress the development of breast cancer. Our results demonstrated that PAD4 citrullinates GSK3β at Arg-344 of the C domain to induce GSK3β nuclear translocation and thus promote CRC progression. Therefore, we speculated that the molecular mechanisms of PAD4 are different among different cancers and specific to cancer type.

Accumulating evidence suggests that CDKN1A expression correlates with tumorigenesis and progression [67,68]. CDKN1A is an important regulator of cell cycle checkpoints and a major transcriptional target of the p53 protein. In general, CDKN1A binds to and inhibits the activity of cyclin dependent kinases Cdk1 and Cdk2 and blocks the cell cycle after DNA damage [69]. Several studies have shown that GSK3β phosphorylates CDKN1A and induces CDKN1A degradation of urothelial carcinoma cells via an ubiquitination-dependent pathway [47,70], which was consistent with ours.

In conclusion, the up-regulation of PAD4 in CRC is a potential indicator of aggressive CRC phenotypes. PAD4 down-regulates CDKN1A expression through the citrullination of GSK3β C domain to induce proliferation and migration of CRC cells. Citrullination of the C terminus is a completely new standpoint in regulation of the location of GSK3β, and we expect future studies committed to explore the function of citrullination of GSK3β in CRC. Moreover, this is the first study to reveal the important role of the PAD4–GSK3β–CDKN1A axis in CRC, which provides a potential therapeutic target in CRC. However, further investigation is needed to search for other regulatory machineries of PAD4 in CRC progression.

Authors' contributions

XL, XZ, and JZ conceived the concept for the study. FH, YP, SC, SX, YG, JL, YX, KD, YC, JQ and XF conducted the study. XF and GH asked for written informed consent and collected CRC samples from patients. SD, XF, JL and XZ performed the sequencing data analysis and statistical analysis. XL, XZ, and ZJ wrote the manuscript. XZ, XL, MS, and ZJ reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

All human tissues were collected from the Department of General Surgery at the First Affiliated Hospital of Shenzhen University with written informed consent. The study was approved by the Clinical Research Ethics Committee.

Consent for publication

Not applicable.

Data availability

Other data that support the findings of this study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The authors declare no potential conflict of interest.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2022.100835.

References

[1] Alhinai EA, Walton GE, Commane DM. The role of the gut microbiota in colorectal cancer causation. *Int J Mol Sci* 2019;**20**.

[2] Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int J Mol Sci* 2017;**18**.

[3] Müller MF, Ibrahim AE, Arends MJ. Molecular pathological classification of colorectal cancer. *Virchows Archiv* 2016;**469**:125–34.

[4] Zhang X, Peng Y, Yuan Y, Gao Y, Hu F, Wang J, Zhu X, Feng X, Cheng Y, Wei Y, et al. Histone methyltransferase SET8 is regulated by miR-192/215 and induces oncogene-induced senescence via p53-dependent DNA damage in human gastric carcinoma cells. *Cell Death Dis* 2020;**11**:937.

[5] Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW. Colorectal cancer tumour markers and biomarkers: recent therapeutic advances. *World J Gastroenterol* 2016;**22**:1745–55.

[6] Sagaert X, Vanstapel A, Verbeek S. Tumor Heterogeneity in colorectal cancer: what do we know so far? *Pathobiology* 2018;**85**:72–84.

[7] Chang X, Fang K. PADI4 and tumourigenesis. *Cancer Cell Int* 2010;**10**:7.

[8] Zhai Q, Wang L, Zhao P, Li T. Role of citrullination modification catalyzed by peptidylarginine deiminase 4 in gene transcriptional regulation. *Acta Biochim Biophy Sin* 2017;**49**:567–72.

[9] Wang S, Wang Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim Biophys Acta* 2013;**1829**:1126–35.

[10] Wang Y, Chen R, Gan Y, Ying S. The roles of PAD2- and PAD4-mediated protein citrullination catalysis in cancers. *Int J Cancer* 2021;**148**:267–76.

[11] Stadler SC, Vincent CT, Fedorov VD, Patsialou A, Cherrington BD, Wakshlag JJ, Mohanan S, Zee BM, Zhang X, Garcia BA, et al. Dysregulation of PAD4-mediated citrullination of nuclear GSK3β activates TGF-β signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proc Nat Acad Sci USA* 2013;**110**:11851–6.

[12] Mondal S, Gong X, Zhang X, Salinger AJ, Zheng L, Sen S, Weerapana E, Zhang X, Thompson PR. Halogen bonding increases the potency and isozyme selectivity of protein arginine deiminase 1 inhibitors. *Angew Chem Int Ed Engl* 2019;**58**:12476–80.

[13] Cherrington BD, Morency E, Struble AM, Coonrod SA, Wakshlag JJ. Potential role for peptidylarginine deiminase 2 (PAD2) in citrullination of canine mammary epithelial cell histones. *PLoS One* 2010;**5**:e11768.

[14] Li G, Hayward IN, Jenkins BR, Rothfuss HM, Young CH, Nevalainen MT, Muth A, Thompson PR, Navratil AM, Cherrington BD. Peptidylarginine deiminase 3 (PAD3) is upregulated by prolactin stimulation of CID-9 cells and expressed in the lactating mouse mammary. *Gland PLoS one* 2016;**11**:e0147503.

[15] Huang B, Zhao Y, Zhou L, Gong T, Feng J, Han P, Qian J. PADI6 regulates trophoblast cell migration-invasion through the hippo/YAP1 pathway in hydatidiform moles. *J Inflamm Res* 2021;**14**:3489–500.

[16] Mondal S, Thompson PR. Protein Arginine Deiminases (PADs): biochemistry and chemical biology of protein citrullination. *Acc Chem Res* 2019;**52**:818–32.

[17] Zhu D, Zhang Y, Wang S. Histone citrullination: a new target for tumors. *Mol Cancer* 2021;**20**:90.

[18] Arita K, Hashimoto H, Shimizu T, Nakashima K, Yamada M, Sato M. Structural basis for Ca(2+)-induced activation of human PAD4. *Nat Struct Mol Biol* 2004;**11**:777–83.

[19] Lewis HD, Liddle J, Coote JE, Atkinson SJ, Barker MD, Bax BD, Bicker KL, Bingham RP, Campbell M, Chen YH, et al. Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nat Chem Biol* 2015;**11**:189–91.

[20] Guo Q, Fast W. Citrullination of inhibitor of growth 4 (ING4) by peptidylarginine deminase 4 (PAD4) disrupts the interaction between ING4 and p53. *J Biol Chem* 2011;**286**:17069–78.

[21] Lee YH, Coonrod SA, Kraus WL, Jelinek MA, Stallcup MR. Regulation of coactivator complex assembly and function by protein arginine methylation and demethylimination. *Proc Nat Acad Sci USA* 2005;**102**:3611–16.

[22] Tanikawa C, Ueda K, Nakagawa H, Yoshida N, Nakamura Y, Matsuda K. Regulation of protein Citrullination through p53/PADI4 network in DNA damage response. *Cancer Res* 2009;**69**:8761–9.

[23] Chang XT, Wu H, Li HL, Li HL, Zheng YB. PADI4 promotes epithelial-mesenchymal transition(EMT) in gastric cancer via the upregulation of interleukin 8. *BMC Gastroenterol* 2022;**22**:25.

[24] Cui YY, Yan L, Zhou J, Zhao S, Zheng YB, Sun BH, Lv HT, Rong FN, Chang XT. The role of peptidylarginine deiminase 4 in ovarian cancer cell tumorigenesis and invasion. *Tumour Biol* 2016;**37**:5375–83.

[25] Fan T, Zhang C, Zong M, Zhao Q, Yang X, Hao C, Zhang H, Yu S, Guo J, Gong R, et al. Peptidylarginine deiminase IV promotes the development of chemoresistance through inducing autophagy in hepatocellular carcinoma. *Cell Biosci* 2014;**4**:49.

[26] Guo J, Yin L, Zhang X, Su P, Zhai Q. Factors associated with promoted proliferation of osteosarcoma by peptidylarginine deiminase 4. *Biomed Res Int* 2021;**2021**:5596014.

[27] Liu C, Tang J, Li C, Pu G, Yang D, Chang X. PADI4 stimulates esophageal squamous cell carcinoma tumor growth and up-regulates CA9 expression. *Mol Carcinog* 2019;**58**:66–75.

[28] Liu M, Qu Y, Teng X, Xing Y, Li D, Li C, Cai L. PADI4–mediated epithelial–mesenchymal transition in lung cancer cells. *Mol Med Rep* 2019;**19**:3087–94.

[29] Wei LN, Luo M, Wang XP, Liang T, Huang CJ, Chen H. PADI4, negatively regulated by miR-335-5p, participates in regulating the proliferation, migration, invasion and radiosensitivity of nasopharyngeal carcinoma cells. *J Biol Regul Homeost Agents* 2021;**35**:117–29.

[30] Zhai Q, Qin J, Jin X, Sun X, Wang L, Du W, Li T, Xiang X. PADI4 modulates the invasion and migration of osteosarcoma cells by down-regulation of epithelial-mesenchymal transition. *Life Sci* 2020;**256**:117968.

[31] Yuzhalin AE, Gordon-Weeks AN, Tognoli ML, Jones K, Markelc B, Konietzny R, Fischer R, Muth A, O'Neill E, Thompson PR, et al. Colorectal cancer liver metastatic growth depends on PAD4-driven citrullination of the extracellular matrix. *Nat Commun* 2018;**9**:4783.

[32] Li P, Yao H, Zhang Z, Li M, Luo Y, Thompson PR, Gilmour DS, Wang Y. Regulation of p53 target gene expression by peptidylarginine deiminase 4. *Mol Cell Biol* 2008;**28**:4745–58.

[33] Saha P, Yeoh BS, Xiao X, Golonka RM, Singh V, Wang Y, Vijay-Kumar M. PAD4-dependent NETs generation are indispensable for intestinal clearance of Citrobacter rodentium. *Mucosal Immunol* 2019;**12**:761–71.

[34] Shen Y, You Q, Wu Y, Wu J. Inhibition of PAD4-mediated NET formation by cl-amidine prevents diabetes development in nonobese diabetic mice. *Eur J Pharmacol* 2022;**916**:174623.

[35] Biron BM, Chung CS, O'Brien XM, Chen Y, Reichner JS, Ayala A. Cl-amidine prevents histone 3 citrullination and neutrophil extracellular trap formation, and improves survival in a murine sepsis model. *J Innate Immunity* 2017;**9**:22–32.

[36] Willis VC, Banda NK, Cordova KN, Chandra PE, Robinson WH, Cooper DC, Lugo D, Mehta G, Taylor S, Tak PP, et al. Protein arginine deiminase 4 inhibition is sufficient for the amelioration of collagen-induced arthritis. *Clin Exp Immunol* 2017;**188**:263–74.

[37] Chen H, Luo M, Wang X, Liang T, Huang C, Huang C, Wei L. Inhibition of PAD4 enhances radiosensitivity and inhibits aggressive phenotypes of nasopharyngeal carcinoma cells. *Cell Mol Biol Lett* 2021;**26**:9.

[38] Lee W, Ko SY, Mohamed MS, Kenny HA, Lengyel E, Naora H. Neutrophils facilitate ovarian cancer premetastatic niche formation in the omentum. *J Exp Med* 2019;**216**:176–94.

[39] Huang C, Wen B. Phenotype transformation of immortalized NCM460 colon epithelial cell line by TGF-β1 is associated with chromosome instability. *Mol Biol Rep* 2016;**43**:1069–78.

[40] Vecchiotti G, Colafarina S, Aloisi M, Zarivi O, Di Carlo P, Poma A. Genotoxicity and oxidative stress induction by polystyrene nanoparticles in the colorectal cancer cell line HCT116. *PLoS One* 2021;**16**:e0255120.

[41] Cao J, Wu X, Qin X, Li Z. Uncovering the effect of passage number on HT29 cell line based on the cell metabolomic approach. *J Proteome Res* 2021;**20**:1582–90.

[42] Katayama M, Nakano H, Ishiuchi A, Wu W, Oshima R, Sakurai J, Nishikawa H, Yamaguchi S, Otsubo T. Protein pattern difference in the colon cancer cell lines examined by two-dimensional differential in-gel electrophoresis and mass spectrometry. *Surg Today* 2006;**36**:1085–93.

[43] Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, Sonbuchner LS, McDonald CH, Cook RG, Dou Y, et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* 2004;**306**:279–83.

[44] Zhang X, Peng Y, Huang Y, Yang M, Yan R, Zhao Y, Cheng Y, Liu X, Deng S, Feng X, et al. SMG-1 inhibition by miR-192/-215 causes epithelial-mesenchymal transition in gastric carcinogenesis via activation of Wnt signaling. *Cancer Med* 2018;**7**:146–56.

[45] Zhang X, Peng Y, Huang Y, Deng S, Feng X, Hou G, Lin H, Wang J, Yan R, Zhao Y, et al. Inhibition of the miR-192/215-Rab11-FIP2 axis suppresses human gastric cancer progression. *Cell Death Dis* 2018;**9**:778.

[46] Yuzhalin AE. Citrullination in cancer. *Cancer Res* 2019;**79**:1274–84.

[47] Lee JY, Yu SJ, Park YG, Kim J, Sohn J. Glycogen synthase kinase 3beta phosphorylates p21WAF1/CIP1 for proteasomal degradation after UV irradiation. *Mol Cell Biol* 2007;**27**:3187–98.

[48] Guerrin M, Ishigami A, Méchin MC, Nachat R, Valmary S, Sebbag M, Simon M, Senshu T, Serre G. cDNA cloning, gene organization and expression analysis of human peptidylarginine deiminase type I. *Biochem J* 2003;**370**:167–74.

[49] Qin H, Liu X, Li F, Miao L, Li T, Xu B, An X, Muth A, Thompson PR, Coonrod SA, et al. PAD1 promotes epithelial-mesenchymal transition and metastasis in triple-negative breast cancer cells by regulating MEK1-ERK1/2-MMP2 signaling. *Cancer Lett* 2017;**409**:30–41.

[50] Kanno T, Kawada A, Yamanouchi J, Yosida-Noro C, Yoshiki A, Shiraiwa M, Kusakabe M, Manabe M, Tezuka T, Takahara H. Human peptidylarginine deiminase type III: molecular cloning and nucleotide sequence of the cDNA, properties of the recombinant enzyme, and immunohistochemical localization in human skin. *J Invest Dermatol* 2000;**115**:813–23.

[51] Malki L, Sarig O, Romano MT, Méchin MC, Peled A, Pavlovsky M, Warshauer E, Samuelov L, Uwakwe L, Briskin V, et al. Variant PADI3 in central centrifugal cicatricial alopecia. *N Engl J Med* 2019;**380**:833–41.

[52] Zhang J, Dai J, Zhao E, Lin Y, Zeng L, Chen J, Zheng H, Wang Y, Li X, Ying K, et al. cDNA cloning, gene organization and expression analysis of human peptidylarginine deiminase type VI. *Acta Biochim Pol* 2004;**51**:1051–8.

[53] Choi M, Lee OH, Jeon S, Park M, Lee DR, Ko JJ, Yoon TK, Rajkovic A, Choi Y. The oocyte-specific transcription factor, Nobox, regulates the expression of Pad6, a peptidylarginine deiminase in the oocyte. *FEBS Lett* 2010;**584**:3629–34.

[54] Xu Y, Shi Y, Fu J, Yu M, Feng R, Sang Q, Liang B, Chen B, Qu R, Li B, et al. Mutations in PADI6 cause female infertility characterized by early embryonic arrest. *Am J Hum Genet* 2016;**99**:744–52.

[55] Esposito G, Vitale AM, Leijten FP, Strik AM, Koonen-Reemst AM, Yurttas P, Robben TJ, Coonrod S, Gossen JA. Peptidylarginine deiminase (PAD) 6 is essential for oocyte cytoskeletal sheet formation and female fertility. *Mol Cell Endocrinol* 2007;**273**:25–31.

[56] Cherrington BD, Zhang X, McElwee JL, Morency E, Anguish LJ, Coonrod SA. Potential role for PAD2 in gene regulation in breast cancer cells. *PLoS One* 2012;**7**:e41242.

[57] Anzilotti C, Pratesi F, Tommasi C, Migliorini P. Peptidylarginine deiminase 4 and citrullination in health and disease. *Autoimmun Rev* 2010;**9**:158–60.

[58] Funayama R, Taniguchi H, Mizuma M, Fujishima F, Kobayashi M, Ohnuma S, Unno M, Nakayama K. Protein-arginine deiminase 2 suppresses proliferation of colon cancer cells through protein citrullination. *Cancer Sci* 2017;**108**:713–18.

[59] Qu Y, Olsen JR, Yuan X, Cheng PF, Levesque MP, Brokstad KA, Hoffman PS, Oyan AM, Zhang W, Kalland KH, et al. Small molecule promotes β-catenin citrullination and inhibits Wnt signaling in cancer. *Nat Chem Biol* 2018;**14**:94–101.

[60] Chen D, Sun Q, Zhang L, Zhou X, Cheng X, Zhou D, Ye F, Lin J, Wang W. The lncRNA HOXA11-AS functions as a competing endogenous RNA to regulate PADI2 expression by sponging miR-125a-5p in liver metastasis of colorectal cancer. *Oncotarget* 2017;**8**:70642–52.

[61] Chang X, Han J. Expression of peptidylarginine deiminase type 4 (PAD4) in various tumors. *Mol Carcinog* 2006;**45**:183–96.

[62] Duan Q, Pang C, Chang N, Zhang J, Liu W. Overexpression of PAD4 suppresses drug resistance of NSCLC cell lines to gefitinib through inhibiting Elk1-mediated epithelial-mesenchymal transition. *Oncol Rep* 2016;**36**:551–8.

[63] Chang X, Han J, Pang L, Zhao Y, Yang Y, Shen Z. Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer* 2009;**9**:40.

[64] Zhang Y, Chandra V, Riquelme Sanchez E, Dutta P, Quesada PR, Rakoski A, Zoltan M, Arora N, Baydogan S, Horne W, et al. Interleukin-17-induced neutrophil extracellular traps mediate resistance to checkpoint blockade in pancreatic cancer. *J Exp Med* 2020;**217**.

[65] Vossenaar ER, Zendman AJ, van Venrooij WJ, Pruijn GJ. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *BioEssays* 2003;**25**:1106–18.

[66] Meares GP, Jope RS. Resolution of the nuclear localization mechanism of glycogen synthase kinase-3: functional effects in apoptosis. *J Biol Chem* 2007;**282**:16989–7001.

[67] Lin J, Song T, Li C, Mao W. GSK-3β in DNA repair, apoptosis, and resistance of chemotherapy, radiotherapy of cancer. *Biochim Biophys Acta Mol Cell Res* 2020;**1867**:118659.

[68] Xiao BD, Zhao YJ, Jia XY, Wu J, Wang YG, Huang F. Multifaceted p21 in carcinogenesis, stemness of tumor and tumor therapy. *World J Stem Cells* 2020;**12**:481–7.

[69] El-Deiry WS. p21(WAF1) mediates cell-cycle inhibition, relevant to cancer suppression and therapy. *Cancer Res* 2016;**76**:5189–91.

[70] Yohn NL, Bingaman CN, DuMont AL, Yoo LI. Phosphatidylinositol 3′-kinase, mTOR, and glycogen synthase kinase-3β mediated regulation of p21 in human urothelial carcinoma cells. *BMC Urol* 2011;**11**:19.