EBioMedicine
Published by THE LANCET

Research paper

Histone-lysine N-methyltransferase SETD7 is a potential serum biomarker for colorectal cancer patients

Baojun Duan [a,b,1], Jun Bai [b,1], Jian Qiu [c,1], Jianhua Wang [c], Cong Tong [c], Xiaofei Wang [a], Jiyu Miao [a], Zongfang Li [d], Wensheng Li [e], Juan Yang [a,*], Chen Huang [a,*]

[a] Key Laboratory of Environment and Disease-Related Gene, Ministry of Education, Department of Cell Biology and Genetics, School of Basic Medical sciences, Xi'an Jiaotong University, Health Science Center, Shaanxi, Xi'an 710061, China
[b] Department of Medical Oncology of Shaanxi Provincial People's Hospital, Shaanxi, Xi'an 710068, China
[c] Department of General Surgery of Shaanxi Provincial People's Hospital, Shaanxi, Xi'an 710068, China
[d] National & Local Joint Engineering Research Center of Biodiagnostics and Biotherapy, the Second Affiliated Hospital of Xi'an Jiaotong University, Shaanxi, Xi'an 710004, China
[e] Department of Pathology of Shaanxi Provincial People's Hospital, Shaanxi, Xi'an 710068, China

ARTICLE INFO

ABSTRACT

Background: There is an urgent need for the identification of new, clinically useful biomarkers of CRC to enhance diagnostic and prognostic capabilities.
Methods: We performed proteomic profiling on serum samples from paired pre- and post-operative CRC patients, colorectal polyps patients and healthy controls using an approach combining magnetic bead-based weak cation exchange and matrix-assisted laser desorption ionization-time of flight mass spectrometry. We next performed liquid chromatography-electrospray ionization-tandem mass spectrometry to identify the proteins and selected potential biomarker based on bioinformatics analysis of the TCGA and GEO dataset. We examined SETD7 expression in serum and tissue samples by ELISA and immunohistochemistry respectively and explored the biological function of SETD7 in vitro.
Findings: 85 differentially expressed peptides were identified. Five peptides showing the most significant changes in abundance across paired pre- and post-operation CRC patients, colorectal polyps patients and healthy controls were identified as peptide regions of FGA, MUC5AC and SETD7. Bioinformatics analysis suggested that the up-regulation of SETD7 in CRC is relatively specific. Validation studies showed that SETD7 expression increased from healthy controls to those with colorectal polyps and finally CRC patients, and decreased after surgery. The sensitivity and specificity of SETD7 were 92.17% and 81.08%, with a high diagnostic value (AUC = 0.9477). In addition, SETD7 expression was significantly correlated with tumor stage and microsatellite instability. Knock-down of SETD7 inhibited cancer cell proliferation, induced G1/S cell cycle arrest and increased apoptosis.
Interpretation: Our data indicate that SETD7 could serve as a potential diagnostic and prognostic biomarker for CRC.

1. Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer and the forth most common cause of cancer-related deaths worldwide [1]. Surgical resection in early disease stages results in a five-year survival rate of >90%, while patients with advanced disease have a five-year survival rate of <10% [2]. Unfortunately, even with surgical resection, there is a high rate of recurrence. Therefore, early diagnosis and monitoring of the efficacy of anti-cancer therapy is important to reduce CRC-associated mortality.

Currently, methods available for the early diagnosis of CRC and monitoring of treatment efficacy include the fecal occult blood test (FOBT),

computed tomography (CT), nuclear magnetic resonance (NMR), endoscopy, and serum carcinoembryonic antigen (CEA). However, the sensitivity and specificity of FOBT, CT and NMR are low, and compliance with endoscopy is poor due to cost and invasiveness. Blood screening is cheaper and easier, however, CEA levels have been reported to fluctuate significantly in healthy populations, and even within the same individual variation can reach 30%, so the value of CEA in screening asymptomatic populations and diagnosing CRC early is controversial [3]. Therefore,

there is an urgent need for the identification of new, clinically useful biomarkers of CRC to enhance current diagnostic and prognostic capabilities.

Mass spectrometry (MS) combined with software-generated models can be used to compare the proteomic profiles of cancer patients and healthy individuals quickly and accurately [4–6]. In the past few years, there have been many reports of the development of diagnosis or prognosis models and screening of new peptide biomarkers of CRC,
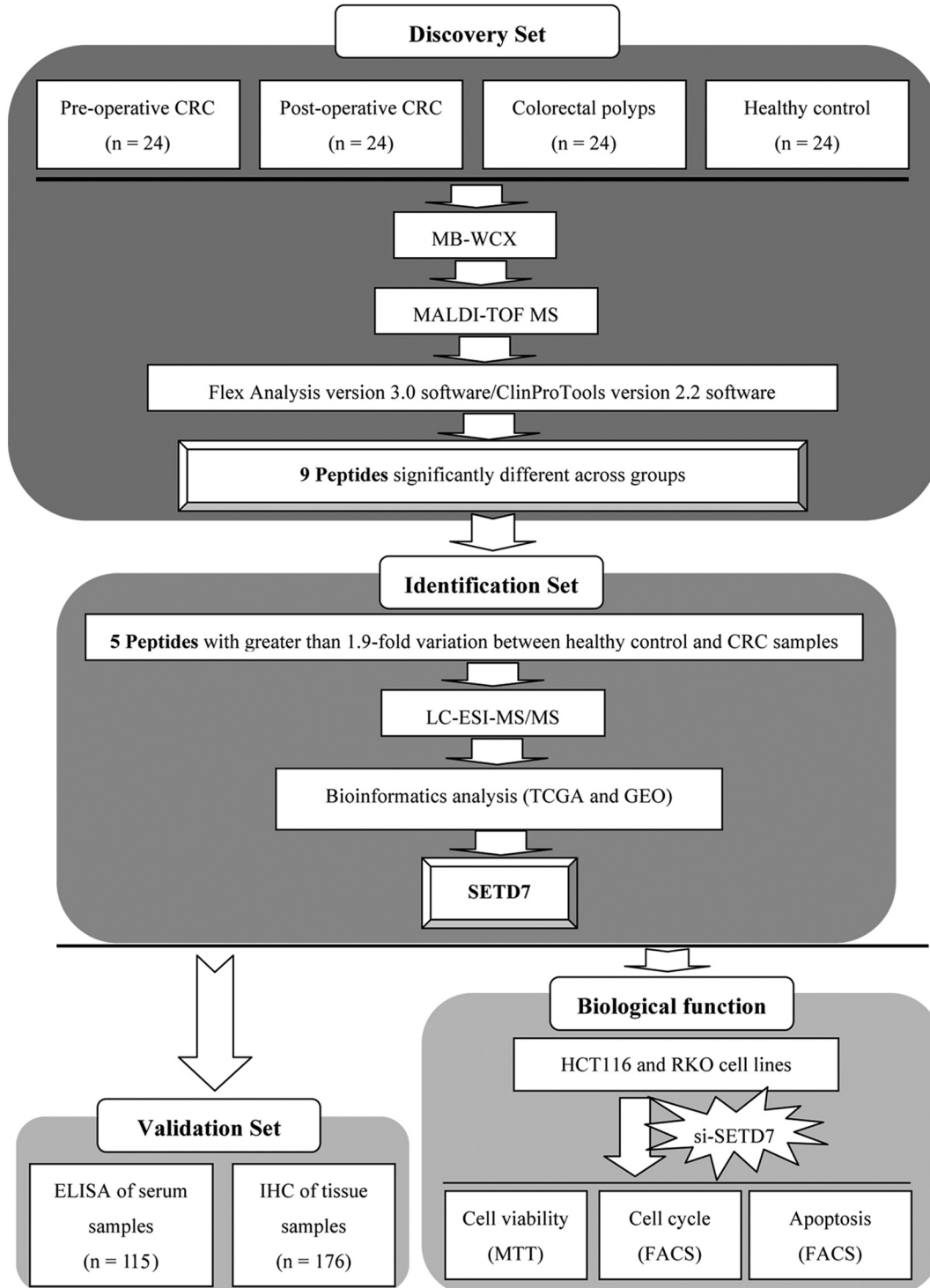
Summary of the workflow

Fig. 1. Workflow used for the discovery, identification, validation and functional analysis of a potential biomarker for colorectal cancer (CRC).

however, these models and biomarkers have not been applied to clinical practice due to variation in sensitivity and specificity [7–11].

In the present study, we used magnetic bead-based weak cation exchange (MB-WCX) purification combined with matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS to conduct proteomic analysis of samples from pre- and post-operative CRC patients, patients with colorectal polyps and healthy controls. Using a bioinformatics approach, we identified SETD7 as a potential diagnostic and prognostic biomarker for CRC. We examined expression of SETD7 in serum and tissue samples and explored the biological function of SETD7 in vitro.

2. Materials and methods

2.1. Ethics committee approval

This study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the Ethics Committee and the Human Research Review Committee of Shaanxi Provincial People's Hospital (No. 20140406). Informed consent was obtained from all of the participants prior to the start of the study. All the applied methods in the study were carried out in accordance with the approved guidelines.

2.2. Patients and sample collection

All samples were collected at the Shaanxi Provincial People's Hospital between September 1, 2014 and September 1, 2016. None of the CRC patients included in this study had any other tumors or inflammatory colorectal disease such as colitis or infectious diseases, and none of the patients received chemotherapy or radiation therapy prior to surgery. Pathological diagnosis was conducted by two different pathologists, and the clinicopathological stage of CRC was determined according to the criteria of the American Joint Committee on Cancer (7th edition).

48 serum samples were obtained from 24 paired pre- and post-operative CRC patients (mean age 65.83 ± 9.14; 54.2% male, 45.8% female), with their post-operative serum samples collected three days after surgery (Supplementary Table S1). 24 serum samples were obtained from patients with colorectal polyps (polyps have not been excised, mean age 60.22 ± 13.61; 58.3% male, 41.7% female) representing benign disease. 24 serum samples were obtained from healthy controls (mean age 63.77 ± 11.83; 50.0% male, 50.0% female). All serum samples were collected in 3 mL vacuum tubes without anticoagulants and kept at 4 °C for 1 h. Samples were then centrifuged at 3000 $\times g$ for 20 min at 4 °C, distributed in 500 μL aliquots and stored at −80 °C until use.

For the enzyme-linked immunosorbent assay (ELISA), serum samples were collected from 115 paired pre- and post-operative CRC patients (mean age 68.54 ± 7.01; 56.5% male, 43.5% female), 38 patients with colorectal polyps (mean age 60.03 ± 5.65; 55.3% male, 44.7% female), and 38 healthy controls (mean age 58.36 ± 10.54; 50.0% male, 50.0% female). Samples were processed and stored as described above.

For immunohistochemistry (IHC) analysis, tissue specimens were obtained from 176 CRC patients (mean age 67.22 ± 8.25; 58.5% male, 41.5%

female), and 20 patients with colorectal polyps (mean age 61.16 ± 4.88; 60.0% male, 40.0% female). We obtained para-cancerous tissue from 20 CRC patients (mean age 65.55 ± 11.09; 50.0% male, 50.0% female). A portion of each tissue specimen was frozen in liquid nitrogen immediately, then stored at −80 °C until use, while the remaining tissue was formalin-fixed and paraffin-embedded, and 4 μm slices prepared.

2.3. Cell lines and culture conditions

Two human colon cancer-derived cell lines (HCT116 and RKO) were included in this study. Authentication of human cell lines was achieved using short tandem repeats (STR, see Supplementary file). Cells were grown in RPMI-1640 medium (Thermo Fisher Scientific, USA) supplemented with 10% fetal bovine serum (Biological Industries, Israel) and 10 mg/mL penicillin streptomycin (1%), and incubated at 37 °C in a humidified atmosphere of 5% CO_2.

2.4. MS analysis

Serum samples were separated using MB-WCX (ClinProt purification reagent sets; Bruker Daltonics, Bremen, Germany) and a magnetic separator, with the magnet lowered, 5 μL serum samples were diluted in 10 μL binding solution in a standard thin wall PCR tube, added to 10 μL of MB-WCX beads and then carefully mixed using the mixing feature of the robot. After thorough stirring, samples were incubated at room temperature for 5 min, then the tubes were placed into the magnetic separator to collect the beads on the wall of the tube until the supernatant was clear (~1 min). The supernatant was then removed and the magnet was lowered again. Following the stepwise application of sample and MB-WCX separation, we eluted the peptide fraction from the magnetic beads with 5 μL of elution solution and 5 μL of stabilization buffer [6]. Eluted peptides were spotted onto the MALDI AnchorChip with 1 μL alpha-cyano-4-hydroxycinnamic acid (Bruker Daltonics) in 50% acetonitrile, and 0.5% trifluoroacetic acid was added twice to the MALDI AnchorChip surface. Each sample was spotted in triplicate in order to evaluate reproducibility.

2.5. MS data analysis

All samples were processed immediately on a calibrated Autoflex III MALDI-TOF MS (Bruker Daltonics), using flexControl version 3.0 software (Bruker Daltonics) with an optimized measuring protocol. A standard calibration mixture of peptides and proteins (mass range: 1–10 kDa) was used for mass calibration. Tests were performed in a blinded manner, including serum analysis of different groups. Data analysis was performed by flexAnalysis version 3.0 software (Bruker Daltonics). Recognition of peptide patterns was performed using ClinProTools version 2.2 software (Bruker Daltonics). Data was processed by a standard workflow, which comprised of spectral pretreatment, peak selection, and peak calculation operations [6].

Table 1
Peaks showing significant differences in abundance across samples from healthy controls, patients with colorectal polyps, pre- and post-operative CRC patients.

Peak	m/z	P value	Healthy controls (n = 24)	Colorectal polyps (n = 24)	Pre-operative (n = 24)	Post-operative (n = 24)
1	2676.04	<0.000001	2.9 ± 0.88	5.32 ± 3.22	9.33 ± 4.15	8.05 ± 5.83
2	2556.90	<0.000001	1.68 ± 0.38	2.33 ± 1.1	2.52 ± 0.76	2.38 ± 1.14
3	2691.34	<0.000001	2.11 ± 0.49	2.68 ± 1.00	3.07 ± 0.93	2.9 ± 1.11
4	5381.05	<0.000001	1.1 ± 0.25	1.44 ± 0.58	1.47 ± 0.52	1.26 ± 0.43
5	5343.69	<0.000001	1.59 ± 0.47	4.41 ± 3.72	7.26 ± 6.09	4.4 ± 2.73
6	4793.31	<0.000001	1.2 ± 0.23	1.97 ± 0.9	2.3 ± 1.23	2.09 ± 1.2
7	3446.91	<0.000001	5.45 ± 3.31	2.8 ± 1.77	2.52 ± 1.13	4.1 ± 2.36
8	3490.67	<0.000001	3.28 ± 1.75	1.97 ± 1.35	1.82 ± 0.65	2.99 ± 1.94
9	2663.37	<0.000001	6.92 ± 4.6	5.69 ± 4.08	3.22 ± 1.76	3.86 ± 3.2

Table 2
Amino acids sequences and identified proteins from the five peaks showing different levels of abundance between healthy controls, patients with colorectal polyps, pre- and post-operative CRC patients.

Peak	m/z	P value	Uniprot ID	Peptide sequence	Identified protein
1	2676.04	<0.000001	TrEMBL A0A0S2Z3M3	K.SYKMADEAGSEADHEGTHSTKRGHA.K	FGA Isoform 1 of fibrinogen alpha chain precursor
5	5343.69	<0.000001	TrEMBL A7Y9J9	G.AAYEDFNIQLRRSQESAAPTLSRVLM*KVDGVVIQLTKGSVLVNGHPVLL.P	MUC5AC Mucin-5 AC precursor (Fragment)
6	4793.31	<0.000001	TrEMBL B5MCZ8	Y.DMFVHPRFGPIKCIRTLRAVEADEELTVAYGYDHSPPGKSGPE.A	SETD7 Histone-lysine *N*-methyltransferase SETD7
7	3446.91	<0.000001	TrEMBL A0A0S2Z3M	R.GKSSSYSKQFTSSTSYNRGDSTFESKSYKMA.D	FGA Isoform 1 of fibrinogen alpha chain precursor
9	2663.37	<0.000001	TrEMBL A0A0S2Z3M	A.DEAGSEADHEGTHSTKRGHAKSRPV.R	FGA Isoform 1 of fibrinogen alpha chain precursor

2.6. Peptide identification

The peptides from plasma samples (the differential peptides are relatively abundant) were eluted from the magnetic beads and were analyzed by nano-UPLC-ESI-MS/MS using a nano Aquity UPLC (Waters Corporation, Milford, USA) coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Samples of 20 μL (the sample was diluted by 2 times) were loaded on a C18 precolumn (Symmetry®C18, 5 μm, 180 μm × 20 mm, nanoAcquity™Column) at 15 μL/min in 5% acetonitrile (Sigma-Aldrich, St Louis, MO, USA), 0.05% trifluoroacetic acid (Sigma-Aldrich) for 3 min. The precolumn was switched online with the analytical column (Symmetry®C18, 3.5 μm, 75 μm × 150 mm, nanoAcquity™ Column) equilibrated in 95% solvent A (5% acetonitrile, 0.1% formic acid; Sigma-Aldrich) and 5% solvent B
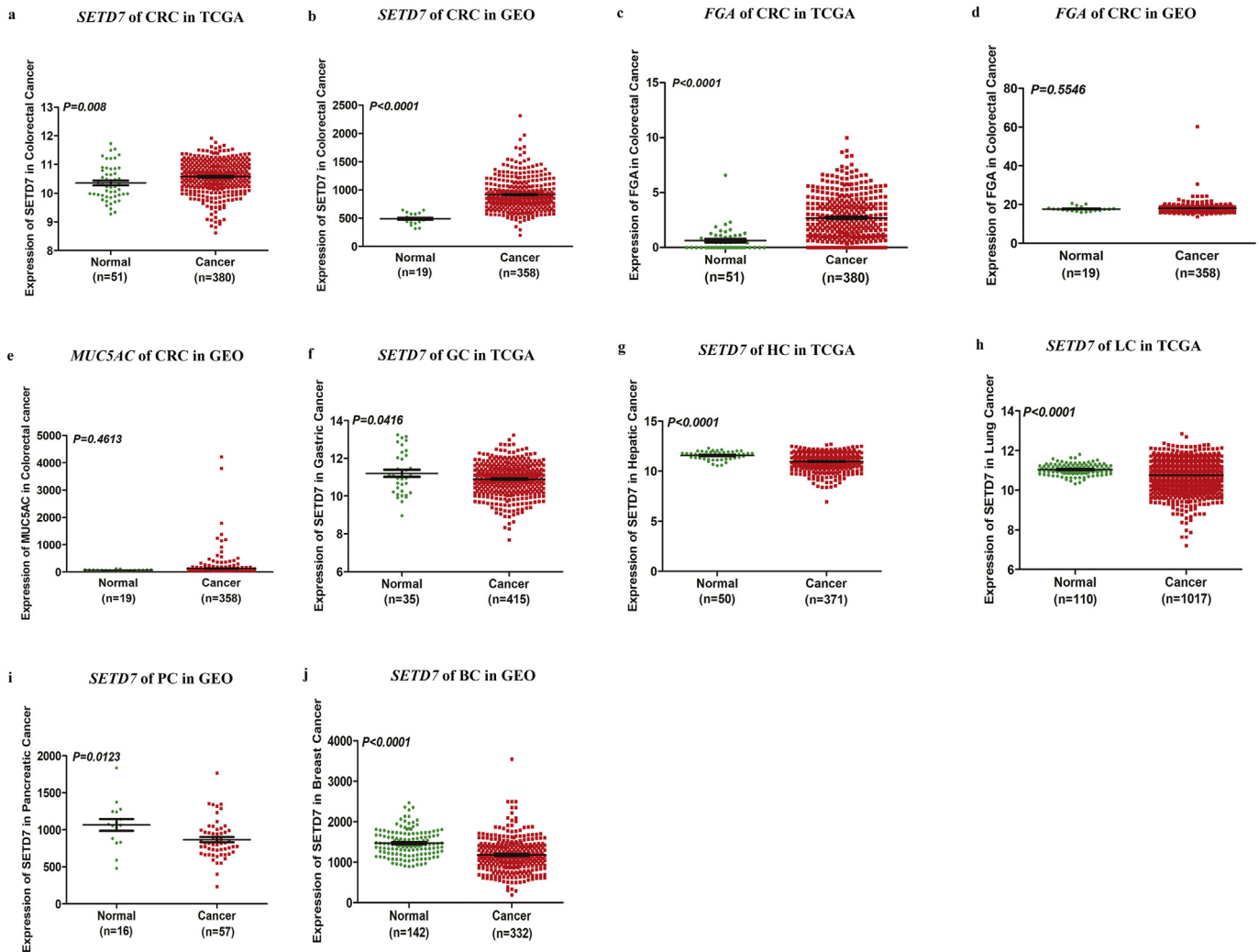

Fig. 2. Bioinformatics analysis. (A, B) Expression pattern of *SETD7* in CRC based on datasets in TCGA and GEO. (C, D) Expression pattern of *FGA* in CRC based on datasets in TCGA and GEO. (E) Expression pattern of *MUC5AC* in CRC based on datasets in GEO. (F) Expression pattern of *SETD7* in gastric cancer based on datasets in TCGA. (G) Expression pattern of *SETD7* in hepatic cancer based on datasets in TCGA. (H) Expression pattern of *SETD7* in lung cancer based on datasets in TCGA. (I) Expression pattern of *SETD7* in pancreatic cancer based on datasets in GEO. (J) Expression pattern of *SETD7* in breast cancer based on datasets in GEO.

(95% acetonitrile, 1.2% formic acid). Peptides were eluted using a 5% to 80% gradient of solvent B over 60 min at a flow rate of 400 nL/min.

The Q Exactive mass spectrometer was operated in the data-dependent mode to switch automatically between MS and MS/MS acquisition. Full-scan survey MS spectra with 2 microscans (m/z 400–2000) were acquired with the Q Exactive with a mass resolution of 100,000 at m/z 400, followed by 10 sequential LC-MS/MS scans. Dynamic exclusion was used with 2 repeat counts, 10 s repeat duration and 60 s exclusion duration. For MS/MS, charge state 1 was rejected and precursor ions were activated using 25% normalized collision energy at the default activation q of 0.25. The mass spectra were searched against the human Uniprot database (https://www.uniprot.org/) using Maxquant software (Version 1.5.2.8). To reduce false positive identification results, a decoy database containing the reverse sequences was appended to the database. The parameters for the search were as follows: no enzyme, the variable modification was oxidation of methionine, peptide tolerance, 20 ppm, MS/MS tolerance, 1.0 Da. Positive protein identification was accepted for a peptide with Xcorr of greater than or equal to 3.20 for triply and 2.86 for doubly charged ions, and all with $\Delta Cn \geq 0.1$, peptide probability ≤2e-3.

2.7. Bioinformatics analysis

Bioinformatics analysis was conducted using The Cancer Genome Atlas database (TCGA, https://cancergenome.nih.gov/) and Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). Gene expression RNAseq dataset of TCGA derived from CRC (380 cancer and 51 normal), Esophageal cancer (184 cancer and 13

normal), Gastric cancer (415 cancer and 35 normal), Hepatic cancer (371 cancer and 50 normal), Pancreatic cancer (178 cancer and 4 normal), Lung cancer (1017 cancer and 110 normal) and Breast cancer (1097 cancer and 114 normal) were analyzed. Gene expression dataset of the GEO was analyzed by Metabolic gEne RApid Visualizer (http://merav.wi.mit.edu/SearchByGenes.html), included CRC (358 cancer and 19 normal), Esophageal cancer (6 cancer and 4 normal), Gastric cancer (15 cancer and 10 normal), Hepatic cancer (14 cancer and 7 normal), Pancreatic cancer (57 cancer and 16 normal), Lung cancer (116 cancer and 10 normal) and Breast cancer (332 cancer and 142 normal).

2.8. ELISA

All serum samples were analyzed blindly, and standards and samples were tested in triplicate. The concentration of SETD7 was quantified using a Human SETD7 ELISA Kit (No. CK-E95510H, Shanghai Elisa Biotech, China) according to the manufacturer's instructions. Optical density was tested at 450 nm.

2.9. IHC

Formalin-fixed, paraffin-embedded sections (4 μm) of CRC, para-carcinoma tissue and colorectal polyps were incubated with rabbit polyclonal anti-SETD7 antibody (1:100, Proteintech Group, Cat# 24840–1-AP, RRID: AB_2737406). After washing with phosphate buffered saline (PBS), the sections were incubated with horseradish peroxidase-conjugated mouse anti-rabbit secondary antibody and treated with 3,3′-diaminobenzidine chromogen substrate buffer. The
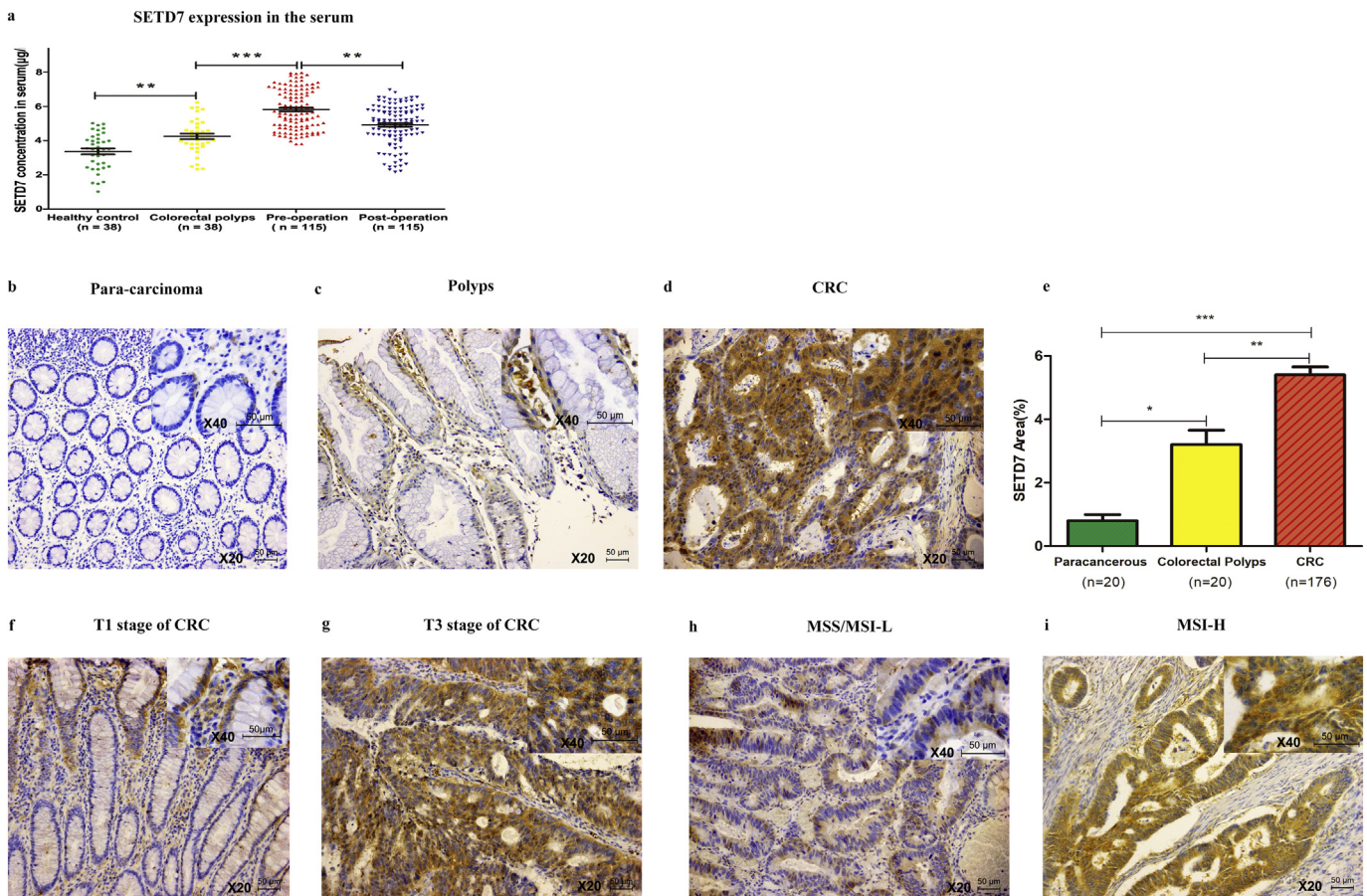

Fig. 3. ELISA and IHC analysis of SETD7 in serum and tissues, respectively. (A) SETD7 expression in the serum of different groups. (B-D) SETD7 expression in para-carcinoma, colorectal polyps and CRC tissue, respectively. (E) Positive expression area of SETD7 in para-carcinoma, colorectal polyps and CRC tissue. (*** indicates $P < .001$, ** indicates $P < .01$, * indicates $P < .05$). (F, G) SETD7 expression in T1 and T3 stage, respectively. (H, I) SETD7 expression in MSS/MSI-L and MSI-H, respectively (figures are magnified ×200, and the upper right corner is magnified ×400).

sections were counterstained with hematoxylin and then mounted on slides. SETD7 expression was analyzed using Image J.

2.10. si-RNA

si-RNA targeting human *SETD7* and a negative control si-RNA were designed and synthesized by Genepharma (Genepharma, Shanghai, China). Transfection of each si-RNA (50 nM) was carried out using Lipofectamine2000 (Thermo Fisher Scientific) according to the manufacturer's recommendations. Sequences of the si-RNA were as follows:

si-*SETD7*–1:
Sense: 5'-GCAUCUACGAUAUGUUUGUTT-3', Antisense: 5'-ACAAAC AUAUCGUAGAUGCTT-3'.

si-*SETD7*–2:
Sense: 5'-CAGUGUACCACUUUGAUAATT-3', Antisense: 5'-UUAUCA AAGUGGUACACUGTT-3'.

Negative control (NC):
Sense: 5'-UUCUCCGAACGUGUCACGUTT-3', Antisense: 5'-ACGUGA CACGUUCGGAGAATT-3'.

2.11. RT-PCR

Total RNA was extracted from the cultured cells using TRIzol Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. RT-PCR analysis was used for relative quantification of *SETD7* mRNA. Total RNA was reverse-transcribed to cDNA using PrimeScript RT Reagent Kit (TAKARA, Japan). cDNA were amplified by real-time PCR on an IQ-5 Real-Time PCR System (Bio-Rad, USA) using SYBR green mix (Thermo Fisher Scientific) with the primers as follows:

SETD7:
Forward: 5'-CGAATTACACACCAAGAGGTT-3', Reverse: 5'-TAGGCA ACGGTGAGCTCTTC-3'.

GAPDH:
Forward: 5'-GTCTCCTCTGACTTCAACAGCG-3', Reverse: 5'-ACCACC CTGTTGCTGTAGCCAA-3'.

All reactions were run in triplicate. Data were analyzed by IQ-5 Series Software. Relative amounts of *SETD7* mRNA were normalized to *GAPDH* mRNA. The relative expression levels were calculated using the $2^{-\triangle\triangle Ct}$ method.

2.12. Western blotting

SETD7 protein levels were quantified by Western blot analysis of whole cell extracts using the antibodies anti-SETD7 (1:1000, Proteintech Group, Cat# 24840–1-AP, RRID: AB_2737406) and anti-GAPDH (1:2000, Abcam, Cat# ab181602, RRID: AB_2630358). GAPDH served as a loading control.

2.13. Cell viability, cell cycle and apoptosis analysis

For cell viability analysis, HCT116 and RKO cells were plated at a concentration of 5×10^3 cells in each well of 96-well plates. After incubation for 24 h, the cells were transfected with si-SETD7 at a final concentration of 50 nM using Lipofectamine2000 (Thermo Fisher Scientific) according to the manufacturer's protocol, and viability was then was assessed using an MTT assay (Boster, Wuhan, China). For cell cycle analysis, HCT116 and RKO cells were plated at a concentration of 5×10^5 cells per well in 6-well plates, and were transfected as above. After incubation for 48 h, cells were harvested, washed with PBS, and fixed in 70% ethanol overnight at −20 °C. Then, cells were treated with RNase A and propidium iodide according to the protocol from the KGA Cell Cycle Detection Kit (Keygen, China). After incubation, the cells were assayed by flow cytometry (Becton, USA). For cell apoptosis analysis, cells were incubated for 48 h following transfection then processed with an Annexin-V FITC Apoptosis Detection Kit (Invitrogen, USA) and

examined by flow cytometry. Data were analyzed using Modfit 3.0 software (Verity Software House, Topsham, ME, USA).

2.14. Statistical analysis

Statistical analysis was performed using GraphPad Prism version 5.0 (GraphPad Software, La Jolla, CA, USA). All data are shown as the mean ± standard deviation. $P < .05$ was considered statistically significant. Chi-square tests were applied to compare categorical variables. Two-tailed Student's t-tests and one-way ANOVA were used to compare quantitative data.

2.15. Data sharing

All raw MS files have been deposited to the ProteomeXchange Consortium (http://proteomecentral. proteomexchange.org) via the iProX partner repository with the dataset identifier PXD011352.

3. Results

3.1. Serum peptidome profiles of different groups

The workflow chart for MALDI-TOF MS-based identification of differentially expressed proteins is shown in Fig. 1. The mass spectra of samples from the healthy controls (Supplementary Fig. S1a), pre-operative patients (Supplementary Fig. S1b), paired post-operative patients (Supplementary Fig. S1c) and patients with colorectal polyps (Supplementary Fig. S1d) showed differences. Component analysis showed some differentially expressed peaks between pre-operative CRC samples (red) and healthy controls (green), while post-operative CRC samples (blue) were similar to healthy controls. Overall, healthy

Table 3
Association between SETD7 expression and clinicopathological features of CRC.

	SETD7 (ELISA)		P value	SETED7 (IHC)		P value
	High (n = 78)	Low (n = 37)		High (n = 151)	Low (n = 25)	
Age (years)			0.6444			0.1398
≥65	58	26		85	18	
<65	20	11		66	7	
Sex			0.2408			0.4748
Male	47	18		90	13	
Female	31	19		61	12	
Histology			0.7917			0.2862
Undifferentiated	12	5		32	3	
Differentiated	66	32		119	22	
MSI			0.0372			0.0277
MSS/MSI-L	56	33		116	24	
MSI-H	22	4		35	1	
RAS[a]			0.1720			0.2652
Wildtype	28	6		25	9	
Mutant	21	10		32	6	
Tumor size[b]			0.0139			0.0451
Tis, T1, T2	45	30		96	21	
T3, T4	33	7		55	4	
Lymph node metastasis			0.9487			0.2893
No	48	23		101	14	
Yes	30	14		50	11	
Distant metastasis			0.4353			0.2725
M0	62	27		123	18	
M1	16	10		28	7	
Stage			0.8413			0.2499
I	19	7		40	6	
II	31	14		61	8	
III	12	6		19	7	
IV	16	10		31	4	

[a] Patients who received the RAS test were analyzed.
[b] For ELISA and IHC analysis, the average tumor size was 43.1 mm and 45.5 mm respectively.

controls, colorectal polyp samples (yellow), pre-operative, and post-operative CRC patient samples showed few overlapping regions (Fig. S2).

3.2. Peaks selection of serum peptidome

A total of 85 peaks were identified among all analyzed groups (see Supplementary data 1), and the abundances of 9/85 peaks revealed statistically significant variation among healthy controls, pre- and post-operative CRC patients, and individuals with colorectal polyps ($P < .000001$, based on the Wilcoxon rank-sum test). Peaks 1–6 tended to be up-regulated in samples from healthy controls, individuals with colorectal polyps and pre-operative patients, and then down-regulated in post-operative patients. In contrast, peaks 7–9 tended to be down-regulated in healthy controls, individuals with colorectal polyps and pre-operative CRC patients, then up-regulated in post-operative patients. Five of these nine peptide peaks (Peak 1, m/z: 2676.04; Peak 5, m/z: 5343.69; Peak 6, m/z: 4793.31; Peak 7, m/z: 3446.91; Peak 9, m/z: 2663.37) with the most significant difference (>1.9 times) between healthy controls and CRC patients were selected for peak identification, and consideration as potential biomarkers for CRC (Table 1). Moreover, these five peaks (Peak 1, m/z: 2676.04; Peak 5, m/z: 5343.69; Peak 6, m/z: 4793.31; Peak 7, m/z: 3446.91; Peak 9, m/z: 2663.37) have significant differences between paired pre- and post-operative CRC patients ($P < .05$, based on the paired t-test).

3.3. Peptide identification

Five peptide peaks (Peak 1, m/z: 2676.04; Peak 5, m/z: 5343.69; Peak 6, m/z: 4793.31; Peak 7, m/z: 3446.91; Peak 9, m/z: 2663.37) with the most significant difference (>1.9 times) between healthy controls and CRC patients were identified using LC-ESI-MS/MS and the Uniprot database. MS/MS spectrum of these peptides identified proteins including FGA, MUC5AC, and SETD7 (Table 2). The MS/MS spectrometry fragment map is shown in Fig. S3.

3.4. Bioinformatics analysis of the potential biomarkers

To determine which candidate biomarker would be further characterized, we examined datasets from TCGA and GEO. Data on MUC5AC were not available in TCGA. The expression of SETD7 but not FGA and MUC5AC is significantly higher in CRC patients compared with normal colorectal mucosa tissue (Fig. 2a-e). In addition, SETD7 is down-regulated significantly in three types of cancer (gastric, hepatic and lung) based on TCGA datasets (Fig. 2f-h), and is down-regulated significantly in two types of cancer (pancreatic and breast) in the GEO database (Fig. 2i-j). Expression data for SETD7 in esophageal, pancreatic and breast cancer based on TCGA datasets and esophageal, gastric, hepatic and lung cancer based on GEO database with no significant difference are shown in Fig. S4.

3.5. Verification of SETD7 expression in serum and tissue samples and analysis of the association between SETD7 expression and clinicopathological features in CRC

The mean serum concentration of SETD7 was 3.364 ± 1.066 ng/mL (range 1.024–5.015 ng/mL) in healthy controls, 4.252 ± 0.987 ng/mL (range 2.332–6.241 ng/mL) in samples from individuals with colorectal polyps, 5.821 ± 1.169 ng/mL (range 3.795–7.967 ng/mL) in pre-operative CRC patients, and 4.917 ± 1.163 ng/mL (range 2.168–6.974 ng/mL) in post-operative CRC patients. The concentration of SETD7 in healthy controls, patients with colorectal polyps and CRC patients showed a gradual increase, and then tended to return to healthy controls values after surgery (Fig. 3a). The data is shown in Supplementary data 2.

Immunoreactivity for SETD7 was detected in the cytoplasm of cells, and the expression area of SETD7 was significantly lower in para-carcinoma samples (Fig. 3b) than in colorectal polyps (Fig. 3c) or CRC (Fig. 3d) tissues ($0.80 \pm 0.87\%$ vs. $3.20 \pm 2.05\%$ vs. $5.40 \pm 3.35\%$, $P < .0001$, Fig. 3e).
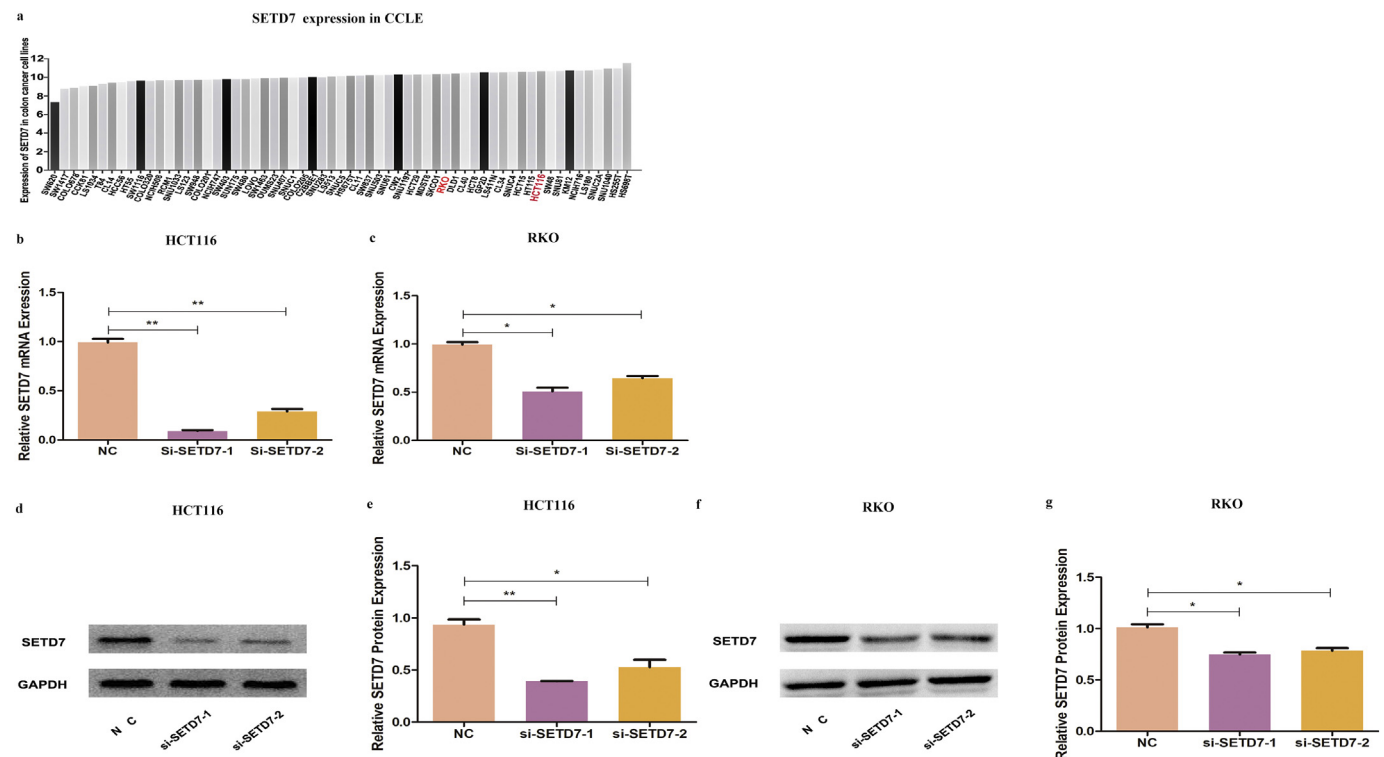

Fig. 4. Knockdown of SETD7 by si-RNA. (A) Expression of SETD7 in colon cancer cell lines in the cancer cell line encyclopedia database. (B-G) Efficiency of si-RNA knockdown of SETD7 in HCT116 and RKO cells. (** indicates $P < .01$, * indicates $P < .05$).

The association between SETD7 expression and clinicopathological features of CRC patients is shown in Table 3. The cutoff values were determined by receiver operating characteristic (ROC) analysis. For ELISA analysis, the sensitivity and specificity of SETD7 were 92.17% and 81.08%, the area under curve (AUC) was 0.9477(95% CI 0.9124–0.9830) and the cutoff value was 5.025 ng/mL. High SETD7 expression was significantly associated with T stage ($P = .0139$) and microsatellite instability (MSI, $P = .0372$). For IHC analysis, the AUC was 0.9357(95% CI 0.8914–0.9799) and the cutoff value was 1.52%. High SETD7 expression was significantly associated with T stage (Fig. 3f, g, $P = .0451$) and MSI (Fig. 3h, i, $P = .0277$).

3.6. Biological roles of SETD7 in colon cancer cell lines

The HCT116 and RKO cell lines express relatively high level of *SETD7*, based on bioinformatics analysis of Cancer cell line encyclopedia (CCLE) (Fig. 4a). Two si-RNA sequences tested in our study demonstrated similar levels of knockdown (Fig. 4b-g). Compared with the control group, *SETD7* knockdown inhibited cell proliferation (Fig. 5a, b), induced G1/S cell cycle arrest (Fig. 5c, d) and prompted apoptosis (Fig. 5e-l).

4. Discussion

In this study, serum proteomic profiles of CRC patients were generated using the combination of MB-WCX fractionation followed by MALDI-TOF MS. Peptides that could distinguish CRC patients (M0/M1 = 91.7%/8.3%) from healthy controls and individuals with colorectal polyps were identified. Moreover, expression of these peptides changed significantly after the removal of the tumor, suggesting that they could be powerful indicators for monitoring the efficacy of anticancer treatment. 5 peptides showing the most significant changes in abundance across paired pre- and post-operation CRC patients, healthy controls and patients with colorectal polyps were chosen for further

identification. Three up-regulated peptides were identified as peptide regions of FGA (Peak 1, *m/z*: 2676.04), MUC5AC (Peak 5, m/z: 5343.69), SETD7 (Peak 6, m/z: 4793.31), and two down-regulated peptides were identified as peptide regions of FGA (Peak 7, m/z: 3446.91, Peak 9, *m/z*: 2663.37).

FGA is human fibrinogen, which is a 340 kDa glycoprotein that is synthesized in the liver and formed by two symmetrical molecules, each of which have three different polypeptide chains. FGA was not only identified as a potential biomarker for CRC in a previous study [12–14], but has also been identified as a potential biomarker for many other cancers including prostate cancer [15], esophageal squamous cell cancer [16], hepatocellular cancer [17] and lung cancer [18].

MUC5AC is a member of the mucin family, a high molecular weight glycoprotein that can be either secreted or membrane bound. MUC5AC has been recently regarded as a potential biomarker for colon adenoma-carcinoma [19], pancreatic cancer [20], cholangiocarcinoma [21] and ovarian cancer [22].

SETD7 (also known as SET7/9) is a 41 kDa lysine-specific SET-domain methyltransferase, which is the only lysine methyltransferases (KMT)7 family member due to its unique enzymatic activity and protein domain architecture [23]. Recent investigations indicated that SETD7 is potentially a biomarker for hepatocellular cancer [24] and breast cancer [25,26], although the role of SETD7 in breast cancer is controversial [25,26].

TCGA and GEO are international public repositories that contain a large number of genomic datasets, and several genes that have diagnostic value in colorectal cancer have been found using these datasets [27–30]. Using bioinformatics analysis, we found that *SETD7*, but not *FGA* and *MUC5AC*, is expressed at significantly higher levels in CRC than in normal colorectal mucosa tissue. Moreover, bioinformatics analysis showed that *SETD7* is down-regulated significantly in several types of the most common cancers, suggesting that the up-regulation of *SETD7* in CRC is relatively specific. Based on the comprehensive analysis
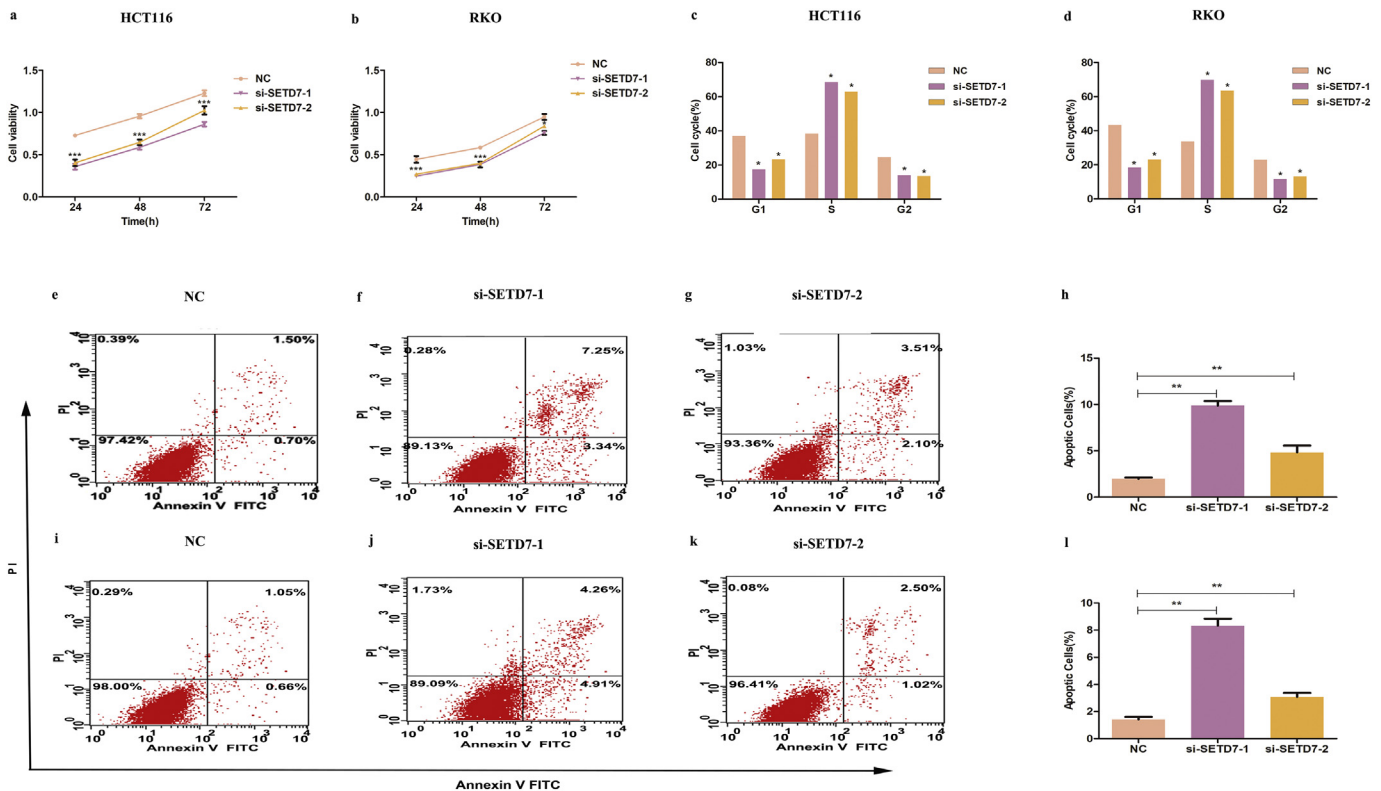

Fig. 5. The effect of *SETD7* knockdown (A, B) Cell viability of *SETD7* knockdown HCT116 and RKO cells. (C, D) Cell cycle analysis of *SETD7* knockdown HCT116 and RKO cells. (*E*-L) Cell apoptosis analysis of *SETD7* knockdown HCT116 and RKO cells. (*** indicates $P < .001$, ** indicates $P < .01$, * indicates $P < .05$).

of our results, SETD7 was selected as the potential serum biomarker for CRC.

Validation studies in serum (M0/M1 = 77.4%/22.6%) and tissue (M0/M1 = 80.1%/19.9%) samples, we found that SETD7 is mainly expressed in the cytoplasm of cells. SETD7 expression increased from healthy individuals to those with colorectal polyps and finally CRC patients, and then showed a tendency to return to healthy controls values after surgery. Furthermore, we found that SETD7 expression was significantly correlated with T stage and MSI in CRC patients. These data suggest that SETD7 may be produced by CRC cells and then released into the blood, and that SETD7 could not only be used as a marker for early diagnosis, but also a candidate marker for prognosis of CRC patients.

As a methyltransferase, numerous non-histone substrates including transcription factors have been described for SETD7, including E2F1, MINT, IRF1, TAF7 and CENPC1 and others [31]. Previous studies have reported that SETD7 is involved in cancer [24,32–34]. Uodhoff et al. reported that SETD7 could control intestinal regeneration and tumorigenesis by regulating the Wnt/β-catenin and Hippo/YAP pathways [33]. Our experiments show that *SETD7* may be involved in CRC by regulating cell cycle and apoptosis in vitro. It is worth noting that we have observed si-SETD7–1 and si-SETD7–2 have similar knockdown effect of *SETD7* in HCT116 and RKO cells, but the apoptosis effects of the two si-RNAs were significantly different. The reason for this difference is not clear, it may be due to the state of the cell culture, the genetic background of the cell and so on, however, the exact mechanism of action requires further study. Together with previous reports, our findings indicate that SETD7 is involved in tumorigenesis and development of CRC.

Our results confirm and extend previous reports indicating that serum peptide signatures could be used for the identification of CRC [34–38]. We propose SETD7 as a potential serum biomarker for CRC. The methods used in this study present an approach to identify potential biomarkers for early diagnosis as well as predicting the effect of treatment in CRC patients. It is worth noting that a single serum biomarker may not have satisfactory discriminating power due to the cancer heterogeneity, and panels of biomarkers may be required [36]. Additionally the number of CRC cases in the discovery set was limited, and further studies will be aimed at verifying biomarker efficacy in a large cohort. In addition, the potential clinical usefulness of this biomarker should be combined with image screening methods to rule out the false positives.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.10.036.

Conflict of interest

The authors declare that they have no conflict of interest.

Authors' contributions

Conception and design: Chen Huang, Baojun Duan, Jun Bai, Jian Qiu and Zongfang Li.

Acquisition of data: Cong Tong, Xiaofei Wang.

Data analysis and interpretation: Juan Yang, Jianhua Wang, Wensheng Li and Jiyu Miao.

Manuscript writing: Baojun Duan, Juan Yang.

Final approval of manuscript: All authors.

References

[1] Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: a review. Updates Surg 2016;68:7–11.

[2] Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016;66:271–89.

[3] Saito G, Sadahiro S, Kamata H, et al. Monitoring of Serum Carcinoembryonic Antigen Levels after Curative Resection of Colon Cancer: Cutoff Values Determined according to Preoperative Levels Enhance the Diagnostic Accuracy for Recurrence. Oncology 2017;92:276–82.

[4] Panis C, Pizzatti L, Souza GF, Abdelhay E. Clinical proteomics in cancer: where we are. Cancer Lett 2016;382:231–9.

[5] Zhou L, Wang K, Li Q, Nice EC, Zhang H, Huang C. Clinical proteomics-driven precision medicine for targeted cancer therapy: current overview and future perspectives. Expert Rev Proteomics 2016;13:367–81.

[6] Yang J, Xiong X, Liu S, et al. Identification of novel serum peptides biomarkers for female breast cancer patients in Western China. Proteomics 2016;16:925–34.

[7] Yu J, Li X, Zhong C, et al. High-throughput proteomics integrated with gene microarray for discovery of colorectal cancer potential biomarkers. Oncotarget 2016;7:75279–92.

[8] Uzozie AC, Selevsek N, Wahlander A, et al. Targeted Proteomics for Multiplexed Verification of Markers of Colorectal Tumorigenesis. Mol Cell Proteomics 2017;16:407–27.

[9] Mori K, Toiyama Y, Otake K, et al. Proteomics analysis of differential protein expression identifies heat shock protein 47 as a predictive marker for lymph node metastasis in patients with colorectal cancer. Int J Cancer 2017;140:1425–35.

[10] Chen Y, Xie Y, Xu L, et al. Protein content and functional characteristics of serum-purified exosomes from patients with colorectal cancer revealed by quantitative proteomics. Int J Cancer 2017;140:900–13.

[11] Zhao C, Su Y, Zhang J, et al. Fibrinogen-derived fibrinostatin inhibits tumor growth through anti-angiogenesis. Cancer Sci 2015;106:1596–606.

[12] Akutekwe A, Seker H, Yang S. In silico discovery of significant pathways in colorectal cancer metastasis using a two-stage optimisation approach. IET Syst Biol 2015;9:294–302.

[13] Karczmarski J, Rubel T, Mikula M, et al. Pre-analytical-related variability influencing serum peptide profiles demonstrated in a mass spectrometry-based search for colorectal and prostate cancer biomarkers. Acta Biochim Pol 2013;60:417–25.

[14] Davalieva K, Kiprijanovska S., Maleva K.I., et al. Comparative Proteomics Analysis of Urine reveals Down-Regulation of Acute phase Response Signaling and LXR/RXR Activation Pathways in Prostate Cancer. Proteomes 2017;6: pii: E1.

[15] Jia K, Li W, Wang F, et al. Novel circulating peptide biomarkers for esophageal squamous cell carcinoma revealed by a magnetic bead-based MALDI-TOFMS assay. Oncotarget 2016;7:23569–80.

[16] Ferrin G, Rodriguez-Peralvarez M, Aguilar-Melero P, et al. Plasma protein biomarkers of hepatocellular carcinoma in HCV-infected alcoholic patients with cirrhosis. PLoS One 2015;10:e118527.

[17] Zhao W, Yang Z, Liu X, et al. Identification of alpha1-antitrypsin as a potential prognostic biomarker for advanced nonsmall cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors by proteomic analysis. J Int Med Res 2013;41:573–83.

[18] Krishn SR, Kaur S, Smith LM, et al. Mucins and associated glycan signatures in colon adenoma-carcinoma sequence: prospective pathological implication(s) for early diagnosis of colon cancer. Cancer Lett 2016;374:304–14.

[19] Kaur S, Smith LM, Patel A, et al. A Combination of MUC5AC and CA19-9 Improves the Diagnosis of Pancreatic Cancer: a Multicenter Study. Am J Gastroenterol 2017;112:172–83.

[20] Pabalan N, Sukcharoensin S, Butthongkomvong K, Jarjanazi H, Thitapakorn V. Expression and Serum Levels of Mucin 5AC (MUC5AC) as a Biomarker for Cholangiocarcinoma: a Meta-analysis. J Gastrointest Cancer 2017;6 pii: E1.

[21] Musrap N, Karagiannis GS, Saraon P, Batruch I, Smith C, Diamandis EP. Proteomic analysis of cancer and mesothelial cells reveals an increase in Mucin 5AC during ovarian cancer and peritoneal interaction. J Proteomics 2014;103:204–15.

[22] Zhang X, Huang Y, Shi X. Emerging roles of lysine methylation on non-histone proteins. Cell Mol Life Sci 2015;72:4257–72.

[23] Chen Y, Yang S, Hu J, Yu C, He M, Cai Z. Increased Expression of SETD7 Promotes Cell Proliferation by Regulating Cell Cycle and Indicates Poor Prognosis in Hepatocellular Carcinoma. PLoS One 2016;11:e154939.

[24] Montenegro MF, Sanchez-Del-Campo L, Gonzalez-Guerrero R, et al. Tumor suppressor SET9 guides the epigenetic plasticity of breast cancer cells and serves as an early-stage biomarker for predicting metastasis. Oncogene 2016;35:6143–52.

[25] Huang R, Li X, Yu Y, et al. SETD7 is a prognosis predicting factor of breast cancer and regulates redox homeostasis. Oncotarget 2017;8:94080–90.

[26] Hao X, Luo H, Krawczyk M, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci U S A 2017;114:7414–9.

[27] Shao Y, Gu W, Ning Z, Song X, Pei H, Jiang J. Evaluating the Prognostic Value of microRNA-203 in Solid Tumors based on a Meta-Analysis and the Cancer Genome Atlas (TCGA) Datasets. Cell Physiol Biochem 2017;41:1468–80.

[28] Sun X, Hu Y, Zhang L, et al. Mining, Validation, and Clinical significance of Colorectal Cancer (CRC)-Associated lncRNAs. PLoS One 2016;11:e164590.

[29] Galamb O, Kalmar A, Bartak BK, et al. Aging related methylation influences the gene expression of key control genes in colorectal cancer and adenoma. World J Gastroenterol 2016;22:10325–40.

[30] Keating ST, El-Osta A. Transcriptional regulation by the Set7 lysine methyltransferase. Epigenetics 2013;8:361–72.

[31] Kim Y, Nam HJ, Lee J, et al. Methylation-dependent regulation of HIF-1alpha stability restricts retinal and tumour angiogenesis. Nat Commun 2016;7:10347.

[32] Oudhoff MJ, Braam M, Freeman SA, et al. SETD7 Controls Intestinal Regeneration and Tumorigenesis by Regulating Wnt/beta-Catenin and Hippo/YAP Signaling. Dev Cell 2016;37:47–57.

[33] Shen C, Wang D, Liu X, et al. SET7/9 regulates cancer cell proliferation by influencing beta-catenin stability. FASEB J 2015;29:4313–23.

[34] Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. Proteomics Clin Appl 2015;9:64–71.

[35] Holm M, Saraswat M, Joenvaara S, Ristimaki A, Haglund C, Renkonen R. Colorectal cancer patients with different C-reactive protein levels and 5-year survival times can be differentiated with quantitative serum proteomics. PLoS One 2018;13: e195354.

[36] Fung KY, Tabor B, Buckley MJ, et al. Blood-based protein biomarker panel for the detection of colorectal cancer. PLoS One 2015;10:e120425.

[37] You J, Kao A, Dillon R, et al. A large-scale and robust dynamic MRM study of colorectal cancer biomarkers. J Proteomics 2018;187:80–92.

[38] Atak A, Khurana S, Gollapalli K, et al. Quantitative mass spectrometry analysis reveals a panel of nine proteins as diagnostic markers for colon adenocarcinomas. Oncotarget 2018;9:13530–44.