



## Homologous proteins SAPCD2X1 and SAPCD2 have significantly different carcinogenic capacities in human colorectal cancer cells based on structural prediction and functional verification

Jingchun Li<sup>1#</sup>, Yuan Chen<sup>1#</sup>, Xiaolong Wang<sup>2</sup>, Wenxi He<sup>1</sup>, Chunli Liu<sup>1</sup>, Feng Feng<sup>1</sup>, Wenwei Zhang<sup>2</sup>, Hongtao Cao<sup>2</sup>, Qixiong Long<sup>3</sup>, Fang Gao<sup>4\*</sup>, Wei Sun<sup>1\*</sup>

<sup>1</sup>Department of Central Laboratory, Huadu District People's Hospital, Guangzhou 510800, China

<sup>2</sup>Department of Gastrointestinal Surgery, Huadu District People's Hospital, Guangzhou 510800, China

<sup>3</sup>Department of Gastroenterology, Huadu District People's Hospital, Guangzhou 510800, China

<sup>4</sup>Department of Critical Care Medicine, Huadu District People's Hospital, Guangzhou 510800, China

# Both authors contributed equally to this work and should be considered as equal first coauthors.

### ARTICLE INFO

#### Original paper

#### Article history:

Received: October 12, 2022

Accepted: November 11, 2023

Published: December 10, 2023

#### Keywords:

Homologous proteins; biological functions; SAPCD2; SAPCD2X1; colorectal cancer

### ABSTRACT

We discussed the expression and biological functions of the SAPCD2X1 protein in the HCT116 CRC cell line by bioinformatics analysis and prediction, and biological function verification. Spatial conformation models of SAPCD2X1 and SAPCD2 were predicted using the threading method, ensemble method, and several other protein structure prediction approaches. The conformational similarity between SAPCD2X1 and SAPCD2 was studied, and their functions were predicted. The biological experiments showed that SAPCD2X1 and SAPCD2 were overexpressed in CRC cells. SAPCD2X1-specific antibodies were prepared. The expressions of SAPCD2X1 and SAPCD2 were localized in cells using the immunofluorescence assay. The SAPCD2 and SAPCD2X1 overexpression models were validated using Western Blot and RT-qPCR. We successfully predicted the structures of the SAPCD2X1 and SAPCD2 proteins and visualized them using the VDM software. It was predicted that the tertiary structure of SAPCD2X1 changed significantly compared with SAPCD2. Alteration of the biological functions of SAPCD2X1 was also predicted due to the changes in the spatial conformation of the protein. Anti-SAPCD2X1 antibody and SAPCD2X1-EGFP and SAPCD2-EGFP recombinant plasmids were established. The overexpression of the two proteins was induced in HCT116 cells using the recombinant plasmids, and verified by RT-qPCR and Western Blot. Meanwhile, the anti-SAPCD2X1 antibody was proved to have a high specificity. The immunofluorescence assay showed that SAPCD2X1 and SAPCD2 are mainly expressed in the cytoplasm. SAPCD2X1 and SAPCD2 exhibited significantly different biological functions in HCT116 cells. SAPCD2 is a carcinogenic protein, while SAPCD2X1 does not affect the proliferation, invasion, and migration of human CRC HCT116 cells.

Doi: <http://dx.doi.org/10.14715/cmb/2023.69.13.2>

Copyright: © 2023 by the C.M.B. Association. All rights reserved.

### Introduction

According to Global Cancer Statistics, colorectal carcinoma (CRC) ranks third among all cancers in terms of incidence, and second in terms of mortality (1). Although the overall incidence of CRC has been decreasing in the United States, CRC has become more common in younger populations than before, and more patients are diagnosed at the advanced stage and in the left colon/rectum (2). Early CRC is usually asymptomatic, and missed diagnosis is common. More than 35% of CRC patients are diagnosed at the advanced stage, thus already missing the best chance for radical surgery. The 5-year survival rate of patients with stage I CRC is 91%, and that of patients with stage IV CRC is only 15% (3). In recent years, studies in genomics, transcriptomics, and proteomics have shown that CRC is associated with a variety of genetic mutations or altered expressions (3), indicating the heterogeneity of CRC. As the pathogenesis of CRC has been partially

understood, targeted drugs have become the mainstream anti-tumor treatment. Targeted therapy has achieved good efficacy and won widespread recognition in clinical practice. Therefore, how to improve the targeting precision and specificity of the targeted therapy, while reducing treatment-related adverse reactions, has become the new trend of development of anti-tumor therapies. Given the above, the major difficulty in CRC diagnosis and treatment lies in understanding the molecular mechanism of tumor occurrence and development and finding tumor-specific biomarkers.

The suppressor APC domain-containing protein 2 (SAPCD2) gene is located on the long arm of chromosome 9 at position 34.3 in the human genome. The carcinogenic effect of the SAPCD2 protein has been supported by in vitro, in vivo, and retrospective clinical studies and the mechanism underlying such carcinogenic effect is partially understood. It was first discovered by Xu et al. as a novel oncogene involved in gastric cancer. SAPCD2 was

\* Corresponding author. Email: [sw3583425@163.com](mailto:sw3583425@163.com); [GF20051014@163.com](mailto:GF20051014@163.com)

found to be specifically expressed in primary gastric cancer tissues, but not expressed in the paired normal gastric mucosa (4). Later, a large number of studies indicated that SAPCD2 expressions in CRC (5), gastric cancer (6), and liver cancer (7) were significantly higher than those in the normal controls. SAPCD2 can promote cancer cell proliferation, migration and invasion, and the malignant transformation of tumor cells (6-8). Some researchers believe that SAPCD2, which is highly expressed in CRC, may be used as an independent prognostic factor in CRC (9). At present, the potential of SAPCD2 as a diagnostic marker and therapeutic target in cancers has been explored and confirmed.

The transcription of the SAPCD2 gene results in two splice variants, namely, SAPCD2 and SAPCD2X1 (suppressor APC domain-containing protein 2 isoform X1), which encode two homologous proteins with a length of 394 and 424 amino acids, respectively. The mRNA sequence of SAPCD2X1 has 90 more nucleotides than SAPCD2. The amino acid sequence of SAPCD2X1 has 30 more amino acids than SAPCD2. The two proteins are completely identical except for the above differences and share highly similar nucleic acid sequences and primary structure. Nevertheless, the functions and biological mechanisms of SAPCD2X1 have not been reported yet. In this study, we verified the expressions and biological functions of SAPCD2X1 in HCT116 cells, based on structural prediction of SAPCD2X1. Our research contributes to the understanding of the diversity of the products and functions of the SAPCD2 gene and lays a theoretical foundation for finding potential targets for CRC treatment.

## Materials and Methods

### Analysis of the primary structures and structural domains of the proteins

The mRNA sequences and amino acid sequences of the SAPCD2 gene were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov/gene/89958>). The primary structures of the proteins were aligned online using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The conserved structural domains of the proteins were analyzed using NCBI's conserved domain database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/>).

### Analysis and prediction of secondary structures of the proteins

The secondary structures were analyzed by MLRC (10) (Multivariate Linear Regression Combination) online ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_server.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html)). This program is an ensemble of three structure prediction methods, namely, GOR4, SIMPA96, and SOPMA (4, 11), and is used for secondary structure analysis of the proteins.

### Analysis and prediction of tertiary structures of the proteins

The fasta files of amino acid sequences of SAPCD2 and SAPCD2X1 were downloaded from NCBI. They were used for the tertiary structure prediction of SAPCD2 and SAPCD2X1 using homology modeling (SWISS-MODEL (12)), threading (I-TASSER (12)), ab initio method (QUARK (13)), ensemble method (ROBETTA (14)), and AlphaFold, successively. After SAVES 6.0 verification, the

models were visualized in VDM. Finally, the models scoring the highest were chosen to predict the spatial conformation of the two homologous proteins.

### GO annotations of sapcd2

The Gene Ontology (GO; <https://www.alliancegenome.org/gene/HGNC:28055#function---go-annotations>) describes the knowledge of the biological domain of sapcd2 with respect to three aspects: Molecular function, Biological process and Cellular component (15).

### Cell culture and transfection

Homologous plasmid construction was performed. The EGFP-labeled SAPCD2X1/SAPCD2-overexpressing plasmids and their respective lentiviral liquids were synthesized by Yunzhou Biotechnology (Guangzhou) Co., Ltd., following the recombinant plasmid construction process. The recombinant plasmids were subjected to enzyme digestion and sequencing before being used for the overexpression experiments. CRC HCT116 cells were cultured at 37°C in a 5% CO<sub>2</sub> incubator. The HCT cells were transfected with the SAPCD2X1- and SAPCD2-overexpressing plasmids using GP-transfect-Mate (G04005, Genepharma, Shanghai, China). The fluorescence intensity was observed under the fluorescence microscope (IX53, Olympus, Tokyo, Japan). Cell pellets were collected at 24 h- 48 h post-transcription for subsequent validation.

### Validation by real-time PCR

Total RNA was extracted from the cell pellets using the RNA extraction kit (AG21017, AG, Changsha, China), and was reverse transcribed into cDNA using the reverse transcription kit (AG, Changsha, China). Real-time qPCR was performed using the Evo M-MLV RT Kit on the qPCR system (CFX-96, Bio-RAD, Hercules, CA, USA), with the following primer sequences:

SAPCD2X1-F: 5'-TTCATGCTTGTGCGCGGTGC-3',  
SAPCD2X1-R: 5'-CCAGCTCCTTCATCTGCTTCAG-3'  
ACTB-F: 5'-CCTGGCACCCAGCACAAT-3', ACTB-R:  
5'-GGGCCGGACTCGTCATAC-3',  
SAPCD2-F: 5'-GAGGTGACCGAGAAGAGTGAG-3',  
SAPCD2-R: 5'-GATGAAGGTGGAATCCAGAGG-3'.  
Gene expressions were calculated using the 2<sup>-ΔΔCt</sup> method based on real-time PCR data.

### Western Blot analysis

The cell pellets were collected and lysed using the RIPA buffer (R0010, Solarbio, Beijing, China). Total proteins were extracted from the cells, and the protein concentration was determined using the BCA Reagent (Thermo, 23225, Waltham, MA, USA). The proteins were then subjected to Western Blot and electrophoresis. Target antibodies: Anti-GAPDH (internal) antibody (60004-1-Ig, Proteintech, USA), anti-EGFP antibody (50430-2-AP, Proteintech, Rosemont, IL, USA), anti-SAPCD2 antibody (NBPI-91740, NOVUS, USA), and anti-SAPCD2X1 antibody (Wuhan Dia-an Biotechnology Co., Ltd., Wuhan, China).

### Immunofluorescence analysis

The cells were transfected with SAPCD2X1- and SAPCD2-overexpressing lentiviral liquids for 48 h. Then, the cells were digested and inoculated to a coverslip for 24 h. After that, the cells were fixed with 4% parafor-

maldehyde and washed with PBS buffer. The cells were stained with DAPI (C1005, Beyotime, Shanghai, China) for 5 min, and washed with PBS buffer. Finally, the coverslip was sealed with anti-quenching agent (H-1000, VECTOR, USA). Confocal microscopy images were taken at the green fluorescent protein excitation wavelength of 488 nm and at the UV excitation wavelength.

### Cell scratch test

HCT116 cells were inoculated to the 6-well plate and transfected with the overexpression plasmids. A line was streaked per well using a yellow pipette tip. The gaps between the line and the wound-bordering cells were observed once every 24 h. Cell images were taken under the inverted microscope (IX53, Olympus, Tokyo, Japan) at the specified time points.

### CCK-8 assay

After the cells were transfected with the overexpression lentiviral liquid for 48 h, the cells were digested with pancreatin, counted and inoculated to the 96-well plate at the concentration of  $2 \times 10^3$  cells per well. Next, the cells were incubated in the incubator for 24 h, 48 h, 72 h, and 96 h, respectively. Cell proliferation was assessed using the Cell Counting Kit-8 (CK04, Dojindo Laboratories, Kumamoto, Japan). Into each well, 10  $\mu$ l of CCK-8 reagent and 100  $\mu$ l of culture medium were added. The cells were incubated at 37°C in an incubator for 1 h. The absorbance was determined at 450 nm using a microplate reader (Elx800, BioTek, Biotek Winooski, VT, USA).

### Plate clonal formation assay

After the cells were transfected with the overexpression lentiviral liquids for 48 h, the cells were digested with pancreatin, counted and inoculated to the 96-well plate at the concentration of  $1 \times 10^3$  cells per well. Next, the cells were incubated in an incubator for 10-14 days and fixed with 4% paraformaldehyde (P0099, Beyotime, Shanghai, China). The cells were washed with PBS buffer and stained with 0.1% crystal violet (G1063, Solarbio, Beijing, China) for 55 min. The cells were then washed with PBS buffer again and air-dried, and images were taken.

### Statistical Analysis

All data were evaluated using SPSS 19.0.0 software (SPSS, Chicago, IL, USA) and presented as mean $\pm$ SD. Statistical analysis was performed using the Student's t-test, and a p-value < 0.05 was considered to indicate a significant difference. All experiments were repeated at least three times.

## Results

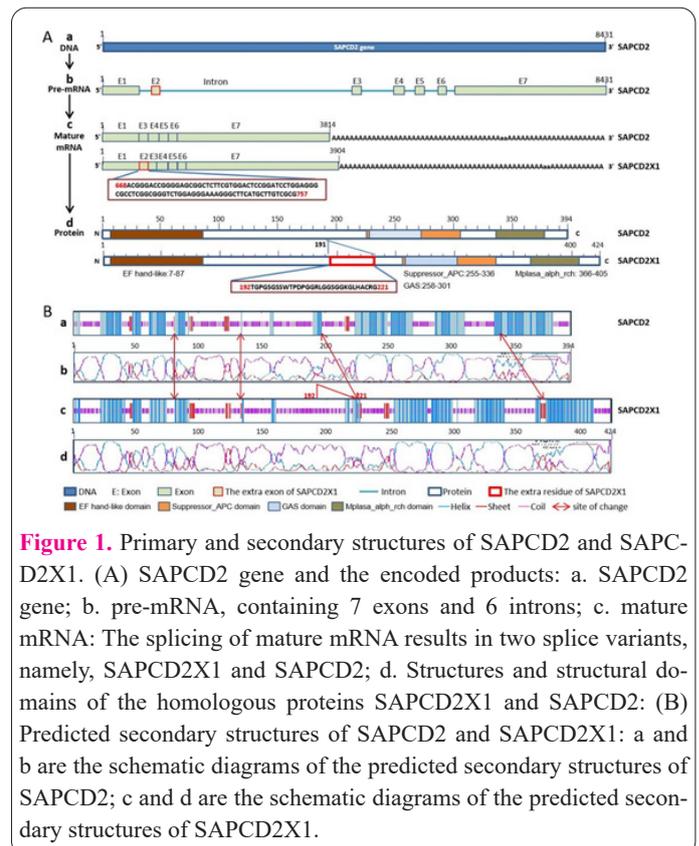
### Analysis of the primary and secondary structures and structural domains of SAPCD2 and SAPCD2X1

According to the downloads from the GenBank database, the full length of the SAPCD2 gene (GeneID: 89958) contains 8431 base pairs. The transcription of the gene results in a pre-mRNA that contains 7 exons, which further encodes two mature mRNAs, one for SAPCD2 and the other SAPCD2X1 (transcript\_id: NM\_178448.4, XM\_011519180.4). Alignment of nucleic acid sequences of SAPCD2 and SAPCD2X1 was performed. It was found that the mature mRNA of SAPCD2X1 is composed of all

seven exons, while that of SAPCD2 is composed of the remaining six exons except for the second exon. Therefore, the mature mRNA of SAPCD2X1 contains 90 more nucleotides compared with SAPCD2. The remaining portion of the nucleotide sequences are identical, and the two mRNA isoforms have highly similar sequences (Figure 1A: a, b, c).

The coding sequences (CDS) of the two mRNA isoforms encode two homologous proteins, SAPCD2 and SAPCD2X1, which have a length of 394 and 424 amino acids, respectively (protein\_id: NP\_848543.2, XP\_011517482.1). Compared with the sequence after position 191 in SAPCD2, SAPCD2X1 contains 30 more amino acids encoded by exon 2 (Figure 2, 192T $\rightarrow$ 221G). The remaining portion of the amino acid sequence is completely the same between SAPCD2 and SAPCD2X1. Therefore, the primary structures of the two homologous proteins are also highly similar. The analysis using NCBI's Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/>) identified three conserved C-terminal domain in the two proteins: suppressor APC domain (Cdd:pfam11414), GAS domain (Cdd:pfam13851), and Mplasa\_alpha\_rch superfamily domain (Cdd:TIGR04523). The GAS domain is contained within the suppressor APC domain. The existing literature reports an EF-hand-like domain from position 7 to position 87 on the N-terminal of the homologous proteins (4, 16). Due to the high identity between the amino acid sequences of the two proteins, both contain all of the above domains (Figure 1A: d). The sequences of the three C-terminal conserved domains are shown in Figure 3.

MLRC was run to predict the secondary structures of SAPCD2 and SAPCD2X1, which also revealed high similarity between the two homologous proteins (Figure 1B). SAPCD2 contained 176  $\alpha$ -helices (h), which accounted for 44.67%; there were 14  $\beta$ -pleated sheets, accounting for 3.55%. SAPCD2 contained 178  $\alpha$ -helices (h), which





**Figure 2.** Comparison of primary and secondary structures of SAPCD2 and SAPCD2X1.

accounted for 41.98%; there were 21  $\beta$ -pleated sheets, accounting for 4.95% (Figure 2). Although the two homologous proteins shared similar secondary structures, the extra 30 amino acids in SAPCD2X1 caused an alteration in the secondary structure. Apart from the insertion of the 30 amino acids, the secondary structure of SAPCD2 changed within a length of 14 amino acids before and after position 191, the insertion site (indicated by the red arrowheads in Figure 2), resulting in more  $\beta$ -pleated sheets.

### Prediction of tertiary structures of the isomeric proteins

SWISS-MODEL was run to predict the three-dimensional structure using homology modeling. The amino acid sequences of the targets SAPCD2 and SAPCD2X1 had an identity of only 22.22%, with the longest consensus sequence in the template database (615j.1.A, Rootletin). The sequence identity was below 30%, indicating that the template was not suitable for homology modeling. To build models using QUARK, the target protein should not be longer than 300 amino acids. Therefore, we used I-TASSER and ROSETTA to predict the tertiary structures of the proteins. According to I-TASSER modeling, the C-score of the first-ranking model of SAPCD2 was -1.39 and the TM-score was 0.54±0.15; the C-score of the first-ranking model of SAPCD2X1 was -1.38 and the TM-score was 0.54±0.15. Both were valid models. The models were then validated by SAVES6.0 but failed to pass the verification. The model quality did not meet the requirements (Supplementary Table 1). The tertiary structures generated by AlphaFold did not pass the SAVES6.0 verification. Among all prediction approaches, the quality of the models built by ROSETTA was the highest. All of the top 3 models for SAPCD2 passed the SAVES6.0 verification. Models 1 and 3 of SAPCD2X1 also passed the SAVES6.0 verification.

Among all models passing the SAVES6.0 verification, model 3 had the highest quality for either SAPCD2 or

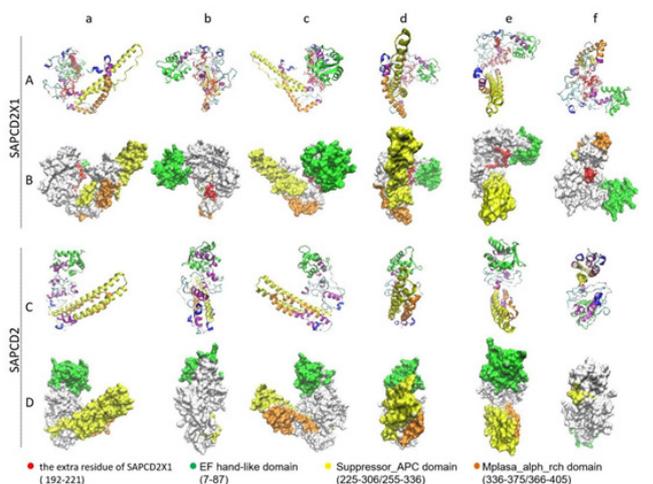
SAPCD2X1. There were 2 errors and 4 passes in PROCHECK. In the Ramachandran plot, the number of amino acid residues reached 91.1% in the most favored regions (A,B,L). For model 3 of SAPCD2X1, there were 0 errors and 5 passes in PROCHECK. In the Ramachandran plot, the number of amino acid residues in the most favored regions (A,B,L) reached 91.1% (Supplementary Table 1). A high-quality model is expected to have a number of amino acid residues above 90% in this region. Therefore, we obtained high-quality models of the tertiary structure generated by ROSETTA for SAPCD2 and SAPCD2X1, respectively. We further analyzed the structures and predicted the functions of SAPCD2 (model\_21142\_3) and SAPCD2X1 (model\_18305\_3) using the models with the highest quality.

As visualized by VMD (Figure 4), SAPCD2 and SAPCD2X1 each had two independent regions, one of which is known as the long arm, mainly composed of the  $\alpha$ -helices contained in the Suppressor<sub>APC</sub> domain and the Mplasa<sub>alph\_rch</sub> domain; the shorter one is known as the short arm, composed of random coils and EF hand-like domain, with the EF hand-like domain located at the end of the short arm.

Although the two proteins shared many similarities, SAPCD2X1 contained 30 more amino acids than SAPCD2, resulting in major changes in the spatial conformation. The long arm of SAPCD2 contained four  $\alpha$ -helices

Name	Source of the Sequence alignment	E-value
Suppressor <sub>APC</sub>	SAPCD2/SAPCD2X1 225/255 DCGLLKQIKELQEKYVLIQGLNMAAGHDYQQQLQRYQEQRRLEGQRASAD	4.17e-32
APC	FGAAGSPRI <sub>GRLLPKVQV</sub> ARCLGELL 306/336	
	Cdd:pfam1414 001DYNLKRKQLEQEKYVLIQGLNMAAGHDYQQQLQRYQEQRRLEGQRASAD	
	FQSDAKQERIEPLLARIQV <sub>NCLGG</sub> LI 82	
GAS	SAPCD2/SAPCD2X1 228/258 LLKQKRELEQEKYVLIQGLNMAAGHDYQQQLQRYQEQRRLEGQRASAD	5.12e-03
	Cdd:pfam13851 097LEKELKDKKWEIVLEQEKYVLIQGLNMAAGHDYQQQLQRYQEQRRLEGQRASAD	141
Mplasa <sub>alph_rch</sub>	SAPCD2/SAPCD2X1 336/366 QQQTILMLKEQNRLTQEVTEKSERITQLQEKSALEKQLFEARALSQQDGGPDLSTFI	2.07e-03
	Cdd:TIGR04523 566KNKEIEELAQKQSLKQKQEKQLIDQEKYKDKLISEI 605	

**Figure 3.** C-terminal conserved domains of SAPCD2 and SAPCD2X1.



**Figure 4.** 360-degree panorama view of the predicted tertiary structures of SAPCD2X1 and SAPCD2. (A) SAPCD2X1 secondary structure model; (B) SAPCD2X1 surface opaque model; (C) SAPCD2 secondary structure model; (D) SAPCD2 surface opaque model; a: initial position of model; b-d: Rotate 90° counterclockwise according to the previous model; e: aerial view of d; f: bottom view of e.

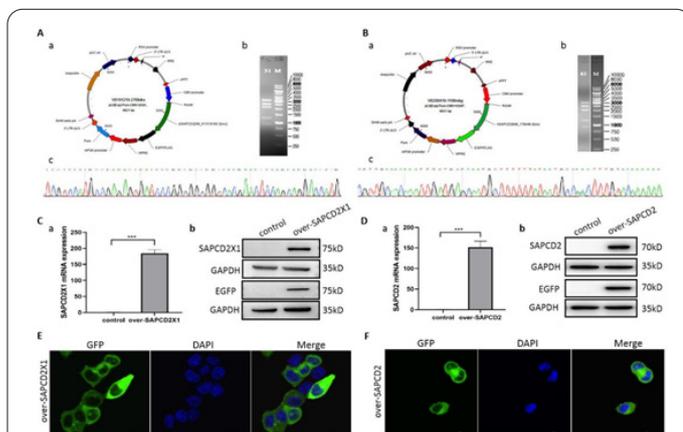
parallel with each other. Suppressor<sub>APC</sub> domain and Mplasa<sub>alph\_rch</sub> domain were parallel to each other, and the long and short arms were located on the same plane. However, the long arm of SAPCD2X1 was composed of three  $\alpha$ -helices, with the Suppressor<sub>APC</sub> domain forming the main axis of the long arm, and the Mplasa<sub>alph\_rch</sub> domain coiling in the middle and lower parts of the main axis. The most important change was that in the presence of 30 extra amino acids in SAPCD2X1, the short arm was rotated by about 90 degrees (Figure 4). As a result, the plane of the long arm was perpendicular to that of the short arm, making the EF-hand-like domain far away from the Suppressor<sub>APC</sub> domain.

## GO annotations

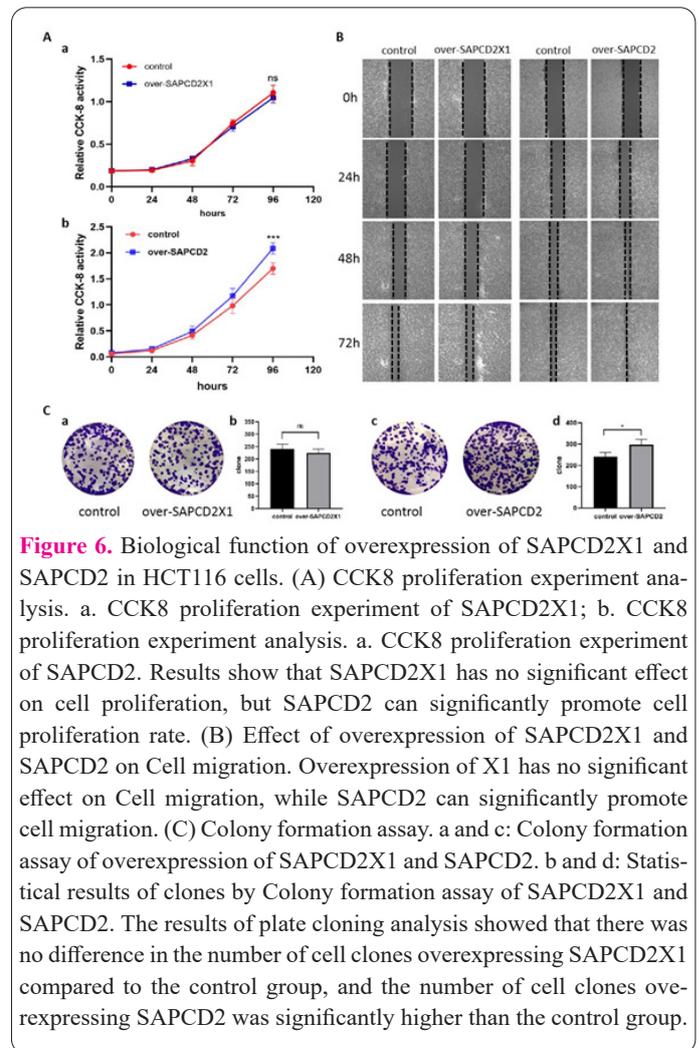
The Gene Ontology describes the knowledge of the biological domain of *sapcd2* with respect to three aspects (Supplementary Table 2): Molecular functions including enzyme regulator activity, DNA-binding transcription factor activity and signaling receptor activity, etc; Biological processes including cell cycle, cell population proliferation and cell death, etc.; Cellular components including the plasma membrane, cytoplasmic vesicle and cytosol, etc.

## Construction, identification, and localization of the SAPCD2X1- and SAPCD2-overexpressing plasmids

The SAPCD2X1- and SAPCD2-overexpressing plasmids were purified, digested with restriction enzymes, identified and sequenced (Figure 5A and B), which verified that the two plasmids were successfully constructed. HCT116 cells were transfected with the SAPCD2X1- and SAPCD2-overexpressing plasmids, and used for real-time



**Figure 5.** Construction of SAPCD2X1 and SAPCD2 recombinant plasmids and overexpression in HCT116 cells. (A) a. Simulation map of SAPCD2X1 plasmid; b. SAPCD2X1 plasmid digestion identification; c. The sequencing results of the SAPCD2X1 plasmid are correct; (B) a. SAPCD2 plasmid simulation map; b. SAPCD2X1 plasmid digestion identification; c. The sequencing results of the SAPCD2X1 plasmid are correct. (C) a. Quantitative real-time PCR, compared with the control group, the expression level of SAPCD2X1 mRNA was significantly increased, \*\*\* $P < 0.01$ ; b. overexpression of SAPCD2X1 plasmid and control plasmid in cells, verified by WB; (D) a. Quantitative real-time PCR, compared with the control group, the expression level of SAPCD2 mRNA was significantly increased, \*\*\* $P < 0.01$ ; b. overexpression of SAPCD2 plasmid and control plasmid in cells, verified by WB. (E) and (F) Subcellular localization of SAPCD2X1 and SAPCD2 verified by immunofluorescence after transfection with lentivirus solution in cells. The results showed that SAPCD2X1 and SAPCD2 were mainly expressed in the cytoplasm.



**Figure 6.** Biological function of overexpression of SAPCD2X1 and SAPCD2 in HCT116 cells. (A) CCK8 proliferation experiment analysis. a. CCK8 proliferation experiment of SAPCD2X1; b. CCK8 proliferation experiment analysis. a. CCK8 proliferation experiment of SAPCD2. Results show that SAPCD2X1 has no significant effect on cell proliferation, but SAPCD2 can significantly promote cell proliferation rate. (B) Effect of overexpression of SAPCD2X1 and SAPCD2 on Cell migration. Overexpression of X1 has no significant effect on Cell migration, while SAPCD2 can significantly promote cell migration. (C) Colony formation assay. a and c: Colony formation assay of overexpression of SAPCD2X1 and SAPCD2. b and d: Statistical results of clones by Colony formation assay of SAPCD2X1 and SAPCD2. The results of plate cloning analysis showed that there was no difference in the number of cell clones overexpressing SAPCD2X1 compared to the control group, and the number of cell clones overexpressing SAPCD2 was significantly higher than the control group.

qPCR and Western Blot. The results showed that the mRNA and protein expressions of SAPCD2X1 and SAPCD2 were significantly higher compared with the controls (Figure 5C and D). The anti-EGFR and anti-SAPCD2X1 antibodies were detected as bands at the same positions, which served as indirect evidence for the good specificity of the anti-SAPCD2X1 antibody. The overexpression plasmids were made into lentiviral liquids and used to transfect HCT116 cells. The cells were observed by confocal microscopy after 48 h of culture. The fluorescence assay showed that SAPCD2X1 and SAPCD2 were mainly expressed in the cytoplasm. The two proteins shared the same subcellular localization.

## Biological functions of the HCT116 cells transfected with the SAPCD2X1- and SAPCD2-overexpressing plasmids

The HCT116 cells overexpressed SAPCD2X1 and SAPCD2. The influences of SAPCD2X1 and SAPCD2 on the functions of the cells were assessed by the CCK-8 assay, cell scratch test, and plate clone formation assay (Figure 6). The results showed that SAPCD2X1 overexpression had no significant impact on the proliferation and migration of HCT116 cells. However, compared with the control group, SAPCD2 overexpression dramatically promoted the migration and proliferation of HCT116 cells. These results suggested that SAPCD2X1 did not exert biological functions in HCT116 cells.

## Discussion

At present, all conclusions that have been drawn concerning the functions of the SAPCD2 gene are based on studies on the SAPCD2 protein. However, we found through sequence alignment, using Blast from NCBI, that SAPCD2X1 and SAPCD2 are homologous proteins encoded by the SAPCD2 gene. The functions of the SAPCD2 gene cannot be fully defined based on those of the SAPCD2 protein alone. Here, we analyzed the biological functions of the SAPCD2X1 protein by GO annotations, which provided a direction for further investigations. The GO annotations are shown in Supplementary Table 2. Based on the structural similarity of SAPCD2 and the published studies, the SAPCD2 gene performs the following molecular functions: enzyme regulator activity, signaling receptor binding, cytoskeletal protein binding, DNA-binding transcription factor activity, transcription factor binding, and metal ion binding. On this basis, the SAPCD2 gene is mainly involved in the following biological processes: cell cycle, cell population proliferation, differentiation, cell death, developmental process, immune system process, and carbohydrate metabolic process. Given the roles of the SAPCD2 gene in the above molecular functions and biological processes, we infer that the SAPCD2 gene may fulfill important functions in tumor biology. An in-depth discussion of the molecular functions and working mechanisms of SAPCD2X1 and SAPCD2 in tumor progression may open up a new pathway for understanding the roles of the SAPCD2 gene in tumor occurrence and development.

The BLAST alignment revealed that SAPCD2X1 has one more exon than SAPCD2. The mature mRNA of SAPCD2X1 contains 90 more nucleotide residues than that of SAPCD2. The amino acid sequence of the former has 30 more amino acids than that of the latter, while the remaining portions of the sequences are identical. No frameshift mutation occurs after the insertion site (Figure 1A). The two proteins are highly similar in primary and secondary structures, which implies functional similarity. Despite the high similarity in the secondary structure of the two proteins, the SAPCD2X1 protein contains a specific insertion of 30 amino acids and exhibits alteration in the predicted secondary structure at 14 amino acid sites (Figure 1B and Figure 2). However, the 30 amino acids specific to SAPCD2X1 do not contain any conserved structural domains. In other words, both proteins contain four common structural domains, namely, Suppressor<sub>APC</sub>, GAS, Mplasa<sub>alph\_rch</sub>, and EF-hand (4, 17-19) (Figure 1A). The Suppressor<sub>APC</sub> domain is sourced from the Adenomatous polyposis coli tumor suppressor protein, and the crystal structure consists of three alpha-helices forming two separate antiparallel coiled coils. The c-terminus of the first helix contains the nuclear export signal (NES), but its function is still unclear. This domain includes the GAS domain. This family is the highly conserved central region of a number of metazoan proteins referred to as growth-arrest proteins, which may have a role in the functioning of motile cellular appendages and cell arrest. Several studies of SAPCD2 have demonstrated its involvement in cell cycle regulation (4, 20-22). Moreover, SAPCD2 may play a vital role in the oriented division of spindles (4, 20-22). Mplasa<sub>alph\_rch</sub> superfamily domain is derived from helix-rich Mycoplasma protein. Sequences are predicted (Jpred 3) to be almost entirely alpha-helical, and the function is still unknown.

The N-terminus of SAPCD2 and SAPCD2X1 contains an EF-hand domain spanning amino acids 7-87 (4). The EF-hand-like domain is characterized by a helix-loop-helix structure and predicted a metal ion binding site involving four amino acids within this region of SAPCD2 (17, 23). The functional importance of the EF-hand-like domain has been recognized through its association with the calcium-bound S100 oncoprotein family, which is involved in a variety of important intracellular regulatory functions, including proliferation, tumor transformation, differentiation, protein phosphorylation, inflammation, and maintenance of intracellular calcium homeostasis. However, whether SAPCD2X1 and SAPCD2 have calcium-regulating functions has not been demonstrated in the laboratory. These structural domains in SAPCD2X1 and SAPCD2 offer further clues for analyzing the functions of the SAPCD2 gene products of SAPCD2X1 and SAPCD2.

Secondary structure and structural domains are important functional bases of the proteins. Some structural domains perform basic biological functions. Therefore, based on the similarity between the two homologous proteins in structural domains, we infer that they may play important roles in tumor cell formation, proliferation and migration. Besides, the biological functions and mechanisms of the SAPCD2 protein have been validated in many solid tumors.

Although similarities in secondary structure and structural domains of the proteins imply similarities in biological functions, changes in secondary structure also imply the possibility of variations in spatial conformation and functions of the protein. Among all high-quality predicted models of tertiary structures of SAPCD2X1 and SAPCD2, the spatial conformation of SAPCD2X1 changes significantly due to the 30 extra amino acids. Its short arm is rotated by about 90 degrees (Figure 4), and its long arm and short arm are no longer located on the same plane, but perpendicular to each other. As a result, the EF-hand-like domain is kept further away from the Suppressor<sub>APC</sub> domain. Meanwhile, in the predicted models, the spatial conformation of the four structural domains also changes for SAPCD2X1. The above changes imply the possibility of different biological functions of the protein. Studies have shown that SAPCD2 promotes the proliferation, migration, invasion and growth of CRC (5, 9), gastric cancer (21, 24), liver cancer (7), lung cancer (25, 26), breast cancer (27, 28), prostate cancer (29), melanoma (30), and kidney cancer (31). In this study, we constructed the SAPCD2X1- and SAPCD2-overexpressing plasmids, using the biological functions of SAPCD2 as control. We arrived at similar conclusions regarding the functions of SAPCD2, compared with our preliminary findings (31) and the results from other research groups. Unlike SAPCD2, the CCK-8 assay did not reveal significant differences between the SAPCD2X1-overexpressing cells and the control cells. The cell scratch test showed that SAPCD2X1 overexpression did not affect the migration capacity of HCT116 cells. The plate clone formation assay did not find significant differences between SAPCD2X1-overexpressing HCT116 cells and the control cells in cell colony formation. We found that SAPCD2X1 did not exhibit the same tumor-promoting effects as SAPCD2, which confirms the hypothesis that changes in spatial conformation do result in functional alterations. Our research sheds some light on how changes in spatial conformation of the protein lead to functional

alterations. Although the spatial conformation of the two proteins needs to be verified by protein crystal X-diffraction, we already obtained the high-quality predicted models of the protein structures using different approaches. We also revealed the close connections between biological functions and spatial conformation of the proteins. There is also the possibility that the spatial distance between the key structural domains determines the biological functions of the proteins. Abundant evidence from laboratory tests has proved that SAPCD2X1 is a non-carcinogenic protein, while SAPCD2 is a carcinogenic protein. The entirely different biological functions of the two homologous proteins prove that the spatial conformation is indeed different between the two proteins. The fact that the SAPCD2 gene encodes two functionally different homologous proteins opens a new pathway for studying and treating CRC involving the SAPCD2 gene. That is, the *in vivo* regulation of the splicing, and modification to promote the transition from SAPCD2X1 mRNA to SAPCD2 mRNA, which may be a switching event for the occurrence of CRC. However, the mechanism underlying such transition remains to be elucidated.

To conclude, the homologous proteins SAPCD2X1 and SAPCD2 have the same subcellular localization in HCT116 cells and share highly similar primary structures and structural domains. However, the two proteins differ in secondary and tertiary structures. SAPCD2 promoted the proliferation, migration and invasion of tumor cells, and acted as a carcinogenic protein. On the contrary, SAPCD2X1 had no significant impact on tumor cell proliferation and migration, which implies the important influence of spatial conformation on the biological functions of the two homologous proteins. The regulatory mechanism underlying the transition from the non-carcinogenic protein SAPCD2X1 to the carcinogenic protein SAPCD2 may provide a new target for the diagnosis and treatment of CRC.

#### Author Contributions

Jingchun Li, Yuan Chen, Fang Gao and Wei Sun conceived the study and designed the experiments. Xiaolong Wang, Wenxi He, Chunli Liu and Feng Feng contributed to the data collection. Wenwei Zhang, Hongtao Cao and Qixiong Long performed the data analysis and interpreted the results. Jingchun Li and Yuan Chen wrote the manuscript. Fang Gao and Wei Sun contributed to the critical revision of the article. All authors have read and approved the manuscript.

#### Acknowledgments

None.

#### Conflicts of Interest

The authors declare that they have no competing interests.

#### Ethical Statement

This article does not contain any studies with human or animal subjects performed by any of the authors.

#### Consent

Not applicable.

#### Funding

This work was supported by the Guangzhou Science and

Technology Project (202102080621); the Guangzhou Health Science and Technology Project (20201A011113, 20231A010080); the People's Hospital of Huadu District Science Project (2021B03, 2021C02); the Key Discipline of Digestive Diseases of Huadu District People's Hospital (YNZDXK202201) and Huadu District Basic and Applied Basic Research Joint Funded Project (23HDQYLH04).

#### References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca-Cancer J Clin* 2021; 71(3): 209-249.
- Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *Ca-Cancer J Clin* 2023; 73(3): 233-254.
- Gorin SS. Care of Cancer Survivors: Special Issues for Colorectal Cancer Survivors. *Fp Essent* 2023; 529: 23-33.
- Xu X, Li W, Fan X, et al. Identification and characterization of a novel p42.3 gene as tumor-specific and mitosis phase-dependent expression in gastric cancer. *Oncogene* 2007; 26(52): 7371-7379.
- Luo Y, Wang L, Ran W, et al. Overexpression of SAPCD2 correlates with proliferation and invasion of colorectal carcinoma cells. *Cancer Cell Int* 2020; 20: 43.
- Cao WJ, Du WQ, Mao LL, Zheng JN, Pei DS. Overexpression of p42.3 promotes cell proliferation, migration, and invasion in human gastric cancer cells. *Tumour Biol* 2016; 37(9): 12805-12812.
- Sun W, Dong WW, Mao LL, et al. Overexpression of p42.3 promotes cell growth and tumorigenicity in hepatocellular carcinoma. *World J Gastroentero* 2013; 19(19): 2913-2920.
- Li M, Wang Z, Guo J, et al. Clinical significance of UGT1A1 gene polymorphisms on irinotecan-based regimens as the treatment in metastatic colorectal cancer. *Oncotargets Ther* 2014; 7: 1653-1661.
- Weng YR, Yu YN, Ren LL, et al. Role of C9orf140 in the promotion of colorectal cancer progression and mechanisms of its upregulation via activation of STAT5, beta-catenin and EZH2. *Carcinogenesis* 2014; 35(6): 1389-1398.
- Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. *Trends Biochem Sci* 2000; 25(3): 147-150.
- Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* 1995; 11(6): 681-684.
- Bienert S, Waterhouse A, de Beer TA, et al. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* 2017; 45(D1): D313-D319.
- Mortuza SM, Zheng W, Zhang C, Li Y, Pearce R, Zhang Y. Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. *Nat Commun* 2021; 12(1): 5011.
- Park H, Kim DE, Ovchinnikov S, Baker D, DiMaio F. Automatic structure prediction of oligomeric assemblies using Robetta in CASP12. *Proteins* 2018; 86 Suppl 1(Suppl 1): 283-291.
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25(1): 25-29.
- Hao Y, Fan T, Nan K. Optimization and Corroboration of the Regulatory Pathway of p42.3 Protein in the Pathogenesis of Gastric Carcinoma. *Comput Math Method M* 2015; 2015: 683679.
- Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017; 45(D1): D200-D203.

18. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015; 43(Database issue): D222-D226.
19. Yang M, Derbyshire MK, Yamashita RA, Marchler-Bauer A. NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. *Curr Protoc Bioinformatics* 2020; 69(1): e90.
20. Zhang J, Lu C, Shang Z, Xing R, Shi L, Lv Y. p42.3 gene expression in gastric cancer cell and its protein regulatory network analysis. *Theor Biol Med Model* 2012; 9(1): 53.
21. Mao L, Sun W, Li W, et al. Cell cycle-dependent expression of p42.3 promotes mitotic progression in malignant transformed cells. *Mol Carcinogen* 2014; 53(5): 337-348.
22. Cao WJ, Mao LL, Zheng JN, Pei DS. p42.3: An abductor of cell cycle. *Anti-Cancer Agent Me* 2015; 15(2): 157-162.
23. Nelson MR, Thulin E, Fagan PA, Forsen S, Chazin WJ. The EF-hand domain: a globally cooperative structural unit. *Protein Sci* 2002; 11(2): 198-205.
24. Gao F, Chen Y, Zhang W, et al. The Generation and Application of Monoclonal Antibodies to Detect SAPCD2 Expression in Pre-cancerous and Malignant Gastric Lesions. *Iran J Immunol* 2023; 20(2): 190-201.
25. Wei D. MiR-486-5p specifically suppresses SAPCD2 expression, which attenuates the aggressive phenotypes of lung adenocarcinoma cells. *Histol Histopathol* 2022; 37(9): 909-917.
26. Zhu B, Wu Y, Niu L, et al. Silencing SAPCD2 Represses Proliferation and Lung Metastasis of Fibrosarcoma by Activating Hippo Signaling Pathway. *Front Oncol* 2020; 10: 574383.
27. Zhang Y, Liu JL, Wang J. SAPCD2 promotes invasiveness and migration ability of breast cancer cells via YAP/TAZ. *Eur Rev Med Pharmacol* 2020; 24(7): 3786-3794.
28. Yang Z, Ding H, Pan Z, Li H, Ding J, Chen Q. YY1-induced activation of lncRNA DUXAP8 promotes proliferation and suppresses apoptosis of triple negative breast cancer cells through upregulating SAPCD2. *Cancer Biol Ther* 2021; 22(3): 216-224.
29. Sun Z, Mao Y, Zhang X, et al. Identification of ARHGEF38, NETO2, GOLM1, and SAPCD2 Associated With Prostate Cancer Progression by Bioinformatic Analysis and Experimental Validation. *Front Cell Dev Biol* 2021; 9: 718638.
30. Liu H, Zhu M, Li Z, et al. Depletion of p42.3 gene inhibits proliferation and invasion in melanoma cells. *J Cancer Res Clin* 2017; 143(4): 639-648.
31. Li PH, Cao WJ, Mao LL, Huang H, Zheng JN, Pei DS. p42.3 promotes cell proliferation and invasion in human Renal-Cell Carcinoma. *Int J Clin Exp Med* 2014; 7(12): 4959-4966.