The translation of a short open reading frame product within the human TUBA1B gene regulates cancer cell proliferation by importin-β

Yaling Tao[1*], Xuefeng Bai[2*], Yinjie Zhou[3], Yue Zhao[2], Liangwei Yang[3], Shun Zhang[3], Yan Sun[1], Xiaochun Huang[4], Edoardo Schneider[2], Anna Zampetaki[2], Andriana Margariti[5], Mauro Giacca[2], James N. Arnold[6], Lingfang Zeng[2#], Ting Cai[1,3#]

[1]Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, 315010, China
[2]School of Cardiovascular and Metabolic Medicine and Sciences, Faculty of Life Science and Medicine, King's College London, London, SE5 9NU, UK
[3]Key Laboratory of Diagnosis and Treatment of Digestive System Tumors of Zhejiang Province, Ningbo No.2 Hospital, Ningbo, 315010, China
[4]GDK Med Ltd, London, SM5 3QB
[5]The Wellcome-Wolfson Institute of Experimental Medicine, Belfast, BT9 7BL, UK
[6]Tumor Immunology Group, School of Cancer & Pharmaceutical Sciences, King's College London, London, SE1 1UL, UK.
*: Tao Y and Bai X equally contribute to this study.
#: Zeng L and Cai T equally contribute to this project.

Correspondence should be addressed to: Dr Lingfang Zeng, School of Cardiovascular and Metabolic Medicine and Sciences, Faculty of Life Science and Medicine, King's College London, London, SE5 9NU, UK, Tel:+44-20784885295; Fax: Tel:+44-20784885296; email: Lingfang.zeng@kcl.ac.uk. Or Professor Ting Cai, Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, 315010, China, email: Caiting@ucas.ac.cn.

Abstract

Understanding cancer biology and the molecular mechanisms underlying cancer progress are crucial for improving cancer treatment strategies. This study revealed a short open reading frame product (sORF1) alternatively translated from the human α-tubulin gene (TUBA1B), which has the completely different amino acids sequence from its main ORF product, α-tubulin 1B chain. sORF1 is highly expressed in cancer cell lines and positive sORF1 cells are found in gastric carcinoma. sORF1 interacts with more than one hundred proteins including importin β. It also acts as a modifier like small ubiquitin-like modifier (SUMO) by covalent binding. Knockdown of sORF1 in cancer cells decreases cell proliferation and tumorigenicity. Further study reveals that the loss of sORF1-importin β connections inhibits importin β guided β-catenin nucleus translocation, thus downregulates Wnt/β-catenin pathway. These findings indicate that gene TUBA1B is a polycistronic gene, able to translate two distinct proteins and the TUBA1B-sORF1 functions as a key tumorigenesis regulator. TUBA1B-sORF1 possesses great potential to be developed as therapeutic target and diagnosis biomarker for cancer.

Introduction

Understanding the complex mechanisms driving cancer progression is crucial for developing innovative therapeutic strategies that can effectively combat tumor recurrence and metastasis[1]. Chemotherapy has long been established as a cornerstone of cancer treatment, aiming to eliminate tumor cells through various mechanisms, including cell cycle arrest, DNA damage, and disruption of cellular structures critical for survival[2]. Microtubules are hollow cylindrical tubes made of heterodimers of α-tubulin and β-tubulin in the cytoplasm and play crucial roles in cell morphology, migration, intracellular transport, organelle stability and cell division. A dynamic transition exists between polymerization and depolymerization via the addition and subtraction of tubulin units, respectively, at both ends of the microtubule structure, depending on cellular function requirements[3]. The microtubule dynamic transition is more active in cancer cells than in normal cells. Among the chemotherapies, microtubule-targeting agents (MTAs) have emerged as a prominent therapeutic approach due to their ability to disrupt the dynamic transition either by inhibiting the addition of tubulin subunits such as Vinca alkaloids or preventing depolymerization such as Taxane, consequently suppressing cell mitosis and eventually leading to cell death[4, 5]. MTAs have been used as a first-line chemotherapy for breast, lung, and ovarian cancer[6]. The response to MTA therapy is generally favorable. However, cancer recurrence frequently occurs after a period of MTA treatment. Numerous studies have aimed to explain the mechanisms behind the recurrence phenomenon, and different theories have been proposed[7, 8, 9]. However, the underlying mechanisms remain unsolved.

Recent high-throughput sequencing studies have reported abnormal levels of the alpha tubulin 1b chain (TUBA1B) in cancer, establishing it as a valuable prognostic and overall survival indicator. Hu et al. found a strong correlation between elevated mRNA expression levels and the overall survival of patients with colorectal cancer[10]. Similarly, Xiong et al. observed significantly heightened TUBA1B mRNA levels in hepatocellular carcinoma[11]. In pancreatic ductal adenocarcinoma, Qin et al. constructed a classifier that identified TUBA1B mRNA as one of the promising biomarkers for patient prediction[12]. Despite its importance in maintaining normal cell function, the underlying mechanism and function of alpha tubulin protein in cancer remains elusive. Intriguingly, Linge et al. demonstrated reduced TUBA1B protein levels in uveal melanomas that metastasized compared to those that did not, indicating the intricate and multifaceted involvement of TUBA1B in cancer progression[13]. The discrepancy between mRNA and protein levels implies other mechanisms are involved in protein translation from the mRNAs.

In the last two decades, great progress has been made in understanding protein translation, including non-AUG codon initiation of translation, stop-codon readthrough, overlapping open reading frames, short ORFs (sORFs), functional ORFs in so-called noncoding RNAs, etc.[14, 15, 16, 17, 18, 19]. These events dramatically increase the cell proteome, possibly being involved in almost all physiological/pathophysiological processes, including those exploited in cancer. sORFs, defined as a subcategory of ORFs consisting of sequences shorter than 300 nucleotides or 100 codons, have

traditionally been overlooked or excluded from datasets due to the lack of their coding potential. In recent years, there has been a marked increase in the identification of sORFs with the capacity to encode proteins, thus driving the databases collecting information on sORFs[20, 21]. In this study, we found a novel sORF product overlapping with the human gene TUBA1B with 50 codons, named TUBA1B-sORF1 or sORF1.

Results

The short open reading frame product is translated in the TUBA1B gene

The aforementioned studies indicate that the TUBA1B gene may play a significant role in cancer. However, although its mRNA expression is upregulated, it is still unknown whether microtubule proteins directly contribute to cancer, and it remains unclear if there is a notable upregulation of α-tubulin protein. We hypothesized that even with an increase in mRNA expression, the level of α-tubulin expression does not significantly upregulated, but the product of its short open reading frame product (sORF1) is translated extensively through non-canonical translation mechanism to exert its function. To test this hypothesis, western blotting (WB) was performed to detect α and β tubulin in gastric and colorectal cancer tissues. As shown in Figure 1A, dramatic variation existed in α-tubulin but not in β-tubulin among different cancer tissues. The analysis of encoding sequences in α-tubulin genes reveals that a highly conserved sORF1 exists in all isoforms from both human and mouse TUBA genes (Figure S1). To assess whether this sORF1 was encodable, proteins with molecular weights between 3 kDa and 25 kDa from the cervical cancer cell line HeLa and gastric cancer cell line SNU-16 were subjected to proteomic analysis to search peptide fragments from the sORF1 sequences of different isoforms. Two peptide fragments corresponding to sORF1 of the TUBA1B gene (GenBank, BC018948) were detected composing the full length of sORF1 sequence (Figure 1B). One fragment contains six amino acids (LVSRLA) upstream start codon (Figure 1B peptide 1). These results suggest that sORF1 can be translated and that some of the translation can be initiated from upstream non-AUG codon. The full-length TUBA1B gene cDNA sequence was thereafter cloned into the pcDNA3.1 vector in which mCherry (red fluorescence protein) and GFP (green fluorescence protein) were inserted in-frame upstream of the stop codon of sORF1 and the main ORF (mORF), respectively, to confirm the alternative translation of the sORF1 product (Figure 1C upper left). Transfection of this plasmid into HeLa cells could detect both mCherry and GFP signals under inverted fluorescence microscopy where the majority of cells exhibit both green and red fluorescence (Figure 1C right). However, the distribution of green and red within the cells is inconsistent and individual cells show variations in the intensity of green and red fluorescence, indicating that the expression ratio of mORF and sORF1 is not uniform among different cells. This directly demonstrated the expression of sORF1 in cancer cells and suggested that the protein translation mechanism of sORF1 may differ from canonical translation. Subsequently, we generated polyclonal antibodies against sORF1 from rabbits by immunopeptides (immunogen: QVTRPLGEEMTPSTC). The sORF1-mCherry fusion protein was detected by Western blotting (WB) with anti-sORF1 and mCherry at the same position (Figure 1C left lower). sORF1 is expected

to have only 50 amino acids, corresponding to a product of approximately 5 kD. However, WB analysis using various cell line lysates revealed multiple bands of different molecular weights for endogenous sORF1, with the main band appearing around 50 kD, while no bands were detected below 10 kD (Figure 1D). Meanwhile, the expression level of sORF1 was higher in cancer cell lines compared to other cell lines. To confirm the specificity of these bands, we performed an overnight antibody blocking with immunopeptide against the antibody. After blocking, no bands were detected, indicating that the high-molecular-weight bands are all specific (Figure S2). Subsequently, immunofluorescence (IF) staining with unblocked and blocked antibodies in HeLa cells revealed the presence of specific sORF1 signals while there was no colocalization with α-tubulin (Figure 1E).

sORF1 expression is essential for cancer cell proliferation and tumor growth.

To unravel the roles of sORF1, we applied a gene editing strategy to induce sORF1 knockdown in SNU-16 cells (MsORF1). CRISPR/Cas9-mediated gene editing was performed with guide RNA (gtaacccaagtgagatccca, located in intron 1) and homologous recombination template (pUC57-TUBA1B-MsORF1) containing TUBA1B exon 1 and intron 1 sequence where AUG codon of sORF1 was removed and stop codons were introduced to disrupt sORF1 and PAM sequence GGG changed to GTG (Figure S3A). Polymerase chain reaction (PCR) plus restriction enzyme BsaAI digestion (Figure S3B), DNA sequencing (Figure S3C) and WB (Figure 2A) confirmed successful gene editing without disturbing alpha-tubulin. Notably, significantly decreased levels of OCT4 were observed, indicating a potential deficiency in pluripotency (Figure 2A). Compared with wild type SNU-16, MsORF1 cells decreased proliferation by 25% (Figure S4) and exhibited diminished ability to form colonies in soft agar assay (Figure 2B). Moreover, SNU-16 MsORF1 cells were unable to form a stable monoclonal cell line. They coexisted with wild-type cells and were gradually overridden by wild-type cells as passages were carried on. To increase the portion of the edited SNU-16 MsORF1 in the cell population, a pcDNA3.1-Neo plasmid was co-transfected during gene editing, and neomycin selection was applied. Thereafter, xenografts were generated in nude mice. WT and MsORF1 cells were inoculated subcutaneously into nude mice, and their tumor-forming ability was investigated. Tumors were palpable at day 10 and day 17 for WT and MsORF1, respectively (Figure 4C). The tumor growth from MsORF1 cells was significantly compromised compared to the parental WT line (Figure 4D). Xenograft tumor sections from both groups showed coincidence of OCT4 and sORF1 (Figure 4E). These results suggest that sORF1 may contribute to cancer cell proliferation and tumor growth both *in vitro* and *in vivo*.

The knockdown of sORF1 in SNU-16 cells resulted in significant regulation of a diverse range of proteins.

To delve deeper into the broader transcriptional alterations caused by the reduction of sORF1 proteins, RNA sequencing was conducted between unedited SNU-16 and SNU-16 MsORF1 cells. The transcriptional program of MsORF1 cells displayed broad re-programming of the cell (Figure 3A). 302 and 100 genes were downregulated and upregulated, respectively (Figure 3B), the majority of which were proliferation related. The downregulated genes are proliferation-promoting genes, while the upregulated

genes are proliferation-inhibitory genes. A full list of the changed genes is provided in Figure S5. The ontologies of the changed genes and their involvement in biological processes, cellular components and molecular functions (Figure 3C & Figure S6, Figure 3D & Figure S7) highlighted a broader reprogramming of the transcriptional processes in these cells as a result of the disruption of ORF1 coding, suggesting that sORF1 plays an important role in cancer cells. Quantitative RT-PCR of a few representative genes confirmed the RNA-seq data (Figure 3E).

sORF1 exhibits a complex interaction network and serves as a posttranslational modifier

As described above, WB analysis of anti-sORF1 detected multiple bands in cancer cells with the main band at approximately 50 kDa. Proteomics analysis could also detect peptide fragments of sORF1 from bands over 50 kDa. Cellular fractionation plus WB analysis revealed different sORF1 band patterns in the cytosol and nucleus which also appeared as high molecular weight complexes (Figure 4A). These observations suggest that sORF1 engages in interactions with a diverse range of proteins, potentially by covalent binding. To confirm the different bands were derived from covalent binding of sORF1 to proteins, a biotin-labeled sORF1-streptavidin agarose beads pulldown assay was first performed in a cell-free system. The bio-sORF1-accosiated proteins were eluted with dithiothreitol (DTT) containing loading buffer. The Coomassie blue staining detected multiple bands from biotin-labeled 50A, different from the BSA control (Figure 4B left). The reduction WB (with DTT) with anti-sORF1 or streptavidin-HRP detected unbound (molecular weight less than 15 kDa) and bound (higher molecular weight bands) sORF1 (Figure 4B middle and right). These results imply that the covalent binding may occur via non-S-S bonds. Proteomics analysis detected more than 100 sORF1-bound proteins, 32 of them are shown in figure 4C.

The system of proximity-dependent labelling of proteins, BioID, was then applied to biotinylate the sORF1-associated protein in living cells where the sequence of sORF1 was inserted into pc3.1-BirA(R118G)-HA[22] to generate the vector pc3.1-sORF1-BirA(R118G)-HA, which encodes a fusion protein of sORF1-BirA(R118G)-HA (Figure 4C). The biotin-labeled proteins covered 15 kDa to 200 kDa within the pc3.1-sORF1-BirA(R118G)-HA-transfected HeLa cells, as revealed by Coomassie blue staining and WB with streptavidin-HRP (Figure 4E), suggesting that sORF1 physically associates with a vast number of proteins. Proteomic analysis revealed more than one hundred proteins, including EGFR, proliferation-regulating proteins, ubiquitin-related ligase, multiple nucleoproteins (Figure 4E and S8). In summary, alongside its intricate protein-protein interaction network, the intracellular sORF1 product acts as a posttranslational modifier, operating in a ubiquitin-like manner to impact cellular functions.

sORF1 interacts with importin β to regulate β-catenin translocation

Previous studies have shown that significant inhibition of malignancy in cancer cells occurs when sORF1 was knocked down. Additionally, sORF1 appeared to exert its effects by influencing other proteins. To investigate the specific mechanisms by which sORF1 maintains cancer cell proliferation and tumorigenicity, we performed enrichment analysis of the protein-protein interaction data obtained through BioID. We found that sORF1 was related in various key signaling pathways, including WNT/β-

Catenin canonical signaling pathway, interferon signaling pathway et a (Figure S9). In HeLa cells with reduced sORF1 expression (M1), we observed enhanced β-Catenin expression in the cytoplasm and decreased expression in the nucleus. Semi-quantitative analysis of three replicates showed a significant reduction of approximately 50% in the nuclear/cytoplasmic ratio of β-Catenin in M1 compared to wild-type cells, indicating that the loss of sORF1 indeed results in the loss of β-Catenin translocation, subsequently affecting downstream signaling pathway activation (Figure 5A). However, we did not find a direct connection between sORF1 and β-Catenin. Instead, we discovered an interaction between sORF1 and importin-β which was verified by peptide pull-down and BirA assays (Figure 5B and 5C). Importin-β can facilitate the nuclear transport of various proteins, including β-Catenin[23]. Subsequently, we performed co-immunoprecipitation (Co-IP) combined with semi-quantitative analysis, and the results showed a significant decrease in the interaction between β-Catenin and importin-β in sORF1-knockdown HeLa cells (Figure 5D). These results suggest that the ability of importin-β to transport β-Catenin into the nucleus is significantly inhibited in the absence of sORF1, highlighting the role of sORF1 in assisting the nucleus translocation of importin-β and in maintaining canonical WNT/β-catenin signaling pathway.

sORF1 protein is upregulated in human gastric cancer tissue and serum

As described above, sORF1 is involved in tumor growth in the xenograft model. Further immunohistochemistry (IHC) staining were performed on formalin-fixed paraffin-embedded (FFPE) sections from 10 groups of gastric carcinoma patients, with both tumor and paratumor tissues being analyzed. A small proportion of sORF1-positive cells were detected in both tumor and paratumor tissues (Figure 6A). The semiquantitative analysis showed a higher ratio of sORF1-positive cells in tumor tissues than in paratumor tissues, with no significant difference, possibly due to the small number of samples (Figure 6B). However, when the tumor grading was evaluated (Figure 6C), a significant increase in sORF1-positive cells was detected in the paratumor area in the undifferentiated-grade tumor group compared to the differentiated-grade group (Figure 6D). Furthermore, an enzyme-linked immunosorbent assay (ELISA) revealed that there was a significantly higher level of serum sORF1 product in cancer patients than that in healthy individuals (Figure 6E). IF staining revealed that sORF1 coexisted with OCT4 but not with α-tubulin in double IF staining of gastric carcinoma tissue (Figure 6F), suggesting a translation switch from α-tubulin to sORF1 in CSCs that may contribute to tumorigenesis. These results highlight the potential of sORF1 as a novel therapeutic target and prognostic biomarker for cancer.

Discussion

In this study, we demonstrated that the sORF1 protein is alternatively translated in cancer cells, functioning independently of its main open reading frame product (mORF), alpha tubulin 1b. Subsequent investigations have revealed that sORF1 exhibits extensive protein-protein interactions and serves as a modifier by forming covalent bonds with other proteins. Remarkably, we observed a significant reduction in

proliferation and tumorigenesis both in vitro and in vivo upon knocking down sORF1 expression in cancer cells. Mechanistically, we identified an interaction between sORF1 and importin beta, which, when disrupted, leads to a notable decrease in the formation of the beta-catenin/importin beta complex. Consequently, this impedes the translocation of beta-catenin into the nucleus and effectively inhibits the downstream Wnt/beta-catenin pathway.

Non-canonical translation or alternative translation widely exists in prokaryotic genomes to increase genome capacity[24]. Recent progress in proteogenomic and ribosomal profiling has revealed that non-AUG codon-initiated translation, especially in sORFs, widely exists in eukaryotic genomes, including mammalian genomes[25, 26, 27]. Increasing genes were discovered that can produce not only its classical protein or main open reading frame (mORF) product, but also an overlapping alternative open reading frame (ORF) or short open reading frame (sORF) product with a completely different amino acid sequence. These genes have been named duo-coding genes due to their ability to encode 2 distinct proteins from the same sequence[16, 28]. Previous studies of duo-coding genes reported that the alternative product from POLG and RPL36 also used CTG and GTG instead of AUG as start codons for translation[28]. Translation initiation from non-AUG codons may involve canonical Kodak sequences[29], internal ribosome entry sites[30] or chemical modification of codons[31]. This alternative translation initiation appears to have a preference for non-AUG codon-guided initiation. Although not extensively studied in this article, the proteomic data shown in Figure 1B detected the presence of the six amino acids (LVSRLA) upstream the first methionine (AUG) of sORF1, where a strong Kodak sequence[32] exists in the CUG codon (ggccaggctggtg). Further detailed investigation is required to clarify this phenomenon and define the function of the additional upstream peptide sequence. This study has demonstrated that the TUBA1B gene is at least duo cistronic.

Translation of sORFs usually occurs only under special conditions such as stimulation[33] or stress[34]. In this study, IF analyses revealed that almost all cancer cells of HeLa and SNU-16 expressed sORF1, while disruption of sORF1 suppressed cancer cell proliferation *in vitro* and tumor growth *in vivo*. These observations suggest that sORF1 may be constitutively translated in cancer cells. Co-translation of sORF and mORF was reported in other genes, such as MUC1[19] and Ataxin-1[18]. The products from sORFs and mORFs may act independently[19] or interact with each other[18]. sORF1 exists in both the cytosol and nucleus, while α-tubulin exists exclusively in the cytosol. No interaction was detected in the IF or IP or BioID studies. This suggests that sORF1 and α-tubulin work independently, similar to MUC1-ARF versus MUC1-TM[19]. Although the products of sORF1 and mORF do not have functional connections, they are translated from the same mRNA transcript. Traditionally, it has been believed that an increase in mRNA expression directly correlates with an increase in protein expression. However, for mRNAs that can produce two different products, it remains unclear which protein level is increased by upregulated levels of mRNA expression. As mentioned earlier, several sequencing studies have identified dysregulated levels of TUBA1B mRNA. However, no research has established the specific function of α-tubulin in cancer. Further investigation is needed to determine whether the high expression level

of TUBA1B leads to increased expression of sORF1 protein, thereby maintaining the necessary conditions for cancer progression. Based on the current findings, the function of sORF1 appears to rely on its interaction with binding proteins to regulate their activity. From the BioID[22] and biotin-peptide assay, we have identified over a hundred proteins associated with sORF1. Additionally, we have discovered that sORF1 can serve as a small modifier covalently attached to larger protein molecules, acting as a post-translational modifier. These findings indicate that sORF1 may play a role in protein regulation and modification through its interactions with binding proteins. We have uncovered a regulatory mechanism of sORF1 involving the modulation of importin-β in nuclear transport function, thereby controlling the translocation of β-catenin. Although in HeLa cells, reducing sORF1 significantly decreased the binding functionality of importin-β, it is still unclear whether sORF1 acts as a facilitator or a necessary small molecule for importin-β to maintain its function. Extensive research is needed to validate the exact functions and mechanisms of sORF1. Posttranslational modification plays a critical role in modulating protein stability, cellular location and function. Ubiquitination[35] and ubiquitin-like modification (SUMOylating)[36] are two types of unique posttranslational modifications that occur by covalently adding one or more protein molecule(s) to target proteins. As ubiquitin/SUMOylation ligase was detected from the sORF1-associated protein, although the GG sequence does not exist in the sORF1 sequence, a non-standard ubiquitination-like mechanism may exist. Further detailed investigation of the identities of the target proteins may elicit the mode of action of TUBA-sORF1 in driving tumor growth.

In clinical research, the expression pattern of sORF1 in gastric cancer tissue differs from that observed in cell lines where sORF1-positive cells are present but in limited quantities. Tumor cells exhibit heterogeneity, and recent single-cell based studies have identified the potential existence of cancer stem cells/cancer-initiating cells and metastasis-initiating cells within the bulk tumor population[37, 38, 39, 40, 41]. Distinguishing these distinct cell populations can aid in developing therapies specific to subgroups of cancer cells[42, 43]. Although a detailed investigation of sORF1-specific cells in cancer tissue is currently lacking, there is promising evidence suggesting their significance from the results of xenograft model where a significant loss of tumorigenic capability is observed when sORF1 is knocked down. Additionally, the presence of elevated levels of sORF1 in serum samples from tumor patients supports this possibility. Such investigations could potentially unlock new avenues for the development of targeted therapies and improve treatment strategies for cancer patients.

This study provides novel insights into non-canonical translation and its functional product in TUBA1B gene suggesting a new molecular mechanism underlying cancer development and highlights new potential targets for therapeutic intervention in cancer treatment.

References

1. Sung H*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).

2. Mitchison TJ. The proliferation rate paradox in antimitotic chemotherapy. *Mol Biol Cell* **23**, 1-6 (2012).

3. Wilson L, Jordan MA. Microtubule dynamics: taking aim at a moving target. *Chemistry & biology* **2**, 569-573 (1995).

4. Dumontet C, Jordan MA. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat Rev Drug Discov* **9**, 790-803 (2010).

5. Wordeman L, Vicente JJ. Microtubule Targeting Agents in Disease: Classic Drugs, Novel Roles. *Cancers (Basel)* **13**, (2021).

6. Čermák V*, et al.* Microtubule-targeting agents and their impact on cancer treatment. *European journal of cell biology* **99**, 151075 (2020).

7. Smith ER, Wang JQ, Yang DH, Xu XX. Paclitaxel resistance related to nuclear envelope structural sturdiness. *Drug Resist Updat* **65**, 100881 (2022).

8. Kavallaris M. Microtubules and resistance to tubulin-binding agents. *Nature Reviews Cancer* **10**, 194-204 (2010).

9. Dan VM, Raveendran RS, Baby S. Resistance to intervention: paclitaxel in breast cancer. *Mini Reviews in Medicinal Chemistry* **21**, 1237-1268 (2021).

10. Hu J*, et al.* Dynamic Network Biomarker of Pre-Exhausted CD8(+) T Cells Contributed to T Cell Exhaustion in Colorectal Cancer. *Front Immunol* **12**, 691142 (2021).

11. Xiong D-d*, et al.* High throughput circRNA sequencing analysis reveals novel insights into the mechanism of nitidine chloride against hepatocellular carcinoma. *Cell death & disease* **10**, 658 (2019).

12. Qin D, Zhao Y, Guo Q, Zhu S, Zhang S, Min L. Detection of Pancreatic Ductal Adenocarcinoma by A qPCR-based Normalizer-free Circulating Extracellular Vesicles RNA Signature. *J Cancer* **12**, 1445-1454 (2021).

13. Linge A*, et al.* Differential expression of fourteen proteins between uveal melanoma from patients who subsequently developed distant metastases versus those who did not. *Investigative ophthalmology & visual science* **53**, 4634-4643 (2012).

14. Adhikari S*, et al.* A high-stringency blueprint of the human proteome. *Nature communications* **11**, 5301 (2020).

15. Wright BW, Yi Z, Weissman JS, Chen J. The dark proteome: translation from noncanonical open reading frames. *Trends in Cell Biology* **32**, 243-258 (2022).

16. Cao X*, et al.* Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24. *Nature Communications* **12**, 508 (2021).

17. Vanderperre B*, et al.* An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein. *The FASEB Journal* **25**, 2373-2386 (2011).

18. Bergeron D, Lapointe C, Bissonnette C, Tremblay G, Motard J, Roucou X. An out-of-frame overlapping reading frame in the ataxin-1 coding sequence encodes a novel ataxin-1 interacting protein. *Journal of Biological Chemistry* **288**, 21824-21835 (2013).

19. Chalick M*, et al.* MUC1-ARF—a novel MUC1 protein that resides in the nucleus and is expressed by alternate reading frame translation of MUC1 mRNA. *PLoS One* **11**, e0165031 (2016).

20. Guerra-Almeida D, Nunes-da-Fonseca R. Small open reading frames: how important are they for molecular evolution? *Frontiers in Genetics* **11**, 574737 (2020).

21. Brunet MA*, et al.* OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Research* **49**, D380-D388 (2021).

22. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of cell biology* **196**, 801-810 (2012).

23. Hwang WY, Kostiuk V, Gonzalez DP, Lusk CP, Khokha MK. Kap-beta2/Transportin mediates beta-catenin nuclear transport in Wnt signaling. *Elife* **11**, (2022).

24. Cao X, Slavoff SA. Non-AUG start codons: expanding and regulating the small and alternative ORFeome. *Experimental cell research* **391**, 111973 (2020).

25. Andreev DE, Loughran G, Fedorova AD, Mikhaylova MS, Shatsky IN, Baranov PV. Non-AUG translation initiation in mammals. *Genome Biology* **23**, 111 (2022).

26. Green KM, Miller SL, Malik I, Todd PK. Non-canonical initiation factors modulate repeat-associated non-AUG translation. *Human molecular genetics* **31**, 2521-2534 (2022).

27. Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in

eukaryotes. *Genes & development* **31**, 1717-1731 (2017).

28. Khan YA*, et al.* Evidence for a novel overlapping coding sequence in POLG initiated at a CUG start codon. *BMC genetics* **21**, 1-16 (2020).

29. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N. Unanticipated antigens: translation initiation at CUG with leucine. *PLoS biology* **2**, e366 (2004).

30. Touriol C*, et al.* Generation of protein isoform diversity by alternative initiation of translation at non‐AUG codons. *Biology of the Cell* **95**, 169-178 (2003).

31. Fujita Y*, et al.* Translational recoding by chemical modification of non-AUG start codon ribonucleotide bases. *Science Advances* **8**, eabm8501 (2022).

32. Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic acids research* **15**, 8125-8148 (1987).

33. Yang J*, et al.* A histone deacetylase 7-derived peptide promotes vascular regeneration via facilitating 14-3-3γ phosphorylation. *Stem cells* **38**, 556-573 (2020).

34. Merry TL*, et al.* Mitochondrial-derived peptides in energy metabolism. *American Journal of Physiology-Endocrinology and Metabolism* **319**, E659-E666 (2020).

35. Pickart CM, Eddins MJ. Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1695**, 55-72 (2004).

36. Johnson E. Protein modification by SUMO. Annu. Rev. Biochem. (2004).

37. Massague J, Ganesh K. Metastasis-initiating cells and ecosystems. *Cancer discovery* **11**, 971-994 (2021).

38. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-737 (1997).

39. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-3988 (2003).

40. Singh SK*, et al.* Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**, 5821-5828 (2003).

41. Shimokawa M*, et al.* Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature* **545**, 187-192 (2017).

42. Takebe N*, et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat Rev Clin Oncol* **12**, 445-464 (2015).

43. Laszlo GS, Estey EH, Walter RB. The past and future of CD33 as therapeutic target in acute myeloid leukemia. *Blood Rev* **28**, 143-153 (2014).

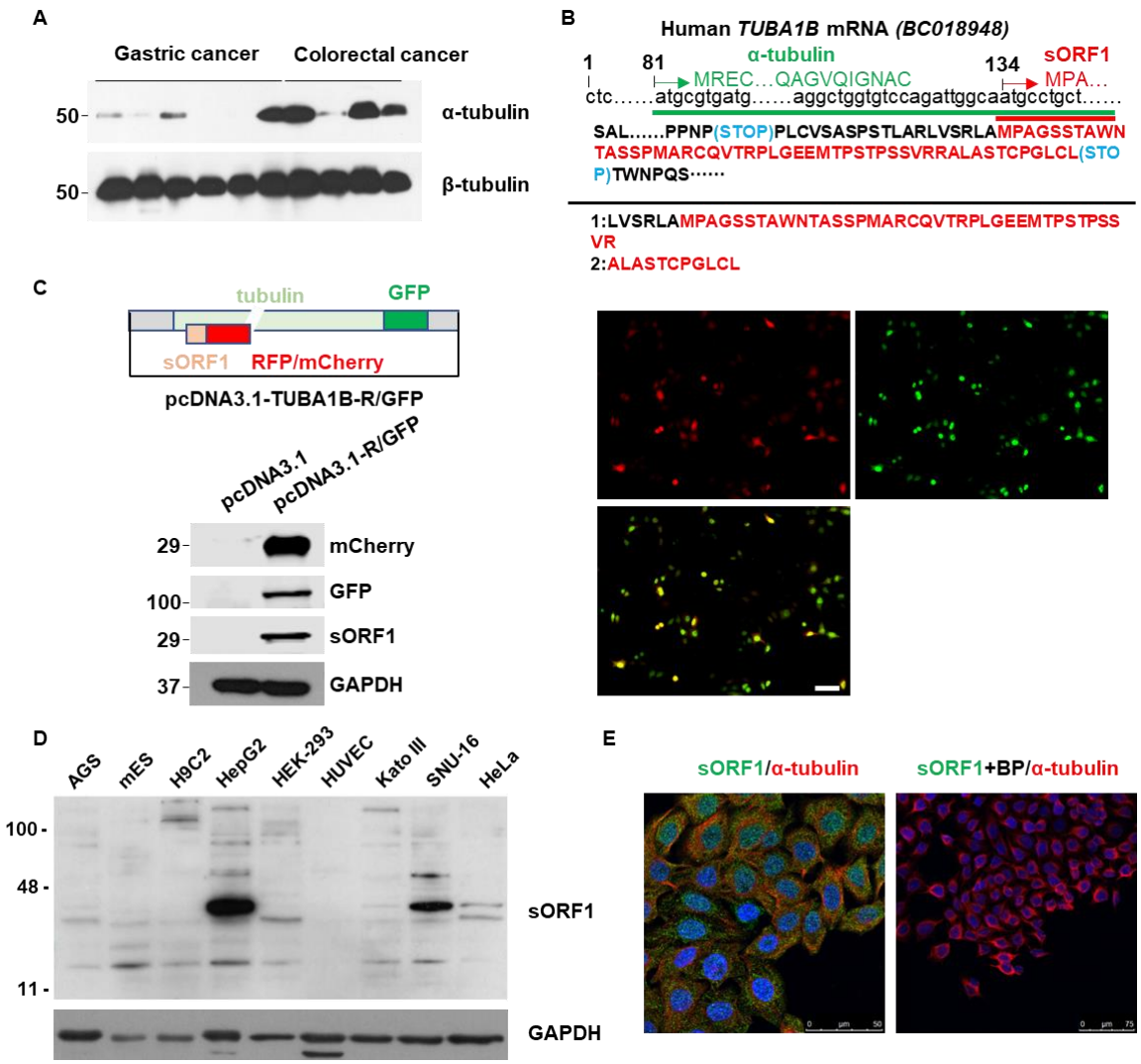


Figure 1: A short open reading frame product was alternatively translated from the human TUBA1B gene. (A) Detection of variable levels of α-tubulin and consistent levels of β-tubulin in gastric and colorectal cancer tissues using WB analysis. (B) Schematic representation illustrating the main ORF (α-tubulin 1B, shown in green) and sORF1 (shown in red) in TUBA1B mRNA (upper panel), along with two peptide fragments identified through proteomic analysis of HeLa and SNU-16 cell lysates (lower panel). (C) Transfection of HeLa cells with pcDNA3.1-TUBA1B-R/GFP plasmid (upper left) and subsequent WB analysis (lower left) with indicated antibodies, and observation under an inverted fluorescence microscope (right). The mCherry signal represents sORF1, while GFP signifies α-tubulin. Scale bar: 100 μm. (D) Identification of TUBA1B-sORF1 in multiple cancer cell lines through blotting with a peptide-generated antibody. Cell lines include AGS (human gastric cancer), mES (mouse embryonic stem cells), H9C2 (rat cardiomyocytes), HepG2 (human liver carcinoma), HEK-293 (human embryonic kidney epithelial cells), HUVEC (human umbilical vein endothelial cells), Kato III (human gastric cancer), SNU-16 (human gastric cancer), and HeLa (human cervical cancer). (E) Immunofluorescent staining demonstrating specific sORF1 signal in HeLa cells. Immunogen blocking of sORF1 peptide to antibody completely diminished fluorescence by incubating the anti-sORF1 antibody overnight with synthetic sORF1 antigen at a ratio of 1μg:10μg in PBS.

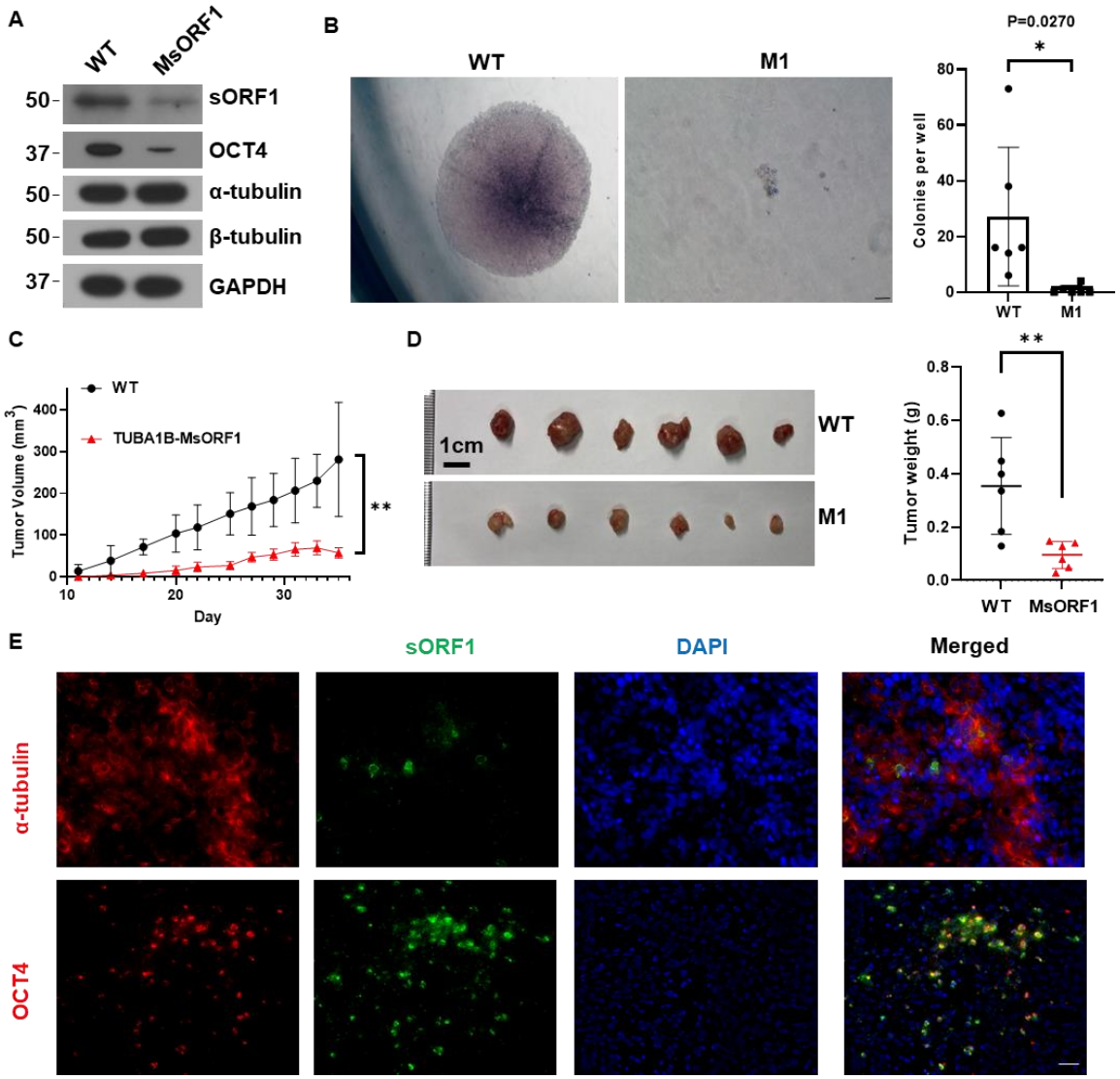


Figure 2: sORF1 was essential for cancer cell proliferation and tumor growth. (A) WB was applied to cell lysates from unedited (WT) and mutant (MsORF1) SNU-16 cells with the indicated antibodies. (B) WT and MsORF1 cells were seeded in soft agarose medium at $1x10^3$ cells/well in 24-well plates (6 wells each). The left panel shows representative images, and the right panel shows colony numbers on day 15. (C-E) SNU-16 WT and MsORF1 cells were xenografted into nude mice, followed by daily measurement of tumor size (C), tumor tissues collected with a photo taken at day 35 (D, left), tumor weight statistical analysis (D, right) and IF with the indicated antibodies (E). *: $p<0.05$; **: $p<0.01$. Scale bar: 100 μm.

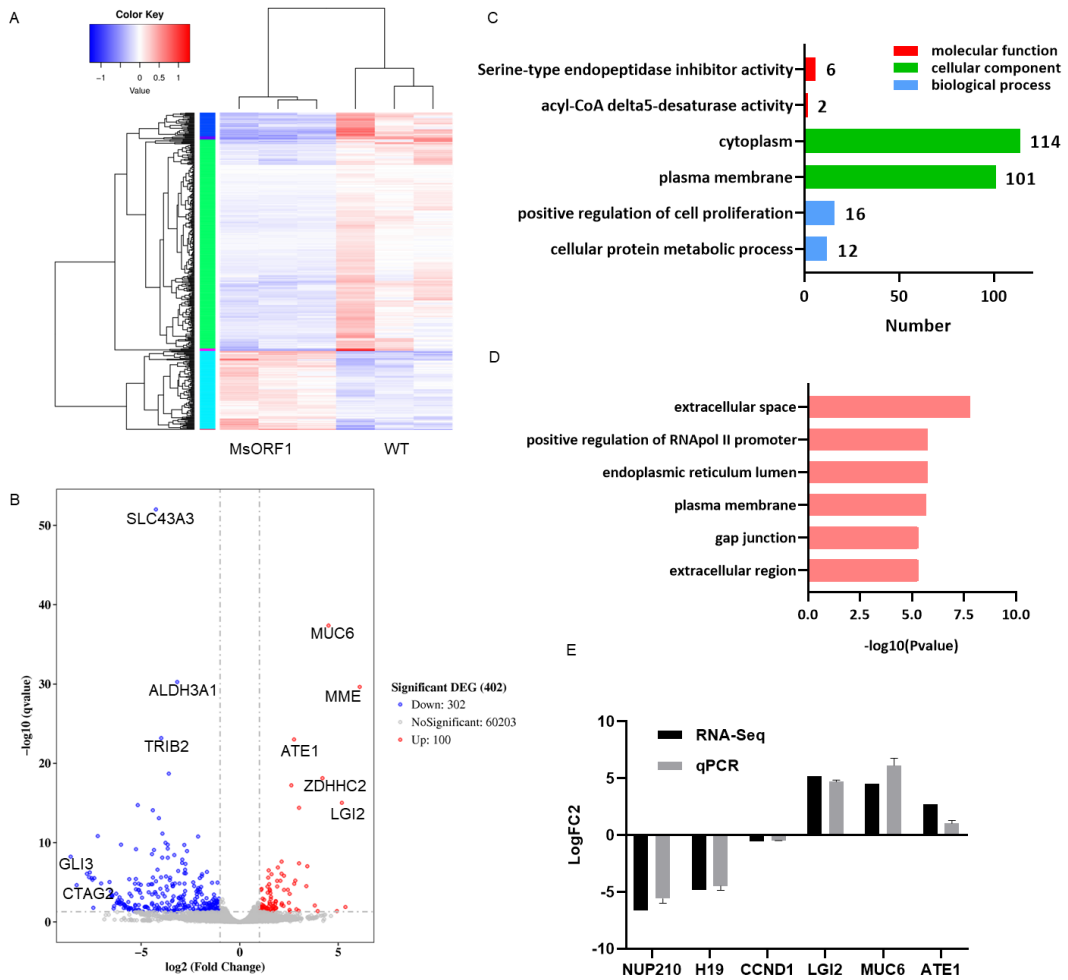
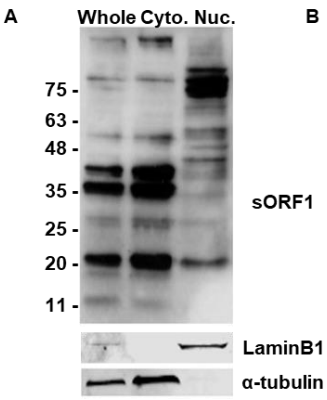
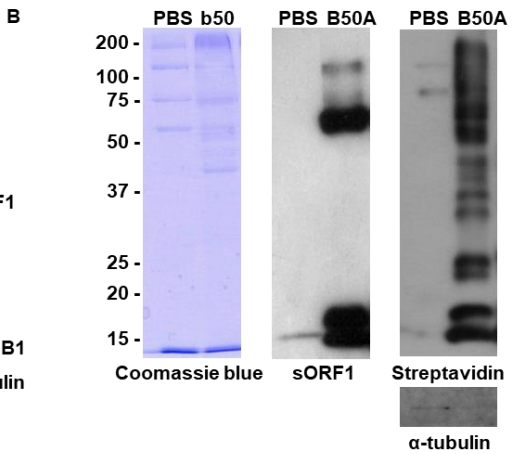


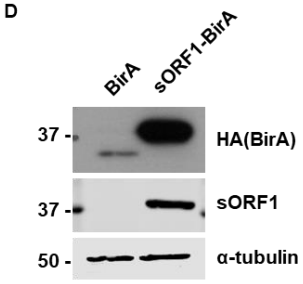
Figure 3: RNA sequencing analysis revealed 402 genes modulated by the disruption of TUBA1B-sORF1 in gene-edited SNU-16 MsORF1 cells. (A, B) Hierarchical clustering diagram (A) and volcano plot (B) of differentially expressed genes between SNU-16 WT and MsORF1 cells from three biological replicates. Downregulated and upregulated genes are indicated in blue and red, respectively. In the volcano plot, the x-axis represents the log2-fold change in gene expression, and the y-axis represents the negative log10 of the q-value. (C) Gene Ontology (GO) annotation and enrichment analysis of differentially expressed genes between SNU-16 WT and MsORF1 cells, where the y-axis represents the enriched GO term, and the x-axis represents the number of differentially expressed genes in that term. Different colors are used to indicate biological processes, cellular components, and molecular functions (top two terms of each shown). (D) Bar chart of GO enrichment P-value, with the y-axis representing the enriched GO term and the x-axis representing the negative log10 of the p-value of that term (top six terms shown). (E) Validation of 6 significantly regulated genes in SNU-16 MsORF1 cells using real-time qPCR.

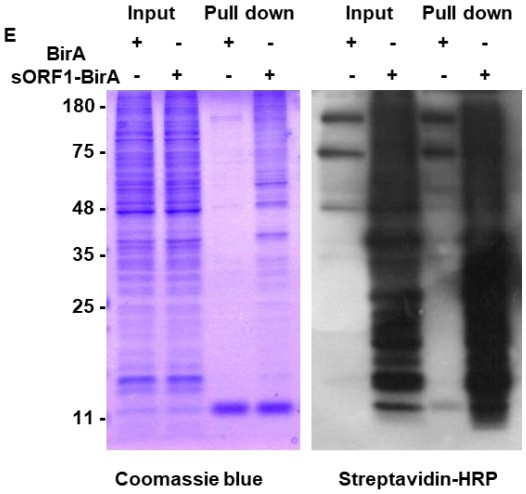




C

Protein	MW	PIP	Protein	MW	PIP
KPNB1	97 kDa	100%	SLC3A2	68 kDa	100%
HPDL	39 kDa	100%	USP39	65 kDa	100%
L1CAM	140 kDa	100%	NXF1	70 kDa	100%
EPRS1	171 kDa	100%	FAM98B	46 kDa	100%
SLC2A1	54 kDa	100%	FBXO21	72 kDa	100%
VPS35	92 kDa	100%	HLA-A	41 kDa	100%
GCLC	73 kDa	100%	HTRA2	49 kDa	100%
SQOR	50 kDa	100%	WDR43	75 kDa	100%
HAO2	39 kDa	99%	TRIM29	66 kDa	100%
ATP5PO	23 kDa	100%	ZADH2	40 kDa	100%
LGALS3BP	65 kDa	100%	BASP1	23 kDa	100%
WASF2	54 kDa	100%	DNM1L	82 kDa	100%
CDKN2AIP	61 kDa	100%	HADHB	51 kDa	100%
BCAP31	28 kDa	100%	HSPB1	23 kDa	100%
PAPSS1	71 kDa	100%	RAVER1	64 kDa	100%
PAPSS2	70 kDa	99%	HYOU1	111 kDa	100%





F

Protein	MW	PIP	Protein	MW	PIP
EGFR	134 kDa	99%	AKAP2	95 kDa	100%
KPNB1	97 kDa	99%	SEC16A	252 kDa	100%
DDX3X	73 kDa	100%	SHANK1	225 kDa	100%
ATP5F1A	60 kDa	100%	IQGAP1	189 kDa	99%
POP1	115 kDa	100%	USO1	108 kDa	99%
PPP1R12A	115 kDa	100%	IARS1	145 kDa	99%
SLC7A5	55 kDa	100%	RAB35	23 kDa	99%
CRYBG3	331 kDa	100%	TUBA1C	50 kDa	100%
RPL37	11 kDa	99%	RNF113A	39 kDa	99%
MISP	75 kDa	100%	WDHD1	126 kDa	99%
CDKN2AIP	61 kDa	100%	SLC3A2	68 kDa	99%
NDUFS1	79 kDa	100%	CBSL	61 kDa	99%
PPFIBP1	114 kDa	100%	MAOB	59 kDa	99%
SEPTIN9	65 kDa	100%	TSFM	35 kDa	99%
CPSF7	52 kDa	100%	RAB8A	24 kDa	99%
TASOR	189 kDa	100%	KARS1	68 kDa	99%
EHMT1	141 kDa	100%	PPP6R3	98 kDa	99%

Figure 4: Exploration of sORF1's protein-protein interactions and its role as a posttranslational modifier. (A) sORF1 exists differently in the cytosol and nucleus. Whole: whole cell lysate; cyto: cytosolic fraction; nuc: nuclear fraction. LaminB1 and α-tubulin were included to indicate nuclear and cytosolic fractions, respectively. (B, C) sORF1 interacted and covalently bound to different proteins. HeLa cell lysate (500 μg) was incubated with biotin-labeled sORF1 (B50A: N-terminal biotinylated sORF1, BSA included as vehicle control), pulled down with streptavidin agarose beads, followed by DTT-reduced SDS-PAGE, stained by Coomassie blue solution (B, left), or WB with anti-sORF1 (B middle), anti-α-tubulin antibodies and streptavidin-HRP (B, right). Pull down proteins were identified by proteomic analysis (C). PIP: protein identification probability. (D-F) HeLa cells were transfected with the pcDNA3.1-TUBA1-sORF1-BirA(R118G)-HA vector followed by WB with the indicated antibodies (D); streptavidin pull-down with Coomassie blue staining and streptavidin-HRP detection (E); and protein identification analysis (F).

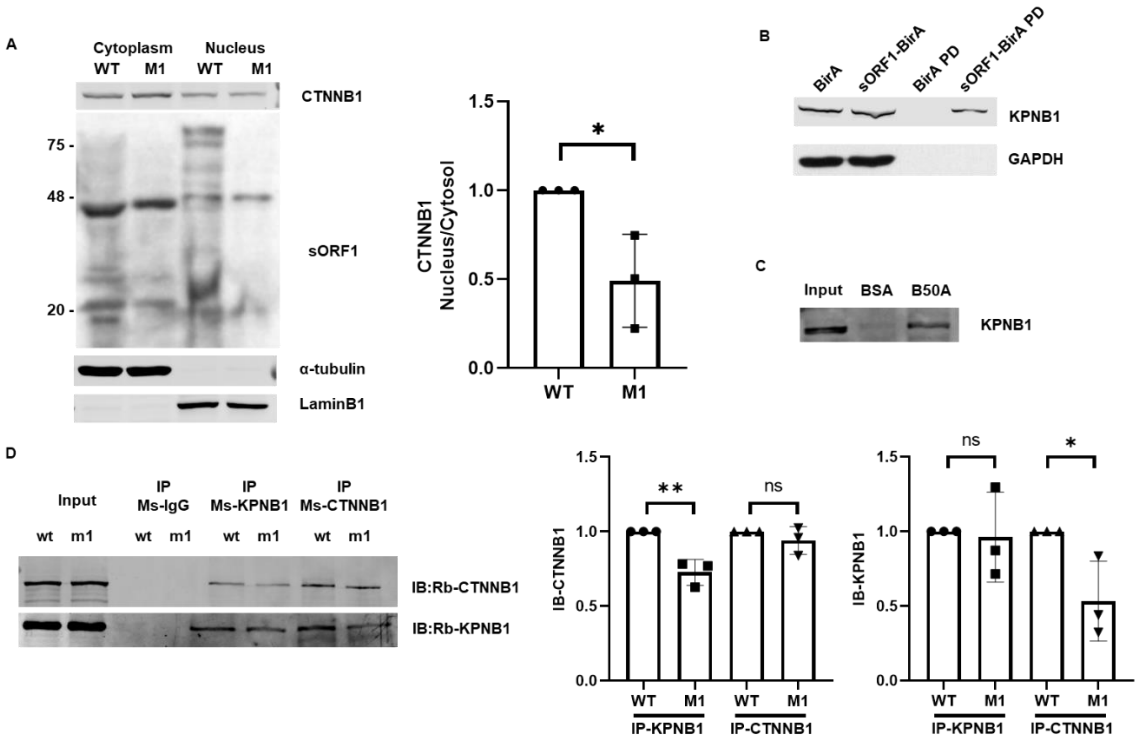
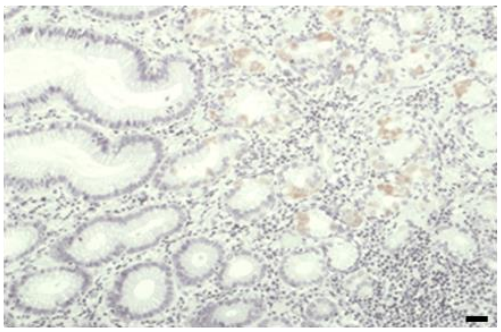


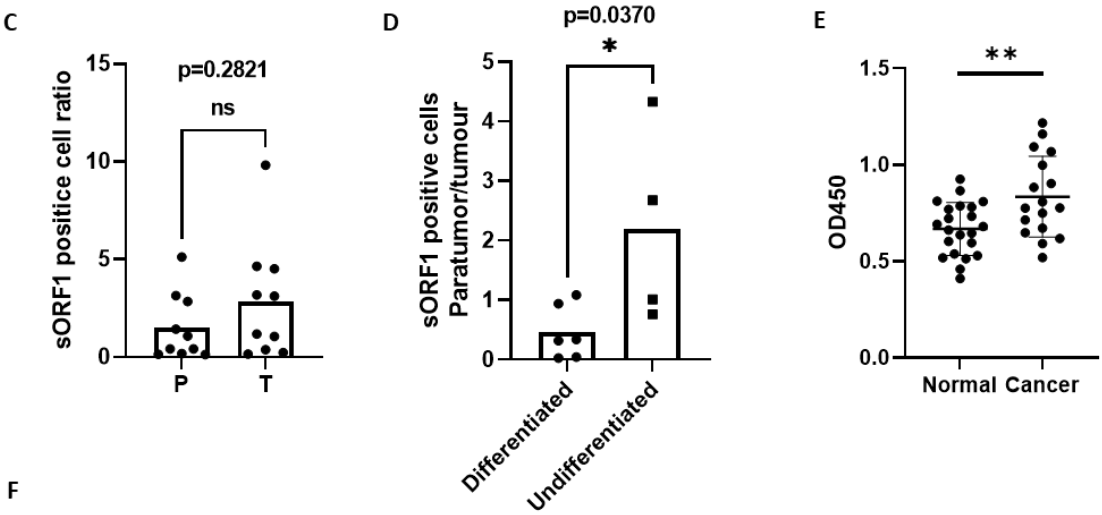
Figure 5: sORF1 knock down disrupted importin-β guided β-catenin nucleal translocation. (A) WB analysis showing the levels of β-catenin and sORF1 in different cellular fractions in the sORF1 knockdown HeLa cells (M1). Semi-quantification analysis indicates a decrease in nuclear/cytosolic fraction of β-catenin (CTNNB1) in the knockdown group. (B) Streptavidin pull-down assay performed on HeLa cells transfected with BirA or BirA-sORF1, followed by western blotting with importin-β (KPNB1). (C) Detection of importin-β in the biotin-labeled sORF1 (B50A) pull-down assay. (D) Co-immunoprecipitation assay displayed a reduced interaction between β-catenin and importin-β in sORF1 knockdown HeLa cells revealed by western blotting (left) and semi-quantification analysis based on three biological replicates (right). *: $p<0.05$.

A



B

Group	Positive cells (%) Paratumor	Tumor	Tumor grading
1	0.142	0.151	Well differentiated
2	3.142	3.115	Undifferentiated
3	0.134	4.645	Moderately differentiated
4	2.845	1.0624	Undifferentiated
5	1.432	4.519	Poorly differentiated
6	5.116	1.181	Undifferentiated
7	1.0762	3.187	Poorly differentiated
8	0.417	9.817	Poorly differentiated
9	0.417	0.385	Poorly differentiated
10	0.178	0.235	Undifferentiated



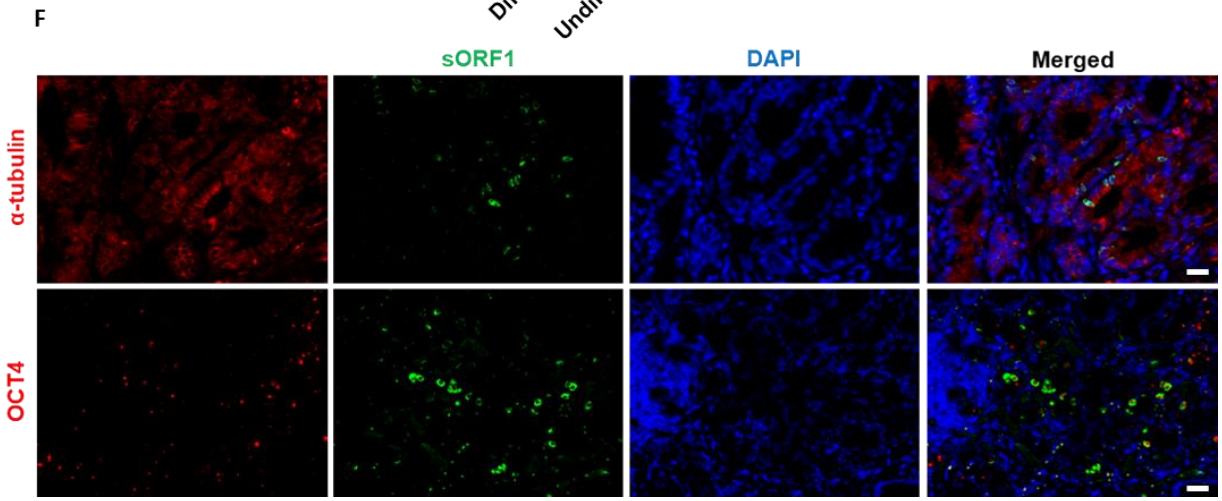


Figure 6: sORF1 protein contributed to tumorigenesis. (A) Representative images of immunohistochemistry staining of sORF1 in gastric carcinoma tissue sections. (B) A comparison of sORF1-positive cells between paratumor (P) and tumor (T) tissues. (C) The percentage of sORF1-positive cells in the ten groups of gastric carcinoma tumor and paratumor sections with tumor grading. (D) A comparison of sORF1-positive cells between differentiated and undifferentiated tumors. *: $p<0.05$. (E) Enzyme-linked immunosorbent assay revealed a higher level of serum sORF1 in cancer patients (n=17) than in healthy individuals (n=21). **: $p<0.01$. (F) Representative images of IF staining with the indicated antibodies. DAPI was included to counterstain nuclei. Scale bar: 100 μm.